Efficient Mechanisms for Level-*k* Bilateral Trading

Vincent P. Crawford

University of Oxford, All Souls College, and University of California, San Diego

Thanks to many people for valuable conversations and advice; and to Rustu Duran for outstanding research assistance.

The research leading to these results received primary funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 339179. The contents reflect only the author's views and not the views of the ERC or the European Commission, and the European Union is not liable for any use that may be made of the information contained therein. The University of Oxford and All Souls College also provided important funding.

Revised 10 May 2018


http://alicesg.blogspot.com/

Introduction

The paper revisits Myerson and Satterthwaite's 1983 *JET* ("MS") classic analysis of the design of incentive-efficient mechanisms for bilateral trading with independent private values, inspired by Chatterjee and Samuelson's 1983 *OR* ("CS") positive analysis.

MS assumed that traders will play any desired Bayesian Nash equilibrium in the game created by the chosen mechanism.

I replace MS's equilibrium assumption with a structural nonequilibrium "level-*k*" model of strategic thinking, meant to describe initial responses to games; and study direct mechanisms.

To focus on nonequilibrium thinking, I maintain standard rationality assumptions regarding decisions and probabilistic judgment.

Motivation

- Mechanism design often creates novel games, weakening the learning justification for equilibrium; yet the design may need to work well the first time

- Even if learning is possible, design may create games complex enough that convergence to equilibrium is behaviorally unlikely

Motivation

- Mechanism design often creates novel games, weakening the learning justification for equilibrium; yet the design may need to work well the first time

- Even if learning is possible, design may create games complex enough that convergence to equilibrium is behaviorally unlikely

We usually assume equilibrium anyway, perhaps because:

- We doubt we can identify a credible basis for analysis among the enormous number of possible nonequilibrium models

- We doubt that any nonequilibrium model could systematically out-predict a rational-expectations notion such as equilibrium

But…

- There is now a large body of experimental research that studies strategic thinking by eliciting subjects' initial responses to games (surveyed in Crawford, Costa-Gomes, and Iriberri 2013 *JEL*)

- The evidence suggests that people's thinking in novel or complex games does not follow the fixed-point or indefinitely iterated dominance reasoning that equilibrium often requires

  (Learning can still make people converge to something that *we* need fixed-point reasoning to characterize; the claim is that fixed-point reasoning doesn't *directly* describe people's thinking)

But…

- There is now a large body of experimental research that studies strategic thinking by eliciting subjects' initial responses to games (surveyed in Crawford, Costa-Gomes, and Iriberri 2013 *JEL*)

- The evidence suggests that people's thinking in novel or complex games does not follow the fixed-point or indefinitely iterated dominance reasoning that equilibrium often requires

  (Learning can still make people converge to something that *we* need fixed-point reasoning to characterize; the claim is that fixed-point reasoning doesn't *directly* describe people's thinking)

- To the extent that people do not follow equilibrium logic, they must find another way to think about the game

- Much evidence points to a class of nonequilibrium "level-*k*" or "cognitive hierarchy" models of strategic thinking

Level-*k* models

In a level-*k* model people follow rules of thumb that:

- Anchor their beliefs in a naïve model of others' response to the game, called *L0*, often uniform random over feasible decisions

and

- Adjust their beliefs via a small number (*k*) of iterated best responses; so *L1* best responds to *L0*, *L2* to *L1*, and so on

Level-*k* models

In a level-*k* model people follow rules of thumb that:

- Anchor their beliefs in a naïve model of others’ response to the game, called *L0*, often uniform random over feasible decisions

and

- Adjust their beliefs via a small number (*k*) of iterated best responses; so *L1* best responds to *L0*, *L2* to *L1*, and so on

People’s levels are usually heterogeneous, and the population level frequencies are treated as behavioral parameters and either estimated from the data or calibrated from previous estimates.

Level-*k* models

In a level-*k* model people follow rules of thumb that:

- Anchor their beliefs in a naïve model of others’ response to the game, called *L0*, often uniform random over feasible decisions

and

- Adjust their beliefs via a small number (*k*) of iterated best responses, so *L1* best responds to *L0*, *L2* to *L1*, and so on

People’s levels are usually heterogeneous, and the population level frequencies are treated as behavioral parameters and either estimated from the data or calibrated from previous estimates.

Estimates vary with the setting and population, but normally the estimated frequency of *L0* is small or zero and the distribution of levels is concentrated on *L1*, *L2*, and *L3*.

- *Lk* (for *k* > 0) is decision-theoretically rational, with an accurate model of the game; it departs from equilibrium only in deriving its beliefs from an oversimplified model of others' responses

- *Lk* (for $k > 0$) is decision-theoretically rational, with an accurate model of the game; it departs from equilibrium only in deriving its beliefs from an oversimplified model of others' responses

- *Lk* (for $k > 0$) respects *k*-rationalizability (Bernheim 1984 *Ecma*), hence in two-person games its decisions survive *k* rounds of iterated elimination of strictly dominated strategies

  Thus *Lk* mimics equilibrium decisions in *k*-dominance-solvable games, but may deviate systematically in more complex games

- *Lk* (for $k > 0$) is decision-theoretically rational, with an accurate model of the game; it departs from equilibrium only in deriving its beliefs from an oversimplified model of others' responses

- *Lk* (for $k > 0$) respects *k*-rationalizability (Bernheim 1984 *Ecma*), hence in two-person games its decisions survive *k* rounds of iterated elimination of strictly dominated strategies

  Thus *Lk* mimics equilibrium decisions in *k*-dominance-solvable games, but may deviate systematically in more complex games.

- A level-*k* model (with zero weight on *L0*) can be viewed as a heterogeneity-tolerant refinement of *k*-rationalizability

- *Lk* (for $k > 0$) is decision-theoretically rational, with an accurate model of the game; it departs from equilibrium only in deriving its beliefs from an oversimplified model of others' responses

- *Lk* (for $k > 0$) respects *k*-rationalizability (Bernheim 1984 *Ecma*), hence in two-person games its decisions survive *k* rounds of iterated elimination of strictly dominated strategies

  Thus *Lk* mimics equilibrium decisions in *k*-dominance-solvable games, but may deviate systematically in more complex games.

- A level-*k* model (with zero weight on *L0*) can be viewed as a heterogeneity-tolerant refinement of *k*-rationalizability

  But unlike *k*-rationalizability, a level-*k* model makes precise predictions, given the population level frequencies: not only that deviations from equilibrium will sometimes occur, but also which settings evoke them and which forms they are likely to take.

- Level-*k* models share the generality and much of the tractability of equilibrium models (contrast *k*-rationalizability's set-valued predictions or quantal response equilibrium's computationally challenging noisy fixed-point predictions)

This allows them to clarify the role of the equilibrium assumption:

- Level-*k* models share the generality and much of the tractability of equilibrium models (contrast *k*-rationalizability's set-valued predictions or quantal response equilibrium's computationally challenging noisy fixed-point predictions)

This allows them to clarify the role of the equilibrium assumption:

- A level-*k* analysis can identify settings where conclusions based on equilibrium are robust to likely deviations from equilibrium

- Level-*k* models share the generality and much of the tractability of equilibrium models (contrast *k*-rationalizability's set-valued predictions or quantal response equilibrium's computationally challenging noisy fixed-point predictions)

This allows them to clarify the role of the equilibrium assumption:

- A level-*k* analysis can identify settings where conclusions based on equilibrium are robust to likely deviations from equilibrium

- A level-*k* analysis can identify settings in which mechanisms that yield superior outcomes in equilibrium are worse in practice than others whose performance is less sensitive to deviations: an evidence-disciplined approach to robustness

- Level-*k* models share the generality and much of the tractability of equilibrium models (contrast *k*-rationalizability's set-valued predictions or quantal response equilibrium's computationally challenging noisy fixed-point predictions)

This allows them to clarify the role of the equilibrium assumption:

- A level-*k* analysis can identify settings where conclusions based on equilibrium are robust to likely deviations from equilibrium

- A level-*k* analysis can identify settings in which mechanisms that yield superior outcomes in equilibrium are worse in practice than others whose performance is less sensitive to deviations: an evidence-disciplined approach to robustness

- A level-*k* analysis might reduce optimal mechanisms' sensitivity to distributional and other details that real mechanisms seldom depend on, as advocated by Robert Wilson (1987) and others

Antecedents

- Crawford and Iriberri's (2007 *Ecma*) level-*k* analysis of bidding behavior in sealed-bid independent-private-value and common-value auctions, which builds on Milgrom and Weber's (1982 *Ecma*) equilibrium analysis
- Crawford, Kugler, Neeman, and Pauzner's (2009 *JEEA*;"CKNP") level-*k* analysis of optimal independent-private-value auctions, which builds on Myerson's (1981 *MathOR*) equilibrium analysis
- Saran's (2011 *GEB*) analysis of MS's design problem with a known population frequency of truthful traders
- de Clippel, Saran, and Serrano's (2015, 2018 *REStud*) analyses of implementation with bounded depth of reasoning, small errors
- Kneeland's (2017) analysis of level-*k* implementation, with illustrations including bilateral trading
- Gorelkina's (2018 *IJGT*) level-*k* analysis of the expected externality mechanism

Outline

- CS's equilibrium analysis of bilateral trading via double auction
- MS's analysis of equilibrium-incentive-efficient mechanisms
- A level-*k* model for direct games with asymmetric information
- Level-*k* analyses of the double auction
  - *L1*s' optimism, aggressiveness, and incentive-*in*effiency
  - *L2*s' pessimism, *un*agressiveness, incentive-*super*effiency
- Design requiring level-*k*-incentive-compatibility (not w.l.o.g.)
  - Observable or predictable levels
  - Level-*k* menu effects and the revelation principle
  - Unobservable and unpredictable, heterogeneous levels
- Design relaxing level-*k*-incentive-compatibility
  - Observable or predictable levels
  - Unobservable and unpredictable, heterogeneous levels

CS's equilibrium analysis of bilateral trading via double auction

CS's model has a potential seller and buyer of an indivisible object, in exchange for money.

Their von Neumann-Morgenstern utility functions are quasilinear in money: thus risk-neutral, with money values for the object.

CS's equilibrium analysis of bilateral trading via double auction

CS's model has a potential seller and buyer of an indivisible object, in exchange for money.

Their von Neumann-Morgenstern utility functions are quasilinear in money: thus risk-neutral, with money values for the object.

Denote the buyer's value V and the seller's value C (for "cost").

V and C are independent, with positive densities f(V) and g(C) on their supports and distribution functions F(V) and G(C).

CS and MS allowed the densities to have any bounded overlapping supports, but without important loss of generality I take the supports to be identical and normalize them to [0, 1].

In the double auction:

- If the buyer's money bid $b \geq$ the seller's money ask a, the seller exchanges the object for a given weighted average of b and a

- CS allowed any weights between 0 and 1, but I take the weights to be equal, so the buyer acquires the object at price $(a + b)/2$, the seller's utility is $(a + b)/2$, and the buyer's is $V - (a + b)/2$

- If $b < a$, the seller retains the object, no money changes hands, the seller's utility is C, and the buyer's utility is 0

- I ignore the possibility that $a = b$, which will have 0 probability

The double auction has many Bayesian equilibria.

When f(V) and g(C) are uniformly distributed, CS identify a linear equilibrium, which also plays a central role in MS's analysis.

Denote the buyer's bidding strategy b(V) and the seller's asking strategy a(C), with * subscripts for the equilibrium strategies.

In the linear equilibrium, with value densities supported on [0, 1],

$$b_*(V) = 2V/3 + 1/12$$

unless V < ¼, when $b_*(V)$ can be anything that precludes trade;

and

$$a_*(C) = 2C/3 + ¼$$

unless C > ¾, when $a_*(C)$ can be anything that precludes trade.

In the linear equilibrium, with value densities supported on [0, 1],

$$b_*(V) = 2V/3 + 1/12$$

unless V < ¼, when $b_*(V)$ can be anything that precludes trade;

and

$$a_*(C) = 2C/3 + ¼$$

unless C > ¾, when $a_*(C)$ can be anything that precludes trade.

Trade occurs if and only if 2V/3 + 1/12 ≥ 2C/3 + ¼, or V ≥ C + ¼: with positive probability the outcome is ex post inefficient.

The ex ante probability of trade is 9/32 ≈ 28% and the expected total surplus is 9/64 ≈ 0.14, less than the maximum individually rational probability of trade 50% and expected surplus 1/6 ≈ 0.17.

MS's analysis of equilibrium-incentive-efficient mechanisms

MS characterized incentive-efficient mechanisms in CS's trading environment, requiring interim individual rationality.

Like CS, MS allowed general, independent value distributions with strictly positive densities on ranges that overlap for the buyer and seller; but I will continue to take both value supports to be [0, 1].

MS assumed that traders will play any desired Bayesian equilibrium in the game created by the chosen mechanism.

A *direct* (or *direct-revelation*) mechanism is one in which players' decisions are conformable to estimates of their values, with the outcome a function of the reported values.

When traders are risk-neutral in money, denoting their value reports v and c (distinct from true values V and C), the payoff-relevant aspects of an outcome are determined by two functions:

● p(v, c), the probability that the object is transferred, and

● x(v, c), the expected monetary payment from buyer to seller

Although these outcome functions depend only on reported values, traders' utilities are determined by their true values.

A *direct* (or *direct-revelation*) mechanism is one in which players' decisions are conformable to estimates of their values, with the outcome a function of the reported values.

When traders are risk-neutral in money, denoting their value reports v and c (distinct from true values V and C), the payoff-relevant aspects of an outcome are determined by two functions:

- p(v, c), the probability that the object is transferred, and
- x(v, c), the expected monetary payment from buyer to seller

Although these outcome functions depend only on reported values, traders' utilities are determined by their true values.

A direct mechanism with outcome functions p(·, ·), x(·, ·) is *incentive-compatible* iff it makes truthful reporting an equilibrium; and is *(interim) individually rational* iff it yields buyer and seller expected utility ≥ 0 for every possible realization of their values.

The revelation principle shows that if traders can be counted on to play any desired equilibrium in the game created by the designer's chosen mechanism, there is no loss of generality in restricting attention to incentive-compatible direct mechanisms:

"We can, without any loss of generality, restrict our attention to incentive-compatible direct mechanisms. This is because, for any Bayesian equilibrium of any bargaining game, there is an equivalent incentive-compatible direct mechanism that always yields the same outcomes (when the individuals play the honest equilibrium)….[w]e can construct [such a] mechanism by first asking the buyer and seller each to confidentially report his valuation, then computing what each would have done in the given equilibrium strategies with these valuations, and then implementing the outcome (transfer of money and object) as in the given game for this computed behavior. If either individual had any incentive to lie to us in this direct mechanism, then he would have had an incentive to lie to himself in the original game, which is a contradiction of the premise that he was in equilibrium in the original game." (MS, pp. 267-268)

MS's Theorem 1 uses the conditions for incentive-compatibility and individual rationality to derive an "incentive budget constraint" (my term, not theirs), subject to which, for traders with quasilinear utility functions, incentive-efficient outcome functions p(·, ·) and x(·, ·) must maximize the sum of traders' ex ante expected utilities.

MS's Corollary 1 shows that no incentive-compatible, individually rational mechanism can assure ex post Pareto-efficiency.

MS's Theorem 2 uses Theorem 1's conditions to characterize the outcome functions associated with incentive-efficient mechanisms.

(The level-*k* counterparts of MS's results will be used below to give a more detailed exposition of MS's analysis.)

In CS's example with uniform value densities, MS's Theorem 2 yields a closed-form solution for an incentive-compatible, incentive-efficient mechanism, which transfers the object if and only if the reported values satisfy $v \geq c + \frac{1}{4}$, at price $(v + c + \frac{1}{2})/3$.

The linear equilibrium of the double auction with uniform densities transfers the object whenever the *true* values satisfy $V \geq C + \frac{1}{4}$.

Thus, even though the double auction is not incentive-compatible, linear equilibrium bidding strategies shade to mimic the outcome of truthful reporting in MS's incentive-efficient mechanism.

(Satterthwaite and Williams 1989 *JET* showed, however, that for generic densities CS's double auction does *not* yield incentive-efficient outcomes; so MS's result for this example is coincidental.)

A level-*k* model for direct games with asymmetric information

Recall that in a level-*k* model people anchor their beliefs in a naïve model of others' responses, *L0*, and adjust their beliefs via iterated best responses: *L1* best responds to *L0*, and so on.

In complete-information games *L0* is usually assumed to make decisions uniformly distributed over the feasible decisions.

A level-*k* model for direct games with asymmetric information

Recall that in a level-*k* model people anchor their beliefs in a naïve model of others' responses, *L0*, and adjust their beliefs via iterated best responses: *L1* best responds to *L0*, and so on.

In complete-information games *L0* is usually assumed to make decisions uniformly distributed over the feasible decisions.

I take *L0*'s decisions to be uniform over the feasible decisions, *independent of own value*.

Specifically, in a direct mechanism for the bilateral trading setting, I take *L0*'s decisions to be uniform over [0, 1].

(Allowing bounded overlapping supports as CS and MS did, my assumption corresponds to assuming that *L0* is uniform on the overlap, which traders have enough information to identify.)

This *L0* yields a hierarchy of rules via iterated best responses.

One can imagine more refined specifications, e.g. with an *L0* buyer's bid (seller's ask) uniform below (above) its value instead of over the entire range (thus eliminating dominated strategies).

But *L0* is not an actual player: It is a player's naïve model of other players—others whose values he does not observe.

It is *logically* possible that players reason contingent on others' possible values, but behaviorally far-fetched: people tend to be informationally naïve ("cursed", Eyster and Rabin 2005 *Ecma*), ignoring links between others' decisions and private information.

The extended level-*k* model captures both this informational naivete and people's aversion to the fixed-point or indefinitely iterated dominance reasoning that equilibrium often requires.

This extended level-*k* model has a long history:

Milgrom and Stokey's (1982 *JET*) "No-Trade Theorem" shows that if traders in an asset market start out in market equilibrium—Pareto-efficient, given their information—then giving them new information, fundamentals unchanged, cannot lead to new trades.

For, any such trades would make it common knowledge that both had benefited, contradicting the efficiency of the initial equilibrium.

This extended level-*k* model has a long history:

Milgrom and Stokey's (1982 *JET*) "No-Trade Theorem" shows that if traders in an asset market start out in market equilibrium—Pareto-efficient, given their information—then giving them new information, fundamentals unchanged, cannot lead to new trades.

For, any such trades would make it common knowledge that both had benefited, contradicting the efficiency of the initial equilibrium.

This result has been called the Groucho Marx Theorem:

> "I sent the club a wire stating, 'Please accept my resignation. I don't want to belong to any club that will accept people like me as a member'."
> —Groucho Marx, Telegram to the Beverly Hills Friars' Club

This extended level-*k* model has a long history:

Milgrom and Stokey's (1982 *JET*) "No-Trade Theorem" shows that if traders in an asset market start out in market equilibrium—Pareto-efficient, given their information—then giving them new information, fundamentals unchanged, cannot lead to new trades.

For, any such trades would make it common knowledge that both had benefited, contradicting the efficiency of the initial equilibrium.

This result has been called the Groucho Marx Theorem:

> "I sent the club a wire stating, 'Please accept my resignation. I don't want to belong to any club that will accept people like me as a member'."
>
> —Groucho Marx, Telegram to the Beverly Hills Friars' Club

In speculating on why zero-sum trades occur despite the theorem, Milgrom and Stokey contrast Groucho's equilibrium-like inference with their rules Naïve Behavior, which keeps its prior but behaves rationally otherwise, as *L1* does; and First-Order Sophistication, which best responds to Naïve Behavior, as *L2* does.

Although Groucho was far from naïve, the informal literature also contains evidence of informational naiveté:

> "Son…One of these days in your travels, a guy is going to show you a brand-new deck of cards on which the seal is not yet broken. Then this guy is going to offer to bet you that he can make the jack of spades jump out of this brand-new deck of cards and squirt cider in your ear. But, son, do not accept this bet, because as sure as you stand there, you're going to wind up with an ear full of cider."
>
> —Obadiah ("The Sky") Masterson, quoting his father in Damon Runyon's *Guys and Dolls: The Stories of Damon Runyon*, 1932, "The Idyll of Miss Sarah Brown"

Although Groucho was far from naïve, the informal literature also contains evidence of informational naiveté:

> “Son…One of these days in your travels, a guy is going to show you a brand-new deck of cards on which the seal is not yet broken. Then this guy is going to offer to bet you that he can make the jack of spades jump out of this brand-new deck of cards and squirt cider in your ear. But, son, do not accept this bet, because as sure as you stand there, you're going to wind up with an ear full of cider.”
>
> —Obadiah (“The Sky”) Masterson, quoting his father in Damon Runyon’s *Guys and Dolls: The Stories of Damon Runyon*, 1932, “The Idyll of Miss Sarah Brown”

Here, Sky’s Dad is worried that Sky is *L1*: rational but for sticking with his prior in the face of an offer that is too good to be true.

More recent work shows that the extended level-*k* model often gives a realistic account of people's nonequilibrium strategic thinking and informational naiveté:

- Camerer et al. (2004 *QJE*) used a cognitive hierarchy analogue of this level-*k* model to explain zero-sum betting

- Crawford and Iriberri (2007 *Ecma*) used this level-*k* model to describes subjects' overbidding and vulnerability to the winner's curse in initial responses in classic auction experiments

- Brown, Camerer, and Lovallo (2012 *AEJ Micro*) used this level-*k* model to explain film-goers' failure to draw negative inferences from studios' withholding weak movies from critics, pre-release

- Brocas, Carillo, Camerer, and Wang (2014 *REStud*) reported very powerful experimental evidence for this level-*k* model from approximately zero-sum betting games

Level-*k* analysis of the double auction

I first apply the level-*k* model to CS's trading environment, focusing on their leading example with uniform value densities.

I assume that a trader's level is independent of its value.

I set *L0*'s frequency to zero, and focus on homogeneous populations of *L1*s or *L2*s, which allows simple illustrations of the main points.

Denote the buyer's bidding strategy $b_i(V)$ and the seller's asking strategy $a_i(C)$, where the subscripts denote levels i = 1,2.

*L1*s' optimism, aggressiveness, and incentive-*in*efficiency

An *L1* buyer believes that the seller's ask is uniformly distributed on [0, 1], independent of its value.

Optimization then yields $b_1(V) = 2V/3$ (in the interior).

Similarly, an *L1* seller's ask $a_1(C) = 2C/3 + 1/3$ (in the interior).

(Compare Crawford and Iriberri's 2007 *ECMA* analysis of *L1* bidding in first-price auctions.)

***L1*s' optimism, aggressiveness, and incentive-*in*efficiency**

An *L1* buyer believes that the seller's ask is uniformly distributed on [0, 1], independent of its value.

Optimization then yields $b_1(V) = 2V/3$ (in the interior).

Similarly, an *L1* seller's ask $a_1(C) = 2C/3 + 1/3$ (in the interior).

(Compare Crawford and Iriberri's 2007 *ECMA* analysis of *L1* bidding in first-price auctions.)

Despite the multiplicity of equilibria in the double auction, a level-*k* model makes generically unique predictions, conditional on the population level frequencies.

L1's optimal strategy is independent of the value densities; unlike *L2*'s, which depends on the other trader's density, or an equilibrium strategy, which depends on both traders' densities.

With $b_1(V) = 2V/3$, an *L1* buyer bids 1/12 more aggressively (that is, bids less) than an equilibrium buyer with $b_*(V) = 2V/3 + 1/12$.

With $a_1(C) = 2C/3 + 1/3$, an *L1* seller asks 1/12 more aggressively (more) than an equilibrium seller with $a_*(C) = 2C/3 + ¼$.

Both have the same 2/3 shading as equilibrium bids or asks.

With $b_1(V) = 2V/3$, an *L1* buyer bids 1/12 more aggressively (that is, bids less) than an equilibrium buyer with $b_*(V) = 2V/3 + 1/12$.

With $a_1(C) = 2C/3 + 1/3$, an *L1* seller asks 1/12 more aggressively (more) than an equilibrium seller with $a_*(C) = 2C/3 + 1/4$.

Both have the same 2/3 shading as equilibrium bids or asks.

For an *L1* buyer and seller, trade takes place iff $V \geq C + 1/2$; ex post efficiency is lost for more values than in equilibrium with $V \geq C + 1/4$; and the ex ante probability of trade is 1/8 = 12.5%, versus the equilibrium 9/32 ≈ 28%.

- Is there a mechanism that enhances efficiency for *L1*s by counteracting their aggressiveness in the double auction?

  I will show that, whether or not *L1*-incentive-compatibility is required, the answer is Yes.

***L2*s' pessimism, *un*aggressiveness, and incentive-*super*efficiency**

An *L2* buyer's bid $b_2(V)$ maximizes over b ϵ [0, 1]

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

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

For instance, if g(C) is uniform, an *L2* buyer believes that the seller's ask $a_1(C) = 2C/3 + 1/3$ is uniformly distributed on [1/3, 1], with density 3/2 there and 0 elsewhere.

An *L2* buyer who believes the seller's ask is distributed on [1/3, 1] believes that trade requires b > 1/3.

For V ≤ 1/3 it is then optimal for an *L2* buyer to bid anything it thinks yields 0 probability of trade: In the absence of dominance among such strategies, I set $b_2(V) = V$ for V in [0, 1/3].

For V > 1/3, an *L2* buyer's bid $b_2(V)$ maximizes over b ϵ [1/3, 1]

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

The second-order condition is always satisfied.

Solving the first-order condition (3/2)(V - b) - (3/4)(V - 1/3) = 0 yields $b_2(V) = 2V/3 + 1/9$ for V ϵ [1/3, 1].

Similarly, an *L2* seller's ask $a_2(C) = 2C/3 + 2/9$ (in the interior).

With $b_2(V) = 2V/3 + 1/9$, an *L2* buyer bids 1/36 less aggressively (more) than an equilibrium buyer with $b_*(V) = 2V/3 + 1/12$ and 1/9 less aggressively (more) than an *L1* buyer with $b_1(V) = 2V/3$.

With $a_2(C) = 2C/3 + 2/9$, an *L2* seller asks 1/36 less aggressively (less) than an equilibrium seller, and 1/9 less than an *L1* seller.

Both again have the same 2/3 shading as equilibrium or *L1*.

With $b_2(V) = 2V/3 + 1/9$, an *L2* buyer bids 1/36 less aggressively (more) than an equilibrium buyer with $b_*(V) = 2V/3 + 1/12$ and 1/9 less aggressively (more) than an *L1* buyer with $b_1(V) = 2V/3$.

With $a_2(C) = 2C/3 + 2/9$, an *L2* seller asks 1/36 less aggressively (less) than an equilibrium seller, and 1/9 less than an *L1* seller.

Both again have the same 2/3 shading as equilibrium or *L1*.

For an *L2* buyer and *L2* seller, trade takes place iff $V \geq C + 1/6$; ex post efficiency is lost for fewer values than in equilibrium with $V \geq C + ¼$ or with *L1*s with $V \geq C + ½$; and the ex ante probability of trade is $25/72 \approx 35\%$, versus the equilibrium 28% or *L1* 12.5%.

With $b_2(V) = 2V/3 + 1/9$, an *L2* buyer bids 1/36 less aggressively (more) than an equilibrium buyer with $b_*(V) = 2V/3 + 1/12$ and 1/9 less aggressively (more) than an *L1* buyer with $b_1(V) = 2V/3$.

With $a_2(C) = 2C/3 + 2/9$, an *L2* seller asks 1/36 less aggressively (less) than an equilibrium seller, and 1/9 less than an *L1* seller.

Both again have the same 2/3 shading as equilibrium or *L1*.

For an *L2* buyer and *L2* seller, trade takes place iff $V \geq C + 1/6$; ex post efficiency is lost for fewer values than in equilibrium with $V \geq C + ¼$ or with *L1*s with $V \geq C + ½$; and the ex ante probability of trade is 25/72 ≈ 35%, versus the equilibrium 28% or *L1* 12.5%.

- Is there a mechanism that does as well or better for *L2*s than the double auction by further exploiting their *un*agressiveness?

  I will show that, whether or not *L2*-incentive-compatibility is required, the answer is No, at least for uniform value densities.

Design for level-*k* traders

Throughout the analysis of level-*k* design, I restrict attention to direct mechanisms, whose decisions can be viewed as estimates of own values; and I ignore the noisiness of people's decisions.

I define incentive-efficiency notions for a designer's correct beliefs; but I derive incentive constraints from traders' level-*k* beliefs.

I use "incentive-compatible" here in the narrow sense, for direct mechanisms in which it is optimal for traders to report truthfully.

But when I relax it traders are still assumed to best respond, even if such responses need not be truthful, as in a first-price auction.

"Level-*k*-incentive-compatibility" and "level-*k*-interim-individual-rationality" parallel the standard notions, "equilibrium-incentive-compatibility" and "equilibrium-interim-individual-rationality".

Recall that MS's Theorem 1 uses conditions for equilibrium-incentive-compatibility to derive an incentive budget constraint.

MS's Corollary 1 then uses that constraint to show that no equilibrium-incentive-compatible, individually rational mechanism can assure ex post Pareto-efficiency with probability one.

MS's Theorem 2 characterizes (for traders with quasilinear utility functions) equilibrium-incentive-efficient mechanisms by deriving conditions for maximizing the sum of traders' ex ante expected utilities subject to Theorem 1's incentive budget constraint.

In CS's example with uniform value densities, Theorem 2 yields a closed-form solution for the incentive-compatible form of an equilibrium-incentive-efficient mechanism; it transfers the object iff the reported values satisfy $v \geq c + \frac{1}{4}$, at price $(v + c + \frac{1}{2})/3$.

Design requiring level-*k*-incentive-compatibility (not w.l.o.g.)

When *Lk*-incentive-compatibility is required, MS's characterization of the incentive-efficient mechanism with uniform value densities is completely robust to level-*k* thinking:

*Theorem 1. With uniform value densities, the set of level-*k-*incentive-compatible mechanisms coincides with the set of equilibrium-incentive-compatible mechanisms for any population of level-*k *traders with* k *> 0. Thus MS's equilibrium-incentive-efficient mechanism is also efficient in the set of* Lk*-incentive-compatible mechanisms, whether or not individual traders' levels are observable or predictable.*

Proof. *L1* traders believe they face a uniform distribution of other's reports, so their incentive-compatibility and individual rationality constraints coincide with those of equilibrium traders. Induction extends this to higher levels, and MS's equilibrium-incentive-efficient mechanism is also efficient in the set of *Lk*-incentive-compatible mechanisms for any population of level-*k* traders.

Level-*k* menu effects and the revelation principle

Comparing Theorem 1 with level-*k* analyses of the double auction reveals a limitation of the revelation principle for level-*k* design.

With uniform value densities, MS's equilibrium-incentive-efficient mechanism is outcome-equivalent to CS's linear double-auction equilibrium, in which traders shade their bids and asks to mimic truthful reporting in MS's mechanism as in the revelation principle.

Level-*k* menu effects and the revelation principle

Comparing Theorem 1 with level-*k* analyses of the double auction reveals a limitation of the revelation principle for level-*k* design.

With uniform value densities, MS's equilibrium-incentive-efficient mechanism is outcome-equivalent to CS's linear double-auction equilibrium, in which traders shade their bids and asks to mimic truthful reporting in MS's mechanism as in the revelation principle.

But in my examples with uniform value densities, *L1*s do worse in the double auction than in MS's mechanism, while *L2*s do better:

- MS's mechanism neutralizes *L1*s' aggressiveness in the double auction by rectifying their beliefs
- MS's mechanism could also neutralize *L2*s' *unagressiveness*, but the non-incentive-compatible double auction uses their non-equilibrium beliefs to do even better

MS's equilibrium-incentive-efficient mechanism in this case is efficient for level-k traders in the set of Lk-incentive-compatible mechanisms only if implemented in its incentive-compatible form.

MS's equilibrium-incentive-efficient mechanism in this case is efficient for level-*k* traders in the set of *Lk*-incentive-compatible mechanisms only if implemented in its incentive-compatible form.

The revelation principle does not work as expected here because the choice between mechanisms that are outcome-equivalent in equilibrium influences the correctness of level-*k* beliefs by changing the mechanism's relationship to *L0*, via CKNP's "level-*k* menu effects".

MS's equilibrium-incentive-efficient mechanism in this case is efficient for level-*k* traders in the set of *Lk*-incentive-compatible mechanisms only if implemented in its incentive-compatible form.

The revelation principle does not work as expected here because the choice between mechanisms that are outcome-equivalent in equilibrium influences the correctness of level-*k* beliefs by changing the mechanism's relationship to *L0*, via CKNP's "level-*k* menu effects".

As a result, the revelation principle may fail for level-*k* players, and requiring level-*k*-incentive-compatibility is not without loss of generality.

Some analysts of design have argued that incentive-compatibility is essential in applications, e.g. for school choice or combinatorial auctions, but mostly in equilibrium analyses where it is w.l.o.g.

Others are willing to consider non-incentive-compatible mechanisms like the Boston Mechanism or first-price auctions.

I don’t try to resolve this issue here; instead I consider both cases.

Some analysts of design have argued that incentive-compatibility is essential in applications, e.g. for school choice or combinatorial auctions, but mostly in equilibrium analyses where it is w.l.o.g.

Others are willing to consider non-incentive-compatible mechanisms like the Boston Mechanism or first-price auctions.

I don't try to resolve this issue here; instead I consider both cases.

My strongest results assume that *Lk*-incentive-compatibility is required, considering separately the cases where traders' levels are observable or predictable and where they are not.

I then consider allowing any direct mechanism but not requiring *Lk*-incentive-compatibility, still assuming that traders best respond.

Design requiring level-*k*-incentive-compatibility, with observable or predictable levels

With general value densities, the payoff-relevant aspects of a direct mechanism are still outcome functions $p(\cdot, \cdot)$ and $x(\cdot, \cdot)$, where buyer and seller report values *v* and *c*, and $p(v, c)$ is the probability the object transfers, for expected payment $x(v, c)$.

For a mechanism (p, x), $f^k(v; p, x)$ and $F^k(v; p, x)$ are the density and distribution function of an *Lk* seller's beliefs and $g^k(c; p, x)$ and $G^k(c; p, x)$ are those of an *Lk* buyer's beliefs.

With *L0* uniform 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 $g^2(c; p, x) \equiv g(\alpha_1^{-1}(c; p, x))$.

For ease of notation, assume each trader population has only one level (the extension to multiple observable levels is immediate).

As in MS's analysis, the buyer's and seller's expected monetary payments, probabilities of trade, and utilities can be written as functions of their value reports, *v* and *c*.

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

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

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

For a given k, the mechanism $p(\cdot, \cdot)$, $x(\cdot, \cdot)$ is Lk-incentive-compatible iff truthful reporting is optimal given Lk beliefs.

That is, if for every V, v, C, and c in [0, 1],

(5.2) $$U_B^k(V,V) \geq U_B^k(V,v) = VP_B^k(v) - X_B^k(v) \text{ and}$$

$$U_S^k(C,C) \geq U_S^k(C,c) = X_S^k(c) - CP_S^k(c).$$

For a given *k*, the mechanism *p*(·, ·), *x*(·, ·) is *Lk*-incentive-compatible iff truthful reporting is optimal given *Lk* beliefs.

That is, if for every *V*, *v*, *C*, and *c* in [0, 1],

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

The mechanism *p*(·, ·), *x*(·, ·) is (interim) *Lk*-individually rational iff for every *V* and *C* in [0, 1],

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

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

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

$$(5.4)\ U_B^i(0,0) + U_S^j(1,1) = \min_{V\in[0,1]} U_B^i(V,V) + \min_{C\in[0,1]} U_S^j(C,C)$$

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

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

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

Proof. The proof follows MS's, adjusted for traders' level-*k* beliefs.

By (5.1), $P_B^i(\cdot)$ is weakly increasing and $P_S^j(\cdot)$ is weakly decreasing for any given (*p*, *x*), which yields necessary and sufficient conditions for incentive-compatibility:

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

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

(5.4) follows because the designer's anticipated expected surplus, with correct beliefs, must suffice to incentivize traders with level-*k* beliefs, but with the cost of doing so evaluated for correct beliefs:

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

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

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

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

where the second-to-last equality follows via integration by parts.

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

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

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

Before stating Theorem 3, consider whether MS's Corollary 1 generalizes when traders are level-*k* with observable levels.

MS's Corollary 1 shows that if traders' values have positive probability densities over [0,1], then no equilibrium-incentive-compatible and equilibrium-interim individually rational trading mechanism can assure ex post efficiency with probability 1.

If the buyer is level *i* and the seller is level *j*, and $p(V,C) \equiv 1$ iff $V \geq C$, the constraint (5.5) reduces to:

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

(5.9) $$= \int_0^1 \int_0^V \left[(V-C)g(C)f(V) - \{1-F(V)\}g^i(C) - G(C)f^j(V) \right] dCdV$$

$$= \int_0^1 \{F(V)-1\}G^i(V)dV + \int_0^1 \{f(V)-f^j(V)\} \int_0^V G(C)dCdV.$$

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

$$(5.9) = \int_0^1 \int_0^V \left[(V-C) g(C) f(V) - \{1-F(V)\} g^i(C) - G(C) f^j(V) \right] dC dV$$

$$= \int_0^1 \{F(V) - 1\} G^i(V) dV + \int_0^1 \{f(V) - f^j(V)\} \int_0^V G(C) dC dV.$$

The first term on the right-hand side of (5.9) is analogous to MS's term, but with the level-*k* beliefs $G^i(V)$; it is again always negative.

The second term on the right-hand side of (5.9) vanishes for correct beliefs and so has no counterpart in MS's analysis; it is positive for some value distributions and negative for others.

To see that the second term can outweigh the first, consider *L1* beliefs: $f^j(V) \equiv 1$ and $g^i(C) \equiv 1$. Then, e.g., $F(\cdot)$ and $G(\cdot)$ with full supports, but with $F(\cdot)$ approximately uniform on [*b*, 1] and $G(\cdot)$ with an approximate spike at *b*, make the right-hand side of (5.9) positive, a contradiction that shows that MS's Corollary 1 does not generalize.

That said, the tendency MS identified for the incentive constraints to hold the probability of ex post efficient trading below 1 persists.

Theorem 3 characterizes mechanisms that are efficient in the set of level-*k*-incentive-compatible mechanisms when buyer's and seller's levels are observable or predictable.

Define, for $\beta \geq 0$,

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

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

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

If feasible, $p_0^{ij}(V,C)$ would yield an ex post efficient allocation; but it may violate the incentive budget constraint (5.5). By contrast, $p_1^{ij}(V,C)$ maximizes the slack in (5.5); but it wastes surplus.

The goal is an optimal compromise between these two extremes.

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

Proof. The proof adapts MS's proof. Fix buyer's and seller's levels *i* and *j*, and consider choosing *p*(·, ·) to maximize traders' ex ante expected total gains from trade subject to 0 ≤ *p*(·, ·) ≤ 1, $U_B^i(0,0) = U_S^j(1,1) = 0$, and (5.5). (5.5) and (5.10) yield:

(5.11) $\max_{\{0 \leq p(\cdot,\,\cdot) \leq 1\}} \int_0^1 \int_0^1 (V - C)\, p(V,C) g(C) f(V) dCdV$

s.t. $0 \leq \int_0^1 \int_0^1 \Psi^{ij}(V,C;1) p(V,C) g(C) f(V) dCdV$.

If a solution *p*(·, ·) to (5.11) yields a $\mathrm{P}_{\mathrm{B}}^{\mathrm{i}}(\cdot)$ that is weakly increasing for all *v* and a $\mathrm{P}_{\mathrm{S}}^{\mathrm{j}}(\cdot)$ that is weakly decreasing for all *c*, then by Theorem 2 that solution is associated with a mechanism that maximizes traders' ex ante expected total gains from trade among all level-*k*-incentive-compatible and level-*k*-interim individually rational mechanisms.

(5.11) is like a consumer's budget problem, with trade probabilities $p(V,C)$ like a continuum of goods with "prices" $\left[\frac{\{1-F(V)\}g^i(C)}{f(V)g(C)} + \frac{G(C)f^j(V)}{g(C)f(V)}\right]$.

Because the $p(V,C)$ enter the objective function and the constraint linearly, solutions of problem (5.11) are "bang-bang", with $p(V,C) = 0$ or 1 almost everywhere and $p(V,C) = 1$ for the (V,C) pairs with the largest expected gain per unit of incentive cost (analogous to the highest marginal-utility-to-price ratios).

Form the Lagrangean:

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

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

The Kuhn-Tucker conditions are:

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

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

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

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

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

Given the continuity and monotonicity of $\Psi^{ij}(V,C;\beta)$, there is a unique λ and $p = p_{\beta}^{ij}(V,C)$, with $\beta = \frac{\lambda}{1+\lambda}$ (equivalently, $\lambda = \frac{\beta}{1-\beta}$), that satisfy $U_B^i(0,0) = U_S^j(1,1) = 0$, (5.5), (5.13), and (5.14). Q.E.D.

Theorem 3's condition that $\Psi^{ij}(V, C; 1)$ is increasing in *V* and decreasing in *C* for all (*p, x*) is the level-*k* analogue of MS's Theorem 2 monotonicity conditions, which are satisfied whenever the true densities fit Myerson's (1981) "regular case", ruling out strong hazard rate variations in the wrong direction.

If traders' beliefs $f^j(V; p, x)$ and $g^i(C; p, x)$ were equal to the true densities $f(V)$ and $g(C)$, Theorem 3's monotonicity condition reduces to MS's Theorem 2 condition.

Otherwise, Theorem 3's condition restricts traders' beliefs and the true densities in a qualitatively similar way.

The proofs of Theorems 2 and 3 show that analogous results would go through for any behavioral model that makes unique predictions that are best responses to beliefs and satisfies analogous monotonicity restrictions.

Comparing level-*k* and equilibrium incentive budget constraints ((5.5) with (2.5) (MS's (2)), and the level-*k* and equilibrium Kuhn-Tucker conditions ((5.14) with (2.8) (MS's p. 274) shows that the design features that foster equilibrium-incentive-efficiency foster efficiency in the set of level-*k*-incentive-compatible mechanisms, with different weights due to incentive effects of different beliefs.

(5.14)'s condition for $p(V, C) = 1$ shows that mechanisms that are efficient in the set of level-*k*-incentive-compatible mechanisms, like MS's equilibrium-incentive-efficient mechanisms, never require commitment to ex post perverse trade for any values.

However, as in MS's analysis, Theorem 2's transfer function (5.8) may sometimes violate ex-post individual rationality by requiring payment from buyers who don't get the object.

The level-*k* Kuhn-Tucker conditions (5.14) also reveal that mechanisms that are efficient in the set of level-*k*-incentive-compatible mechanisms must generally tacitly exploit traders' predictably incorrect beliefs ("TEPIB"):

- "Predictably" via the level-*k* model
- "Exploit" in the benign sense that traders' incorrect beliefs are used only for their benefit
- "Tacitly" in that the mechanism does not actively deceive traders

Relative to an equilibrium-incentive-efficient mechanism, TEPIB favors trade at (*V, C*) combinations for which traders' beliefs make the "prices" $\left[\frac{\{1-F(V)\}g^{i}(C)}{f(V)g(C)}+\frac{G(C)f^{j}(V)}{g(C)f(V)}\right]$ low compared to MS's equilibrium "prices" $\left[\frac{1-F(V)}{f(V)}+\frac{G(C)}{g(C)}\right]$.

For *L2* and higher levels, TEPIB also tends to favor mechanisms that increase the advantages of the first two effects.

More generally, the above analysis suggests that if people's levels of thinking are observable or predictable, viewing "robust mechanism design" as implementing equilibrium-incentive-efficient outcomes under weaker behavioral assumptions is too narrow.

To use an optimal-auctions example (CKNP):

- A second-price auction seems more robust than an equilibrium-revenue-equivalent first-price auction, because it yields the equilibrium level of revenue for any mixture of level-*k* bidders
- But revenue-equivalence breaks down for level-*k* bidders; and design for *L1*s tends to favor first-price auctions, which make *L1*s overbid relative to equilibrium, over second-price auctions, which make *L1* bidders mimic equilibrium bidders (Crawford and Iriberri 2007, CKNP)

By contrast, I will show that if traders' levels are unobservable and unpredictable, a level-*k*-incentive-compatible mechanism cannot screen values and levels at the same time, and must be robust.

I now give some examples of mechanisms that are efficient in the set of level-*k*-incentive-incentive-compatible mechanisms.

As in MS's analysis, I have closed-form solutions only with uniform value densities, for which TEPIB has no influence (Theorem 1).

To illustrate TEPIB, I report computed trading regions for such mechanisms for *L1*s and combinations of linear densities.

(Figure 1 in the paper reports *L1*s' trading regions for a coarse subset of linear density combinations, excluding only extreme combinations that violate Theorems 2-3's monotonicity conditions.

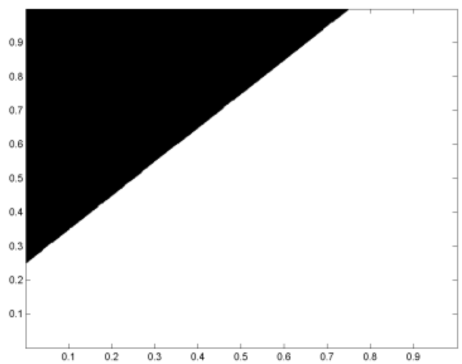
For *L2*s, with $f^2(v) \equiv f(\beta_1^{-1}(v; p, x))$ and $g^2(c) \equiv g(\alpha_1^{-1}(c; p, x))$, (5.5) and (5.14) depend on both the transfer function $x(\cdot, \cdot)$ and $p(\cdot, \cdot)$, making the dimensionality of search too high.)

The examples show that mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms are qualitatively quite similar to equilibrium-incentive-efficient mechanisms.

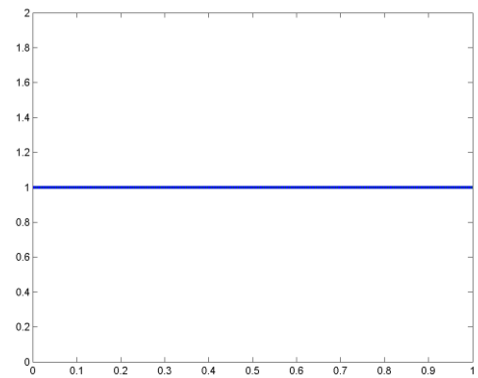
From Figure 1. Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms

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

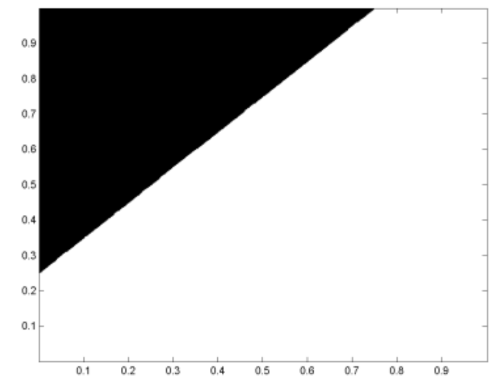
Equilibrium: 1.0, 1.0



Buyer (—) and seller (···)

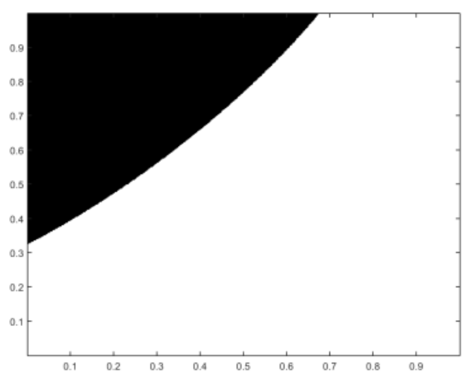


L1: 1.0, 1.0

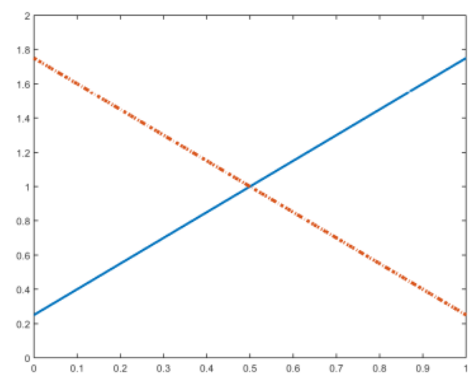


Uniform value densities

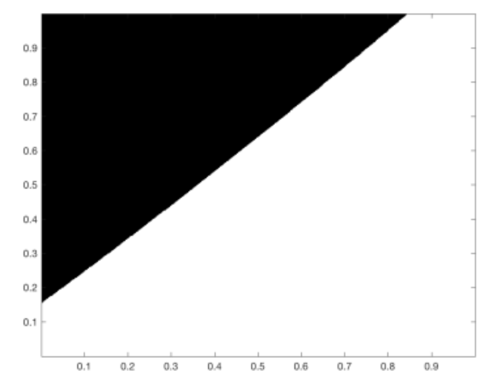
Equilibrium: 0.25, 1.75



Buyer (—) and seller (···)



L1: 0.25, 1.75

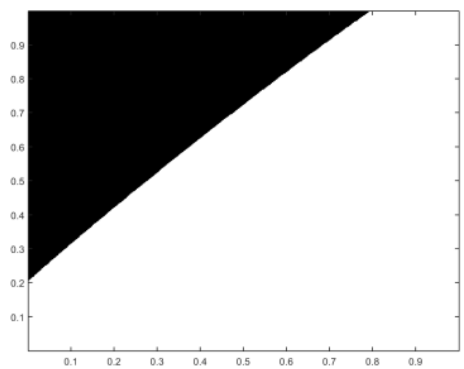


Pessimism makes *L1* trading region larger than for equilibrium

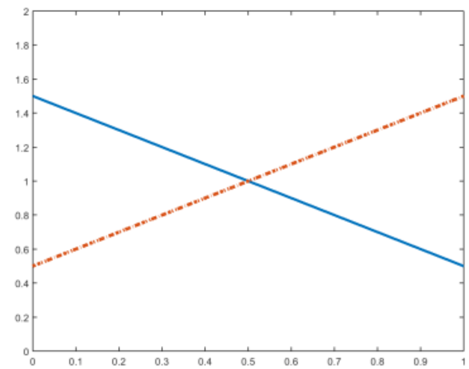
From Figure 1. Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms

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

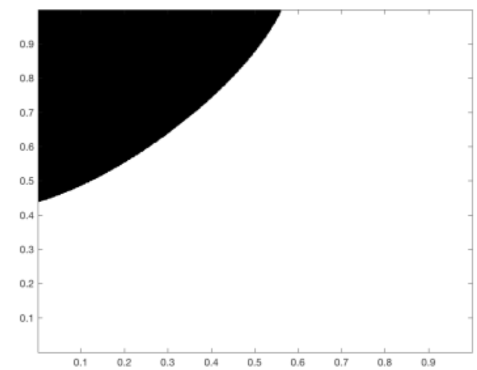
Equilibrium: 1.5, 0.5



Buyer (—) and seller (···)

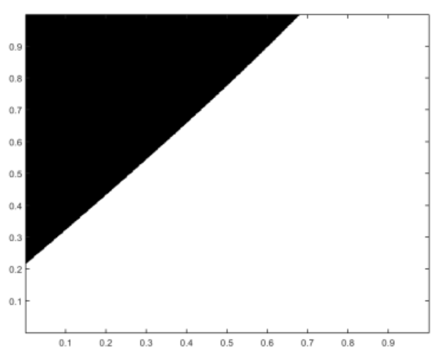


L1: 1.5, 0.5

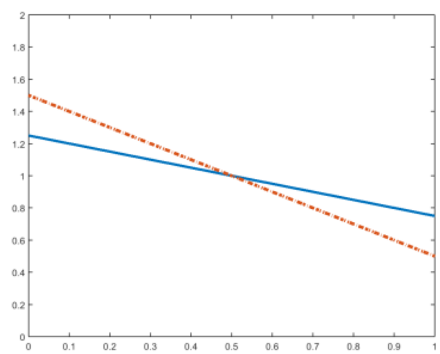


Optimism makes *L1* trading region smaller than for equilibrium

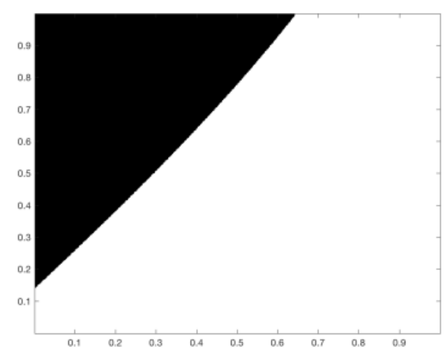
Equilibrium: 1.25, 1.5



Buyer (—) and seller (···)



L1: 1.25, 1.5



Design requiring level-*k*-incentive-compatibility, with unobservable, heterogeneous levels

I now relax the assumption that traders' levels are observable or predictable, continuing to require level-*k*-incentive-compatibility.

Theorem 1 shows that with uniform value densities, even in this case, MS's equilibrium-incentive-efficient direct mechanism is efficient in the set of level-*k*-incentive-compatible mechanisms.

But with general value densities, unobservable and unpredictable levels pose significant new difficulties for design.

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

A “deterministic posted-price mechanism” is a random posted-price mechanism for which the distribution over prices is a spike.

Theorem 4. *Assume that traders' levels are unobservable and unpredictable, but the designer knows that the population level distributions for buyers and sellers include* L1 *and at least one higher level with positive probabilities. Then, if* $f(\cdot)$, $g(\cdot) \neq 1$ *almost everywhere, level-*k*-incentive-compatibility requires that the mechanism is equivalent to a random posted-price mechanism.*

*Further, a mechanism that maximizes traders' true expected total gains from trade among level-*k*-incentive-compatible and interim individually rational mechanisms is equivalent to a deterministic posted-price mechanism with* $U_B^i(0,0) = U_S^j(1,1) = 0$ *for all levels i in the buyer population and j in the seller population.*

Finally, the optimal posted price, π, is unique, characterized by the first-order condition:

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

Proof. By Theorem 2, (5.6) must hold for any levels in the buyer and seller populations. If a mechanism is *L1*-incentive-compatible, the proof of Theorem 1 shows that it is *Lk*-incentive-compatible for all *k* > 1 if and only if it is also equilibrium-incentive-compatible:

(5.6) $$U_B^i(V,V) = U_B^i(0,0) + \int_0^V P_B^i(\hat{v})d\hat{v} \text{ for all } V \text{ and}$$

$$U_S^j(C,C) = U_S^j(1,1) + \int_C^1 P_S^j(\hat{c})d\hat{c} \;\; for\ all\ C$$

Standard arguments show that this is a contradiction unless the mechanism is equivalent to a random posted-price mechanism. A random posted-price mechanism is level-*k*-incentive-compatible for any *k*. By linearity, there are always "bang-bang" $p(v,c)$ solutions with $p(V,C) = 0$ or 1 almost everywhere, so $\mu(\cdot) \equiv 1$ w.l.o.g. An optimal deterministic posted-price mechanism solves:

(5.25) $$\max_{\{0 \le \pi \le 1\}} \int_\pi^1 \int_0^\pi (V - C)\, p(V,C) g(C) f(V) dCdV.$$

The second-order condition is satisfied, so the unique optimal posted price satisfies the first-order condition (5.15). Q.E.D.

Theorem 4 shows that if traders' levels are unobservable and unpredictable, a mechanism that is efficient in the set of level-*k*-incentive-compatible mechanisms must be a deterministic posted-price mechanism with a particular price, which mechanism makes truthful revelation of values a dominant strategy for all levels.

Thus, in that case, a mechanism's inability to screen levels and values simultaneously rules out the sensitivity to reported values of mechanisms that are equilibrium-incentive-efficient or efficient in the set of level-*k*-incentive-compatible mechanisms.

This result would extend to any well-behaved nonequilibrium model where strategic thinking falls into identifiable discrete classes and decisions are unique best responses to some beliefs.

Theorem 4 derives dominant-strategy implementation from the incentive constraints and the heterogeneity of strategic thinking, rather than assuming it as in robust mechanism design analyses.

To assess the cost of giving up sensitive dependence on reported values, suppose that value densities are approximately uniform.

The equilibrium-incentive-efficient mechanism then approximately yields ex ante probability of trade 9/32 ≈ 28% and expected surplus 9/64 ≈ 0.14. The optimal posted price with uniform densities is ½, which then yields probability of trade 1/4 = 25% and expected surplus 1/8 = 0.125: a modest cost for robustness.

Such static posted-price mechanisms come closer to satisfying Wilson's (1987) desideratum, in that their *rules* are distribution-free. However, the optimal posted price is determined, via (5.15), by conditional means that are sensitive to the full value densities.

However, Čopič and Ponsatí's (2016 *JET*) continuous-time double auction with bids revealed to traders only once they are compatible avoids such dependence on the densities, in that any (even random) posted-price mechanism can be so implemented.

Design relaxing level-*k*-incentive-compatibility, with observable or predictable levels

Relaxing level-*k*-incentive-compatibility, one can still define a general class of feasible direct mechanisms; and with quasilinear utilities the payoff-relevant aspects of a mechanism are still described by outcome functions $p(\cdot, \cdot)$ and $x(\cdot, \cdot)$.

However, even a direct mechanism's incentive effects can no longer be tractably captured via incentive constraints. Instead they must be modeled via level-*k* traders' best responses to it.

For tractability, I focus on double auctions with reserve prices chosen by the designer, and on uniform value densities.

Reserve prices have no benefit when *L0* is uniform random on the full range of possible values [0, 1], as assumed so far.

But a restricted menu might make *Lk* players anchor beliefs instead on the correspondingly restricted range of bids or asks, which can make reserve prices useful (CKNP).

For example, in the double auction with uniform value densities, *L1* traders believe they face bids or asks uniformly distributed on [0, 1], which leads to incentive-*in*efficient outcomes.

To implement the outcome of MS's equilibrium-incentive-efficient direct mechanism via the double auction, *L1* traders have to believe that they face bids or asks uniform on [1/4, 3/4], the range of "serious" bids or asks in CS's linear double-auction equilibrium.

If *L1* traders anchor on the restricted menu, those beliefs can be induced by restricting bids to [1/4, 3/4] and asks to [1/4, 3/4]. (The upper ask limit could be moved to 1 and the lower bid limit to 0.)

Thus with uniform value densities, for *L1s* a double auction with reserve prices can mimic MS's equilibrium-incentive-efficient mechanism, whose direct form is then efficient in the set of *L1*-incentive-compatible mechanisms.

(MS's general specification of feasible mechanisms implicitly allows reserve prices, and their analysis therefore shows that if equilibrium is assumed, reserve prices are not useful here.)

For *L2*s with uniform value densities, my analysis of the double auction without reserve prices shows that it can improve upon a mechanism that is efficient in the set of *L2*-incentive-efficient mechanisms, or MS's equilibrium-incentive-efficient mechanism.

It can be shown that reserve prices allow no further improvement, so the double auction without reserve prices is optimal for *L2*s.

Design relaxing level-*k*-incentive-compatibility, with unobservable and unpredictable, heterogeneous levels

With unobservable and unpredictable, heterogeneous levels, suppose for tractability that the designer knows that the population includes multiple levels, but that one's frequency is very high.

Then a mechanism that would be efficient in the set of level-*k*-incentive-compatible mechanisms for the frequent level, or perhaps a level-*k*-incentive-efficient mechanism (relaxing level-*k* incentive-compatibility) for that level, can generally improve upon a mechanism that is efficient in the set of level-*k*-incentive-compatible mechanisms for the full heterogeneous population.

Such a mechanism gains the benefits of sensitive dependence on reported values for most traders, at bounded cost for the rest.

For example, in the case of approximately uniform value densities a mechanism that is efficient in the set of level-*k*-incentive-compatible mechanisms is a posted-price mechanism, with approximate optimal posted price ½, probability of trade 1/4 = 25%, and expected total surplus 1/8 = 0.125.

By contrast, a mechanism that is efficient in the set of *L1*-incentive-compatible mechanisms with uniform value densities yields ex ante probability of trade 9/32 ≈ 28% and expected surplus 9/64 ≈ 0.14 for almost all traders, a significant gain.

Alternatively, the *L2*-incentive-efficient double auction without reserve prices yields ex ante probability of trade 25/72 ≈ 35% and expected surplus 11/72 ≈ 0.15, an even larger gain.

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