Robust Bilateral Trade and an Essay on Awareness as an Equilibrium Notion

Thesis by

Jernej Čopič

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2006

(Submitted June 1, 2006)

Acknowledgements

To my parents, and to Mamajrena

I wish to thank to all who have helped me along the way. Debrah has been a source of happiness, and without her uncompromising love and support, I would have been miserable. I would also have been deported for failing to do my tax reports, and in general I would have been a mess. She has also helped me immensely with my research, reading and discussing with me an infinite number of versions of my work. Matt Jackson, my adviser, has been relentless in trying to extract the most out of me. His energy and help have been immeasurable, and I will not forget the wonderful bike rides through the San Gabriels. John would patiently meet with me and then make me excited with his love for theory. Bill has not only taught me how to write, how to present, and how to engage in a scholastic discussion without being intimidated. He has also supported me at every step, professionally, as a tutor, and as a friend, from the first moment we met. Federico has also been very helpful, and has given me numerous comments. The interaction with my committee has been a privilege. Clara played a very important role in my thesis which is clear from chapters 1 and 2. She has been a great co-author and a great friend. I have also learnt a great deal from my friend Mimmo, who he has unselfishly sacrificed his sleep for our joint project-see chapter 3. With his immense enthusiasm, working with him has been great fun. I am indebted to Ken, who is one of the most impressive people I have ever met, and has helped me greatly at many moments. I am grateful to my friends from Autonoma, Laura, Carmelo, Julius, and Jo. Clara needs to be mentioned again, as my being at Autonoma is a consequence of her. I am grateful to a few other people at Caltech, in particular Peter, Jonathan, Preston, Colin, Jey, and Elena. I also want to thank those from elsewhere who have given me opportunities through various invitations: Rani, Ariel, and Hugo. Again, I must feel privileged for having met those wonderful people. Finally, I thank my friends who have been throwing me off the mountain tops, over the waterfalls, down the canyons, and from the airplanes, all of which I have enjoyed greatly.

Abstract

The aim of this thesis is to analyze various effects of informational constraints. In chapters 1 and 2 we consider a robust model of bilateral trade where traders have private reservation values and utility functions are common knowledge. In chapter 1 we study direct-revelation mechanisms. Under incentive and participation constraints, we define the notion of *ex-post* constrained efficiency, which does not depend on the distribution of types. Given *ex-post* incentive and participation constraints, a sufficient condition for constrained efficiency is simplicity: the outcome is a lottery between trade at one type-contingent price and no trade. For constant-relative-risk-aversion environments, we characterize simple mechanisms. Under risk neutrality they are equivalent to probability distributions over posted prices. Generically, simple mechanisms converge to full efficiency as agents' risk aversion goes to infinity. Under risk neutrality, *ex-ante* optimal mechanisms are deterministic, and under risk aversion, they are not.

In chapter 2 we address indirect implementation. We define Mediated Bargaining Game-a continuous-time double auction with a hidden order book. It is the optimal bargaining game in the sense that its *ex-post* Nash equilibria in weakly undominated strategies constitute the Pareto-optimal frontier of the set of all *ex-post* Nash equilibria of all bargaining games. In Mediated Bargaining Game, type-monotone Bayesian equilibria coincide with *ex-post* Nash equilibria. The inefficiency due to incomplete information is manifested through delay. In contrast with the direct revelation mechanisms, in Mediated Bargaining Game the mechanism designer does not need to know the agents' risk attitudes.

Informational constraints may also be a result of agents' subjective knowledge of the economic situation. In chapter 3 we study normal-form games, where each player may be aware of a subset of the set of possible actions, and has a set of possible awareness architectures. Awareness architecture is given by agents' perceptions, and an infinite regress of conjectures about others. Awareness equilibrium is a steady state where neither actions nor awareness architectures can change. We provide conditions under which awareness equilibria exists and study a parametrization of the set of possible awareness architectures.

Contents

Acknowledgements **iii**

Abstract **v**

1 *Ex-Post* Constrained-Efficient Bilateral Trade with Risk-averse Traders **1**

1.1 Introduction 1

1.2 The problem 4

1.2.1 Efficiency and *ex-post* constrained efficiency 7

1.2.2 Sufficient conditions for *cexpiric*; Simple mechanisms 10

1.2.3 Differentiable mechanisms and first-order conditions 12

1.3 Constant relative risk-aversion environments 13

1.3.1 Risk neutrality 13

1.3.2 Risk aversion 16

1.3.3 *Ex-ante* optimality 17

1.4 Conclusion 20

Bibliography **32**

2 Optimal Robust Bargaining Games **34**

2.1 Introduction 34

2.2 Related literature 38

2.3 Dynamic bargaining games and robustness 39

2.4 Direct revelation mechanisms 42

2.5 Mediated Bargaining Game 43

2.6 Risk aversion and unequal discount rates. 51

2.7 Separating PBE are REQ 55

Bibliography **67**

3 Awareness as an Equilibrium Notion **70**

3.1 Introduction 70

3.2 Awareness equilibrium in normal-form games 72

3.3 A model of cognitive bounds 77

3.4 Conclusion 82

Bibliography **86**

Chapter 1

Ex-Post Constrained-Efficient Bilateral Trade with Risk-averse Traders

The work presented in this chapter has been done jointly with Clara Ponsatí.

1.1 Introduction

Bilateral trade is a fundamental problem of economics. A unit of an indivisible commodity is to be traded between a seller and a buyer. The seller has a private cost of producing the good, and the buyer has a private valuation, these are traders' *types*. Traders may be risk averse, the general shape of their utility functions determines the *environment*, which is common knowledge. [1] A desirable model of this situation ought to be robust, that is, not too sensitive to the details of the specification of traders' information.[2] If trade is voluntary, agents have incentives to misrepresent their private information, and efficient exchange is impossible. We provide optimality bounds for the allocations that can be achieved in equilibrium.

Our work brings two innovations. First, we introduce *ex-post* constrained efficiency as the optimality criterion which is congruent with robustness. Second, our

[1] Myerson and Satterthwaite [1983] address this problem under the assumptions that agents are risk neutral, and that types are drawn independently from distributions which are common knowledge.

[2] Wilson [1987] advocates detail-free approach to mechanism design. Robust mechanisms for the bilateral-trade problem were first discussed by Hagerty and Rogerson [1987].

method allows us to analyze risk-averse environments, and consequently allows for making comparative statics across different environments. Clearly, risk aversion plays an important role in many bilateral settings, such as wage bargaining or real-estate markets. In the absence of noncooperative theories, applied economists have used cooperative bargaining solutions to analyze such settings. Requiring robustness simplifies the mechanism-design problem, and allows for the analysis of general environments with risk aversion.

There are two aspects to the question of what are the second-best allocations that may be achieved in equilibrium. First, since under robustness the traders need not know the distribution of types, the appropriate equilibrium constraints are the *ex-post* incentive and participation constraints. In an environment with private values (such as the present one), this observation is due to Ledyard [1978]. More recently, Bergemann and Morris [2005], Chung and Ely [2005], and Jehiel et al. [2005], provide foundational work for robust implementation. Second, when the mechanism designer does not know the distribution of agents' types he can only Pareto maximize traders' *ex-post* utility allocations, and the class of mechanisms that obtain is the class of *ex-post* constrained-efficient mechanisms. Under *interim* incentive and participation constraints, an analogous notion is the *ex-post* incentive efficiency, due to Holmstrom and Myerson [1983], and *ex-post* constrained efficiency is its natural extension. *Ex-post* constrained efficiency is defined via *ex-post* Pareto domination, where a mechanism in order to dominate some other mechanism, must allocate better utilities to *all* draws of types. In comparison, *ex-ante* domination requires that the mechanism dominate another mechanism *on average*, which results in a much stronger notion of constrained efficiency but depends on the distribution of agents' types.[3] *Ex-post* constrained efficiency is a Paretian criterion, allowing for general statements about risk-averse environments, where utility is nontransferrable.

In Section 2, we provide sufficient conditions for *ex-post* constrained efficiency, un-

[3]Much of the discussion in Holmstrom and Myerson[1983], relating *ex-post* incentive efficiency to other notions of efficiency, applies also to the present context with *ex-post* incentive and participation constraints, e.g. every *ex-ante* constrained efficient mechanism has to satisfy *ex-post* constrained efficiency, but the reverse need not be the case.

der *ex-post* incentive and participation constraints. In additions to these constraints, it suffices that the mechanism be *simple*, and that trade takes place with probability one when reports are the lowest-cost and the highest-valuation. Simplicity means that the mechanism can be described by two functions of traders' reports: a probability of transferring the object and the price at which to trade, conditional on the object being transferred. Under risk neutrality, all mechanisms are representable in a simple form, but in general environments, not all mechanisms are simple.[4] An example of a simple mechanism is a posted price, but except for risk-neutral environments, randomizations over posted prices are neither simple, nor *ex-post* constrained efficient. In a simple mechanism, the traders have incentives to report truthfully, as a result of a tradeoff between the probability of trade, and the price that they obtain.

In Section 3, we analyze constant relative risk aversion environments. Under risk neutrality, we first provide a general characterization of *ex-post* constrained-efficient mechanisms, under *ex-post* incentive and participation constraints, as lotteries over posted prices.[5] We then provide a characterization of simple mechanisms for environments with risk aversion, and these are no longer representable by lotteries over posted prices. When agents become infinitely risk averse, the allocations generically converge to full *ex-post* efficiency.

We conclude our analysis by an example of *ex-ante* constrained efficiency. That example shows how the characterization of *ex-post* constrained efficiency can be used as a tool in the analysis of *ex-ante* welfare. Under risk neutrality, for a given type distribution, the *ex-ante* constrained-efficient mechanism is a posted price, while under risk aversion, it is a mechanism in which the trading price depends on traders' valuations. Assuming that in a world with stationary uncertainty, only *ex-ante* constrained-efficient exchanges should be observed, this provides a positive observation. In markets with large risk, relative to agents' wealth, we observe dispersed prices (correlated

[4]Under risk aversion, a nonsimple mechanisms may satisfy incentive constraints, but when it is recast into a simple form (in equilibrium, via certainty equivalents), the resulting simple mechanism need not satisfy incentive constraints.

[5]This is a generalization of the Hagerty and Rogerson [1987] results, who prove that under additional conditions, mechanisms are representable as lotteries over posted prices. The conditions they impose preclude efficiency assessments.

to agents' valuations), while in markets where risk is small, we observe posted pricing. Anecdotal evidence, corroborating this observation is abundant: objects of small value are generally exchanged at posted prices, while in markets for objects with large values, such as real estate, the prices are generally negotiated; in markets of the underdeveloped world, where there is arguably more risk, many more goods are bargained at bazaars.

Section 4 provides a short conclusion.

1.2 The problem

A seller, s, and a buyer, b, bargain over the price of an indivisible commodity. A trader i's payoff from trading at a price $p \in [0,1]$ is given by utility function $u_i(v_i, p)$, and traders obtain 0 if no trade takes place. We assume that for $i = s, b$, $u_i(v_i, p) : [0,1] \times [0,1] \to R$ is twice continuously differentiable, and $u_i(p,p) = u_i(v_i, \omega_{NT}) = 0$, where ω_{NT} denotes the no-trade outcome. Furthermore, $u_s(v_s, p)$ is increasing in p, decreasing in v_s, concave in each parameter, and satisfying the single-crossing condition $\frac{\partial^2 u_s}{\partial v_s \partial v_p} \leq 0$. Similarly, $u_b(v_b, p)$ is decreasing in p, increasing in v_b, concave in each parameter, and satisfying the single-crossing condition $\frac{\partial^2 u_b}{\partial v_b \partial v_p} \leq 0$. For instance, if $u_s(v_s, p) = u_s(p - v_s)$ and $u_b(v_b, p) = u_s(v_b - p)$, $u_i : [0,1] \to R, i = s, b$, are increasing, concave, and twice differentiable, then the above assumptions are satisfied. We denote $u = (u_s, u_b)$, and we call u the *environment.*

Parameters v_s and v_b are traders' private reservation values, or *types.* The interpretation is that v_s is the seller's cost of producing the good, and v_b is the buyer's valuation of the good. It is common knowledge that pairs of types $v = (v_s, v_b)$ are drawn from $[0,1] \times [0,1]$, according to *some* continuous joint distribution function F, with a strictly positive density f on $[0,1] \times [0,1]$.[6] We stress that congruent with the notion of robustness F need not be common knowledge, so that traders may have different beliefs about F, and different beliefs about the beliefs of the other trader

[6] We could generalize our analysis to environments where vector v is drawn from $[\underline{v}, \overline{v}] \times [\underline{v}, \overline{v}]$, $\underline{v} < 1/2 < \overline{v}$. This is equivalent to the requirement that F has support on $[\underline{v}, \overline{v}] \times [\underline{v}, \overline{v}]$. Note also that common knowledge of the support of F is sufficient for our analysis, but it might not be necessary.

and so on. We abstract from these considerations by simply assuming that the details of F are unknown to the traders, and we use the appropriate equilibrium notion consistent with this assumption.

A *direct revelation mechanism* (from now on a *mechanism*) is a game form, mapping agents' reports of their reservation values into outcomes. Denote by p_s the payment received by the seller, and by p_b the price charged to the buyer. We assume that outcomes have to be feasible, so that $p_s \leq p_b$; ω_{NT} is always feasible. We denote the vector of agents' reports by $\tilde{v} = (\tilde{v_s}, \tilde{v_b})$. Given agents' reports, an outcome is given by a lottery $\mu[\tilde{v}]$ over the feasible set, $\{(p_s, p_b) \mid p_s \leq p_b\} \cup \{\omega_{NT}\}$. Note that $\mu[\tilde{v}]$ is the lottery which the agents face *ex post*, after having reported their types. A mechanism m is thus a collection of lotteries $m = \{\mu[\tilde{v}] \mid \mathbf{supp}(\mu[\tilde{v}]) \subset \{(p_s, p_b) \mid p_s \leq p_b\} \cup \{\omega_{NT}\}, \tilde{v} \in [0,1]^2\}$.

Given a mechanism m, when agents report $\tilde{v}$, the expected utility of agent i with the reservation value v_i is

$$U_i^m(\tilde{v}, v_i) = E_{\mu[\tilde{v}]} u_i(v_i, p_i), i = s, b,$$

where E_μ denotes the expectation with respect to the measure μ. We slightly abuse the notation and denote by $U_i^m(v)$ the expected utility of agent i, $i = s, b$, when both agents report truthfully, $U_i^m(v) = U_i^m(v, v_i)$. We again stress that the expectation operator E_μ has nothing to do with the distribution of agents' types: utility allocations $U_i^m(\tilde{v}, v_i)$ are *ex-post* expected utilities that traders obtain in a mechanism m when they report $\tilde{v}$. Measure μ refers to the randomization over prices for a given reported vector of valuations $\tilde{v}$.

We consider mechanisms which are *ex-post* individually rational (XPIR) and *ex-post* incentive compatible (XPIC). We thus require that trade be always voluntary *ex post*, and reporting the reservation values truthfully be a dominant-strategy equilibrium.

(XPIR) Ex-post Individual Rationality.

Mechanism m is *ex-post individually rational* if

$supp(\mu[\tilde{v}]) \subset \{(p_s, p_s) \mid \tilde{v}_s \leq p_s \leq p_b \leq \tilde{v}_b\} \cup \{\omega_{NT}\}, \forall\ \tilde{v} \in [0,1]^2.$

(XPIC) Ex-Post Incentive Compatibility.

m is *ex-post incentive compatible* if

$$U_i^m(v) \quad \geq \quad U_i^m(\tilde{v}_i, v_j, v_i) \forall\ v_i \forall\ v_j \forall\ \tilde{v}_i, i = s, b, j \neq i.$$

From now on, whenever we write *XPIRIC mechanism* we mean a direct revelation mechanism satisfying XPIR and XPIC. We remark that since we are in a private-value setting XPIC is equivalent to dominant-strategy equilibrium of the game form defined by the direct-revelation mechanism. An example of XPIRIC mechanism is a *posted price.*

Example 1.2.1. In a posted price, the price is deterministic and is independent of the agents' reports. Once the traders observe the price, they can trade if they both find it optimal to do so. Formally,

$$\pi(v) = p \in [0,1], \varphi(v) \quad = \quad \begin{cases} 1 \text{ if } v_b \geq p \geq v_s, \\ 0 \text{ otherwise.} \end{cases}$$

In a posted price, it is clearly optimal for each trader to report his valuation truthfully, regardless of the report of the other trader, so that XPIC holds; XPIR is obviously satisfied.

It is well known and immediate to prove that XPIC implies monotonicity. It is also well known and easy to verify that XPIRIC imply uniform continuity of expected utility of trader i w.r.t. i's type (see Hagerty and Rogerson [1987], Theorem 1, for the case when traders are risk neutral). We provide the proof of monotonicity.

Lemma 1.2.2. *Let m be XPIC. Then $U_s^m(v)$ is strictly decreasing in v_s, whenever $U_s^m(v) > 0$; and $U_b^m(v)$ is strictly increasing in v_b, whenever $U_b^m(v) > 0$.*

Proof. We provide the proof for the seller. Let $U_s^m(v_s, v_b) > 0$, for some $0 < v_s < v_b$ and let $v_s' < v_s$. Then it must be that $\mu[v]$ assigns a positive probability to some feasible prices, so that by strict monotonicity of u_s in v_s, we have $U_s^m(v_s', v_b, v_s) > U_s^m(v_s, v_b, v_s)$. Hence, by XPIC,

$$U^m(v_s', v_b, v_s') \geq U^m(v_s', v_b, v_s) > U^m(v_s, v_b, v_s).$$

□

1.2.1 Efficiency and *ex-post* constrained efficiency

Next, we define the efficiency requirements. *Ex-post* efficiency is a standard requirement, albeit a very strong one.

(EFF) Ex-Post Efficiency.
m is *ex-post efficient* if the allocation $(U_s^m(v), U_b^m(v))$ is Pareto-optimal for each $v \in [0,1]^2$.

Example 1.2.1, *continued.* No posted price satisfies EFF, since in a posted price it can always happen *ex post* that either $p > v_b > v_s$ or $v_b > v_s > p$.

XPIRIC and EFF mechanisms do not exist. The following proposition is a simple extension of the Myerson and Satterthwaite [1983] impossibility result to the present setup. The *ex-post* incentive and participation constraints are stronger than the *interim* constraints considered in Myerson and Satterthwaite [1983]. For that reason, the proof of the impossibility result is very simple in the present setup. Note that the impossibility result stated here is general and relies only on u being monotonic; Myerson and Satterthwaite [1983] result requires risk neutral traders.

Proposition 1.2.3. *There does not exist a XPIRIC bilateral-trade mechanism satisfying EFF.*

Proof. Let $m = \{\mu[v]; v \in [0,1]^2\}$ be an XPIRIC and EFF mechanism. We show that this is impossible. For $v \in [0,1]^2$ s.t. both traders are risk neutral on $\mathbf{supp}(\mu[v])$, define $\bar{\pi}(v) = E_{\mu[v]}[p]$. Clearly, for all such v, $U_i^m(v) = u_i(\bar{\pi}(v), v_i)$. Next, for all $v \in [0,1]^2$, s.t. at least one trader has a strictly concave utility function on $\mathbf{supp}(\mu[v])$, it has to be that $\mathbf{supp}(\mu[v])$ is a singleton. Otherwise the allocation under lottery $\mu[v]$ would not be Pareto efficient. Denote the price at which trade occurs by $\bar{\pi}(v)$, and again $U_i^m(v) = u_i(\bar{\pi}(v), v_i)$, for all such v. By Lemma 1.2.2, $\bar{\pi}(v)$ is increasing in both v_s and v_b. By XPIR, it must be that $\bar{\pi}(x,x) = x, \forall\ x \in [0,1]$. Now take a $v = (v_s, v_b), v_s < v_b$. If $\bar{\pi}(v) > v_s$, then b would miss-report to $v_b' = v_s$, so XPIC for b would be violated. If $\bar{\pi}(v) < v_b$, then s would miss-report to $v_s' = v_b$, a contradiction. □

Since EFF is not possible we consider XPIRIC mechanisms that attain constrained-efficient allocations. The constrained-efficiency criterion that we propose is the *ex-post constrained efficiency.* This notion is an extension of the *ex-post* incentive efficiency, due to Holmstrom and Myerson [1983].

(XPCE) Ex-Post Constrained Efficiency:

Denote the set of incentive and participation constraints by $\mathcal{IP}$ (these could be either *ex-post*, *interim*, or any other set of participation and incentive constraints). A mechanism m, satisfying $\mathcal{IP}$, is *ex-post dominated, under* $\mathcal{IP}$, by another mechanism m', $m' \succ_{xp|\mathcal{IP}} m$, if m' satisfies $\mathcal{IP}$, and

$$U_s^{m'}(v, v_s) \geq U_s^m(v, v_s) \text{ and } U_b^{m'}(v, v_b) \geq U_b^m(v, v_b), \forall\ v,$$

with a strict inequality for an open set of v's, for at least one of the traders. A mechanism m is *ex-post constrained efficient* under $\mathcal{IP}$, if there does not exist a mechanism m' s.t. $m' \succeq_{xp|\mathcal{IP}} m$. We call XPCE mechanisms, under XPIRIC, *cexpiric* mechanisms.

The notion of *ex-post* constrained efficiency is tailored to our assumption that the joint distribution of traders' valuations has a full support and is continuous (regardless of the exact shape of the distribution). The requirement that the strict inequality hold for an open set of types is then equivalent to requiring that the event in which at least one player is strictly better off have a nonzero probability. Equivalently we could require that for at least one trader, the Lebesgue measure of the set of types that are strictly better off be positive.

Example 1, *continued.* Let $\bar{p}$ and $\bar{\bar{p}}$ be two posted prices, $0 \leq \bar{p} < \bar{\bar{p}} \leq 1$. Then neither $\bar{\bar{p}} \succ_{xp|\mathcal{IP}} \bar{p}$ nor the other way around. To see for instance the former, observe that under $\bar{p}$ the draws of types v s.t. $v_s < \bar{p} < v_b$ all obtain a strictly positive utility, while under $\bar{p}$ these pairs of traders obtain a 0 utility.

When $\mathcal{IP}$ are the *interim* incentive and participation constraints, this constrained efficiency notion is equivalent to the *ex-post* incentive efficiency as defined by Holmstrom and Myerson [1983]. Per se, XPCE does not depend on the specification of the distribution of agents' types, so that this is an optimality criterion that is suitable for robustness. Moreover, since it is a Paretian criterion, no assumptions are made on the interpersonal utility comparisons, which is important for the environments with risk aversion (i.e., environments with nontransferable utility).

Clearly,

$$\emptyset = \{m \mid m \text{ XPIRIC and EFF } \} \subset \{m \mid m \ \mathit{cexpiric}\},$$

where the first equality follows from Proposition 1.2.3. The requirements under XPIR and XPIC defined above are the strongest participation and incentive-compatibility criteria, but XPCE is the weakest constrained-efficiency notion; XPIR and XPIC are stronger than their *interim* analogs, while XPCE is weaker than *interim* constrained efficiency, which in turn is weaker than the *ex-ante* constrained efficiency. In other words, regardless of what is specified by $\mathcal{IP}$, the sets of the *ex-ante* and the *interim* constrained-efficient mechanisms are supersets of EFF mechanisms, and subsets of XPCE mechanisms.

1.2.2 Sufficient conditions for *cexpiric*; Simple mechanisms

We start by a simple sufficient condition for *cexpiric*.

Proposition 1.2.4. *Posted prices are cexpiric.*

Proof. Let m be a posted price, given by some $p^* \geq 0$. That m satisfies XPIRIC is obvious. We show that there are no XPIRIC mechanisms which *ex-post* dominate posted prices.

Suppose there exists a m' s.t. $m' \succeq_{xp} p^*$, (we use p^* to refer both to the mechanism m and to the posted price). Since on the set $\{v \mid v_s \leq p^* \leq v_b\}$ the allocation under m is Pareto optimal, the allocation under m' has to coincide with the allocation under p^* on that set. In particular, on the line segments $v_s = p^*$ and $v_b = p^*$, m' is identical to p^*, since otherwise the XPIR constraints for m' would be violated. Thus, by monotonicity of $U_i^{m'}$ w.r.t. v_i, $U_s^{m'}(v) = 0$ for $v_s > p^*$, and $U_b^{m'}(v) = 0$ for $v_b < p^*$. Since $m' \succeq_{xp} p^*$, by definition of *ex-post* constrained efficiency, there exists either an open rectangle $\Gamma \subset \{v \mid p^* \geq v_b > v_s\}$ s.t. $U_s^{m'}(v) > 0$ for $v \in \Gamma$, or an open rectangle $\Gamma' \subset \{v \mid p^* \leq v_s < v_b\}$ s.t. $U_b^{m'}(v) > 0$ for $v \in \Gamma'$. Both of these two cases are treated in exactly the same way so we consider the first possibility. Since $U_b^{m'}(v) = 0$ for $v_b < p^*$ (by monotonicity of $U_b^{m'}$ and $U_b^{m'}(v_s, p^*) = 0$), it is clear that $U_b^{m'}(v) = 0$ for $v \in \Gamma$, and since for $v \in \Gamma$, $U_s^{m'}(v) > 0$, it must be that on Γ, m' is a mechanism that for the buyer randomizes between one price $\pi'(v) = v_b$ and ω_{NT}, and the probability on $\pi'(v) = v_b$ must be positive. Denote this probability by $\varphi_b'(v)$. So fix a $\bar{v} \in \Gamma$. Clearly, $v_b = p^*$ has incentives to report $\bar{v}_b$ instead of p^* since $\varphi_b'(\bar{v})u_b(\bar{v}_b, p^*) > 0 = U_b^{m'}(v_s, p^*)$, a contradiction. $\square$

On a more abstract level, one can think of a mechanism as an assignment of feasible *ex-post* utility payoffs. Under risk neutrality, the standard parametrization of these payoffs is by specifying, at each vector of reports, the probability of transferring the object, and the expected monetary transfer between the traders. Such parametrization is without loss of generality only under risk neutrality, if XPIRIC hold. A slightly different parametrization is more convenient here. In general environments, we parametrize agents' expected utilities by the *probability of trade* and the *price at which to trade, conditional on trade taking place* (both are functions of agents' reports). Under risk neutrality, this parametrization is equivalent to the standard one. In general environments, we call the mechanisms that can be parametrized in this way *simple mechanisms*.[7]

We define a *simple mechanism* m as a mechanism where each $\mu[\tilde{v}]$ is a binary lottery between one price and ω_{NT}. A simple mechanism m is represented by a pair of functions $(\pi, \varphi) : [0,1]^2 \to [0,1]^2$. Given agents' reports, $\pi(\tilde{v})$ is the price at which the agents trade, $\varphi(\tilde{v})$ is the probability of trading at that price, and $1 - \varphi(\tilde{v})$ is the probability of ω_{NT}.

Example 1, *continued.* A posted price $\bar{p}$ is a simple mechanism: $\pi(v) = \bar{p}, \forall\, v \in [0,1]^2$; $\varphi(v) = 1$ if and only if $v_b \geq \bar{p} \geq v_s$, and $\varphi(v) = 0$ otherwise.

As we have shown in Proposition 1.2.4, one (very strong) sufficient condition for *cexpiric* is that a mechanism is a posted price. We can relax this condition considerably. Under a mild assumption on the utility functions, simple mechanisms satisfying XPIRIC, and s.t. the lowest-cost seller and the highest-valuation buyer trade *ex-post* with certainty are *cexpiric*.

Theorem 1.2.5. *Let* $u_s(v_s, p) = u_s(p - v_s)$ *and let* $u_b(v_b, p) = u_b(v_b - p)$*, let* $u_i''(x) < 0$*,* $\forall\, x \in [0,1]$ *for at least one* i*, and let the following condition hold:*

$$u_i(x) \not\equiv x \Rightarrow u_i'''(x) \geq 0, i = s, b.$$

If $m = (\pi, \varphi)$ *is a simple, XPIRIC mechanism for the environment specified by* u*, and* $\varphi(0,1) = 1$*, then* m *is cexpiric.*

Proof. See the Appendix. □

In the following example, we provide two mechanisms. One mechanism is simple,

[7] Clearly, if $\mathcal{IP}$ are not imposed then such parametrization is always without loss of generality. However, in a general environment, when a mechanism m satisfying XPIRIC, but which is not *simple*, is reparametrized into a simple form, it may no longer satisfy XPIRIC.

the other is not. In this example, the simple mechanism satisfies the assumptions of Theorem 1.2.5 and it *ex-post* dominates the nonsimple mechanism.[8]

Example 1.2.6. Let $u_s(p, v_s) = (p - v_s)^\gamma, u_s(p, v_b) = (v_b - p)^\gamma, \gamma \in (0, 1]$. Consider the following two mechanisms. Let m be simple and given by $\pi(v_s, v_b) = \frac{v_s + v_b}{2}$ and $\varphi(v_s, v_b) = \max\{0, (v_b - v_s)^\gamma\}$. It is easy to check that m is XPIRIC. Next, let $\bar{m}$ be given by lottery $F_p \equiv U[0, 1]$ over posted prices, where $U[0, 1]$ denotes the uniform distribution over $[0, 1]$. In other words, $\bar{m}$ is a mechanism where the price is drawn randomly from a uniform distribution, and the traders trade if it is individually rational for both. When traders are risk neutral, i.e., $\gamma = 1$, $U_i^m(v) = U_i^{\bar{m}}(v), \forall\ v, i = s, b$, so that m, $\bar{m}$ are equivalent. When $\gamma < 1$, $\bar{m}$ is not simple, and it is *ex-post* dominated by m. We return to this in Section 1.3.

1.2.3 Differentiable mechanisms and first-order conditions

If in a mechanism m the expected utilities of agents are differentiable, then the XPIC can be specified as a first-order condition (FOC). In this subsection, we show that if a simple mechanism is differentiable, then this FOC is necessary and sufficient, so that all simple differentiable XPIRIC mechanisms are given as all possible differentiable solutions (π, φ) to the FOC.

Given a mechanism m, we denote by S^m the set of types where both traders obtain a strictly positive expected utility under truthful reports. When $m = (\pi, \varphi)$, $S^{\pi,\varphi}$ can be written as

$$S^{\pi,\varphi} = \{v \mid v \in [0, 1] \times [0, 1], \varphi(v) > 0, v_s < \pi(v) < v_b\}.$$

(DIFF) Differentiability. A mechanism m is differentiable if $U_i^m(v)$ are differentiable on S^m.

We remark that a simple XPIRIC mechanism (π, φ) is differentiable if and only

[8] Our conjecture is that under the above assumptions on the environment, simplicity is also necessary for *ex-post* constrained efficiency. Insofar we have been unable to prove this.

if π and φ are both differentiable, which follows from the Implicit Function Theorem and the fact that XPIC implies $U_i^m(v)$ is strictly monotonic in v_i on S^m. *Cexpiric* mechanisms that satisfy DIFF will be referred to as *dexpiric*.

Proposition 1.2.7. *A simple and DIFF mechanism $m = (\pi, \varphi)$ is XPIRIC if and only if,* $\forall\ v \in S^{\pi,\varphi}$,

$$\begin{aligned} \frac{\partial\varphi(v)}{\partial v_s} u_s(\pi(v), v_s) &= -\varphi(v)\frac{\partial u_s(\pi(v),v_s)}{\partial p}\frac{\partial\pi(v)}{\partial v_s}, \\ \frac{\partial\varphi(v)}{\partial v_b} u_b(\pi(v), v_b) &= \varphi(v)\frac{\partial u_b(\pi(v),v_b)}{\partial p}\frac{\partial\pi(v)}{\partial v_b}. \end{aligned} \tag{1.1}$$

Proof. We derive the FOC for the seller. It is necessary that

$$\frac{\partial U_s^m(v, v_s')}{\partial v_s'} \mid_{v_s'=v_s} = 0,$$

which gives the desired condition. For sufficiency see the Appendix. □

The interpretation of this FOC is that the agents are provided with the correct incentives by a marginal tradeoff between the price and the probability of trade.

1.3 Constant relative risk-aversion environments

In this section, we analyze symmetric *constant relative risk aversion* (CRRA) environments. CRRA utility functions are specified by $u_s(p, v_s) = (p - v_s)^{\gamma_s}$ and $u_b(p, v_b) = (v_b - p)^{\gamma_b}$, where $\gamma_i \in (0, 1], i = s, b$, and by symmetry we mean that $\gamma_s = \gamma_b = \gamma$. Note that when $\gamma = 1$ this is the standard risk-neutral environment, and as γ tends to 0, agents' risk aversion tends to infinity.

1.3.1 Risk neutrality

When traders are risk neutral, the set of *cexpiric* mechanisms is equivalent to the set of probability distributions over posted prices, in terms of utility allocations to the traders. A distribution over posted prices is a mechanism given by some distribution function $F_p : [0, 1] \to [0, 1]$. The posted price p is drawn at random according to F_p, independently from trader's reports, and the traders then trade at p if and only if trading at p is individually rational for both of them. Theorem 1.3.1 is a generalization of the Hagerty and Rogerson [1987] results.[9]

Note that if F_p is a probability distribution over posted prices (i.e., $F_p(1) = 1$) and $(\pi_{F_p}, \varphi_{F_p})$ is a simple representation of this mechanism, then $\varphi(0,1) = 1$. The lowest-cost seller and the highest-valuation buyer trade with probability 1, regardless of the specification of F_p. Since no types can trade with a probability higher than 1, $F_p(1) \leq 1$ is a feasibility restriction on the distributions.

Theorem 1.3.1. *For $u_i(x) = x, i = s, b$, a mechanism (π, φ) is XPIRIC if and only if there exists a distribution function F_p over posted prices, i.e., an increasing $F_p : [0,1] \to [0,1]$, with $F_p(0) = 0$ and $F_p(1) \leq 1$, such that*

$$\pi(v) = E_{F_p}[\omega \mid \omega \in (v_s, v_b)]$$

$$\varphi(v) = \max\left\{F_p(v_b) - F_p(v_s), 0\right\}.$$

Proof. First, a distribution $F_p(.)$ over posted prices satisfies XPIRIC, since every posted price is XPIRIC and F_p is independent of traders' reports. The *simple* representation of the mechanism given by F_p is $\varphi(v) = \max\left\{F(v_b) - F(v_s), 0\right\}$ and $\pi(v) = E_{F_p}[\omega \mid \omega \in (v_s, v_b)]$, i.e., expected payoffs are the same as those generated under F_p (it is very easy to verify this).

For the converse, take an XPIRIC (π, φ). It is enough to show that $\varphi(v) = \varphi(0, v_b) - \varphi(0, v_s)$ since we can then define $F_p(\omega) = \varphi(0, v_b)$ and it follows immediately that $\pi(v) = E_{F_p}[\omega \mid \omega \in (v_s, v_b)]$.

XPIRIC implies that $\varphi(.)$ is nonincreasing in v_s and nondecreasing in v_b. By Lemma 1.2.2, $U_i^{\pi,\varphi}(v)$ is monotonic in v_i, whenever $U_i^{\pi,\varphi}(v)$ is strictly positive. Take

[9] They establish that for a mechanism satisfying XPIRIC, and s.t. either (φ, π) are twice continuously differentiable, or the image of φ is $\{0, 1\}$, there exists a payoff-equivalent distribution over posted prices.

the seller and let $v_s' > v_s$. By XPIC,

$$\varphi(v_s, v_b)(\pi(v_s, v_b) - v_s) \geq \varphi(v_s', v_b)(\pi(v_s', v_b) - v_s), \text{ and}$$

$$\varphi(v_s', v_b)(\pi(v_s', v_b) - v_s') \geq \varphi(v_s, v_b)(\pi(v_s, v_b) - v_s').$$

By subtracting first the rhs, and then the lhs of the second inequality from the first inequality, we obtain

$$\varphi(v_s, v_b)(v_s - v_s') \geq U_s^{\pi,\varphi}(v_s, v_b) - U_s^{\pi,\varphi}(v_s', v_b) \geq \varphi(v_s', v_b)(v_s - v_s').$$

Thus, φ is weakly decreasing in v_s, and since φ is bounded this implies that $U_s^{\pi,\varphi}(v_s, v_b)$ is Lipschitz in v_s for each v_b, so that it is absolutely continuous. Hence, it is an integral of its derivative. Again, by above inequalities, $\frac{\partial U_s^{\pi,\varphi}(z,v_b)}{\partial v_s} = \varphi(v)$, whenever this derivative exists. Thus, $U_s^{\pi,\varphi}(v)$ can be expressed as

$$U_s^{\pi,\varphi}(v) = \int_{v_s}^{v_b} \frac{\partial U_s^{\pi,\varphi}(z, v_b)}{\partial v_s} dz = \int_{v_s}^{v_b} \varphi(z, v_b) dz.$$

Similarly, we obtain $U_b^{\pi,\varphi}(v) = \int_{v_s}^{v_b} \varphi(v_s, z)dz$, and adding the two equations yields

$$\varphi(v) = \frac{1}{v_b - v_s} \int_{v_s}^{v_b} \varphi(v_s, z) + \varphi(z, v_b) dz, \forall\ v \in [0,1]^2 .$$

The claim now follows from the following theorem. $\square$

Theorem 1.3.2. *Consider a function $\varphi(v_s, v_b)$, which is bounded, increasing in v_s, decreasing in v_b, and nonnegative, for $(v_s, v_b) \in [0,1]^2$. Let $\varphi(v_s, v_b)$ satisfy,*

$$\varphi(v_s, v_b) = \frac{1}{v_b - v_s} \int_{v_s}^{v_b} \varphi(\tau, v_b) + \varphi(v_s, \tau) d\tau, \forall\ (v_s, v_b) \in [0,1]^2. \tag{1.2}$$

Then $\varphi(v_s, v_b) = \tilde{\varphi}(v_b) - \tilde{\varphi}(v_s), \forall\ v_b \geq v_s$, where $\tilde{\varphi}(.)$ is some increasing function.

Proof. See the Appendix. $\square$

Each distribution function over posted prices is XPIRIC, but only the probability distributions are *ex-post* undominated. The proof follows directly from Theorem 1.3.1.

Corollary 1.3.3. *A mechanism m is cexpiric if and only if m can be represented as a probability distribution over posted prices.*

Proof. A distribution over posted prices, which is not a probability distribution, is *ex-post* dominated by some probability distribution. On the other hand, a probability distribution over posted prices is not *ex-post* dominated by another probability distribution over posted prices, the proof of which is the same as the proof that two posted prices do not *ex-post* dominate each other. □

We remark that by Corollary 1.3.3, the *dexpiric* mechanisms under risk neutrality are given simply by continuously differentiable probability distributions F_p over posted prices. Under risk neutrality, *dexpiric* mechanisms are therefore generic within the class of *cexpiric* mechanisms. Nonetheless, if the distribution of types were known, then the *ex-ante* optimal XPIRIC mechanism under risk neutrality is a degenerate distribution over posted prices (see Section 1.3.3), which is a discontinuous mechanism.

1.3.2 Risk aversion

We first treat the symmetric case, when both agents have the same risk-aversion parameter, i.e., when $u_i(x) = x^\gamma, i = s, b, \gamma \in (0, 1]$. Then we can explicitly compute all simple XPIRIC mechanisms. We use this result to show that in a sequence of symmetric environments, when relative risk aversion of traders converges to ∞ pointwise, every simple *cexpiric* m, satisfying $S^m = \{v \mid v_s < v_b\}$, converges to *ex-post* efficiency (EFF).

Proposition 1.3.4. *Let $u_i(x) = x^\gamma$ for $\gamma \in [0, 1], i = s, b$. Then a simple mechanism $m = (\pi, \varphi)$ is cexpiric if and only if*

$$\varphi(v) = \begin{cases} \left(\int_{v_s}^{v_b} dF(z)\right)^{\gamma}, & \textit{if } v_b \geq p \geq v_s, \\ 0, & \textit{otherwise}, \end{cases}$$

$$\pi(v) = \frac{1}{F(v_b) - F(v_s)} \int_{v_s}^{v_b} x dF(x), \ \textit{if } v_s < v_b,$$

(and $\pi(v) = v_s$, for $v_s \geq v_b$), for some probability distribution $F : [0,1] \to [0,1]$.

Proof. For $u_i(x) = x^{\gamma}, i = s, b$, wherever (π, φ) are differentiable, the first-order conditions (1.1) are,

$$\begin{array}{l} \frac{\partial \varphi(v)}{\partial v_s}(\pi(v) - v_s) + \gamma \frac{\partial \pi(v)}{\partial v_s}\varphi(v) = 0, \\ \frac{\partial \varphi(v)}{\partial v_b}(v_b - \pi(v)) - \gamma \frac{\partial \pi(v)}{\partial v_b}\varphi(v) = 0. \end{array} \tag{1.3}$$

By setting $\varphi(v) = \bar{\varphi}(v)^{\gamma}$ we obtain exactly the same system of equations for $(\pi, \bar{\varphi})$ as under risk neutrality, and the claim follows from Theorem 1.3.1. □

The following corollary is immediate.

Corollary 1.3.5. *For $u_i(x) = x^{\gamma}, \gamma \in (0,1), i = s, b$, no lottery over posted prices is XPCE. Conversely, a simple IRIC mechanism m that is not a posted price is not representable by a lottery over posted prices.*

Observe that Proposition 1.3.4 implies that every mechanism m, with $S^m = \{v \mid v_s < v_s\}$, satisfies the property that whenever agents become infinitely risk averse, the allocation converges to full efficiency. Under risk neutrality, such mechanisms are precisely probability distributions over posted prices with a full support.

1.3.3 *Ex-ante* optimality

We provide an example to illustrate the usefulness of the characterization in Proposition 1.3.4 in order to make statements about the *ex-ante* constrained-efficient mechanisms under risk aversion. We make two remarks. First, in order to perform *ex-ante* welfare analysis, one has to know the type distribution. The interpretation in the context of robustness is that this is a positive observation: if there is an underlying distribution of types, and we expect to observe only the constrained-efficient mechanisms, then *ex-ante* constrained efficiency is an appropriate notion. As we mentioned before, this class is a subclass of *cexpiric* mecanisms.

Second, we only proved that if incentive and participation constraints are met and trade assured for the maximum valuation and minimum cost pair, simplicity is sufficient - we did not prove that it is necessary. Thus, a nonsimple *ex-post* constrained efficient mechanism may exist, and it may be that such mechanism is *ex-ante* optimal. What the example shows is that necessarily, when agents are risk averse, the *ex-ante* optimal mechanism is not deterministic (trade may happen with positive probability not equal to 1). Namely, among the simple mechanisms the *ex-ante* optimal one is a lottery, and the only deterministic mechanisms are posted prices.[10]

The analysis of the present CRRA example, with $\gamma_s = \gamma_b = \gamma$, is simple because we have the closed-form solutions for all simple *cexpiric* mechanisms. A similar exercise could be performed more generally, but the computations would be numerical at all steps of the analysis.

We assume that the traders' types are i.i.d., uniformly distributed on $[0,1]$, so that $f(v_s, v_b) = 1, \forall\ v_s, v_b \in [0,1]$, where $f(.,.)$ is the density of F, the traders' distribution of types. To keep things simple we look at a utilitarian *ex-ante* social welfare function,

$$W^m = \int_{v_s \in [0,1]} \int_{v_b \in [0,1]} \left(U_s^m(v_s, v_b) + U_b^m(v_s, v_b)\right) f(v_s, v_b) dv_b dv_s,$$

where m is a mechanism.

The problem of designing the *ex-ante* optimal simple IRIC mechanism can be

[10]Another way to view this result is in terms of linear programing. Solving for the *ex-ante* optimal mechanism under risk neutrality is to solve a linear program on the convex set of *cexpiric* mechanisms, so that it is not surprising that a solution is generically a "corner" of this set, i.e., a posted price. When agents are risk averse, the *ex-ante* optimization is no longer a linear program, and the optimal mechanisms are more interesting.

written as

$$\begin{aligned} &\max\nolimits_m W^m \\ &\text{s.t. } m \text{ XPIRIC and } \text{ simple.} \end{aligned} \tag{1.4}$$

Every m which is *ex-ante* constrained efficient has to be *cexpiric*. Hence, it is in expression (1.4) enough to optimize over all *cexpiric* mechanisms. By Proposition 1.3.4 and since $f(.,.) \equiv 1$, the problem (1.4) can be written as

$$\max_{F_p} \int_0^1 \int_0^1 \left(\varphi(v_s, v_b)\left[(\pi(v_s, v_b) - v_s)^\gamma + (v_b - \pi(v_s, v_b))^\gamma\right]\right) dv_b dv_s,$$

where $\varphi(v_s, v_b) = (\max\{F_p(v_b) - F_p(v_s), 0\})^\gamma$ and $\pi(v_s, v_b) = E_{F_p}[p \mid v_s \leq p \leq v_b]$. This can be rewritten as

$$\max_{f_p} \int_0^1 \int_{v_s}^1 \left[\left(\int_{v_s}^{v_b} (t - v_s) f_p(t) dt\right)^\gamma + \left(\int_{v_s}^{v_b} (v_b - t) f_p(t) dt\right)^\gamma\right] dv_b dv_s.$$

Denoting $G(t) = \int_0^t F_p(\tau) d\tau$, and letting

$$\nu(v_s, v_b) = \left[G(v_b) - G(v_s)\right](v_b - v_s),$$

we can rewrite the above expression (integrate by parts each of the two innermost integrals and compute the appropriate derivatives) as

$$\max_{\nu} \int_0^1 \int_{v_s}^1 \left[\left(-2\frac{\nu(v_s, v_b)}{v_b - v_s} + \frac{\partial \nu}{\partial v_b}\right)^\gamma + \left(2\frac{\nu(v_s, v_b)}{v_b - v_s} + \frac{\partial \nu}{\partial v_s}\right)^\gamma\right] dv_b dv_s \tag{1.5}$$

The maximization problem (1.5) is a manageable optimization problem. We can in principle compute its solutions, using the calculus of variations. Except when $\gamma = 1$, we cannot compute the solutions in closed form. When $\gamma = 1$ the problem simplifies substantially since only the terms involving the derivatives of ν remain. It is then straightforward to compute that the *ex-ante* optimal mechanism is a posted price at $p = \frac{1}{2}$, which we can also easily deduce directly: there is no reason to randomize over suboptimal prices.

When $\gamma < 1$ the *ex-ante* optimal mechanism is not a posted price. To see this, compute the necessary first-order condition of (1.5),

$$\nabla . \mathcal{H}_{\nabla \nu} = \mathcal{H}_{\nu},$$

where $\mathcal{H} = \int_0^1 \int_{v_s}^1 \left[\left(-2\frac{\nu(v_s,v_b)}{v_b - v_s} + \frac{\partial \nu}{\partial v_b} \right)^{\gamma} + \left(2\frac{\nu(v_s,v_b)}{v_b - v_s} + \frac{\partial \nu}{\partial v_s} \right)^{\gamma} \right] dv_b dv_s$, and ∇ is the gradient operator.[11] The expression for the first-order condition is somewhat messy, but it is immediate that, when $\gamma < 1$, a constant function does not solve this equation, so that no posted price is a solution when $\gamma < 1$. Since the *ex-ante* optimal simple mechanism is not a posted price it then follows that *ex-ante* optimal mechanism must be probabilistic. It is also clear that a lottery over posted prices cannot be optimal since when $\gamma < 1$ simple mechanisms dominate such lotteries by the representation of Proposition 1.3.4. Thus an *ex-ante* optimal mechanism under risk aversion necessarily has the feature that prices depend on agents valuations, so that if there is dispersion of valuations we should also observe dispersion in prices.

1.4 Conclusion

We focused on the simplest exchange with a two-sided incomplete information. The key to our analysis is the use of the distribution-free concept of *ex-post* constrained efficiency, in conjunction with Theorem 1.2.5. These methods apply more generally. Immediate is the extension to the problem of providing a public good with private valuations, analogous to Mailath and Postlewaite [1990], but incorporating robustness and risk aversion.

Present results provide lower bounds for efficiency of optimal Bayesian mechanisms. We remark, however, that the results here hold for environments with correlated types, whereas under *interim* incentive and participation constraints, with correlation, full efficiency is possible (see Cremer and Maclean [1985,1988] and Mcafee and Reny [1992]).

[11] Here, $\mathcal{H}_{\nabla \nu}$ denotes the vector of partial derivatives of $\mathcal{H}$ w.r.t. all the components of $\nabla \nu$, and $\nabla . \mathcal{H}_{\nabla \nu}$ is the dot product of gradient operator with $\mathcal{H}_{\nabla \nu}$ - i.e., it is the sum of components of $\mathcal{H}_{\nabla \nu}$, each differentiated w.r.t. the appropriate component of v.

In the present paper, we addressed the case where the mechanism designer knows the shape of the agents utilities. This information is necessary for the designer to know, in order to be able to construct incentive-compatible direct-revelation mechanisms. In chapter 2, we show that when the mechanism designer does not know the shape of the traders' utility functions, but this information is known to the traders, the mechanism designer can construct an optimal indirect game form, the *mediated bargaining game*. The equilibria of the mediated bargaining game implement the *ex-post* constrained-efficient allocations described here.

Appendix

PROOF OF THEOREM 1.2.5.

We will need the following Lemma to show that a nonsimple mechanism cannot *ex-post* dominate a simple one.

Lemma 1.4.1. *Assume utilities depend only on the net surplus,* $u_s(v_s, p) = u_s(p - v_s)$ *and* $u_b(v_b, p) = u_s(v_b - p)$*,* $u_i : [0, 1] \to R, i = s, b$*. Also assume that* $u_i''(y) < 0, \forall\ y \in [0, 1]$*, for at least one* i*, and* $u_i'''(y) \leq 0, \forall\ y \in [0, 1], i = s, b$*. Next, let* μ *be a measure with supp*$(\mu) \subset [0, 1]$*, let Let* u_i *satisfy* $u_i''(y) < 0, \forall\ y \in [0, 1]$*, for at least one* i*, and* $u_i'''(y) \leq 0, \forall\ y \in [0, 1], i = s, b$*. Next, let* μ *be a measure with supp*$(\mu) \subset [0, 1]$*, let*

$$U_s^\mu = E_\mu[u_s(y)],$$

$$U_b^\mu = E_\mu[u_b(y)],$$

and define $p, \sigma \in [0, 1]$ *by* $\sigma u_s(p) = U_s^\mu, \sigma u_b(1 - p) = U_b^\sigma$*. Then at least one of the following must be true:*

1. μ *is a degenerate point-mass at* p *and* $\sigma = \mu[\{p\}]$*,*

2. $\sigma u_s'(p) < E_\mu[u_s'(y)]$*, or*

3. $\sigma u_b'(1 - p) < E_\mu[u_b'(1 - y)]$*.*

Proof. Suppose μ is nondegenerate. First note that p and σ are uniquely defined. Next, we can assume without loss of generality that by normalization, $\mu([0,1]) = 1$. Since $u_i''(y) < 0$, it follows by Jensen's inequality that

$$u_s\left(E_\mu y\right) \geq E_\mu[u_s(y)] = \sigma u_s(p),$$

$$u_b\left(E_\mu y\right) \geq E_\mu[u_b(y)] = \sigma u_b(1-p),$$

where at least one of the inequalities is strict, and $\sigma < 1$. If $E_\mu y \leq p$, then by convexity of u_s' and Jensen's inequality,

$$E_\mu u_s'(y) \geq u_s'(E_\mu y) \geq u_s'(p) > \sigma u_s'(p).$$

Alternatively, if $E_\mu y \geq p$, then by convexity of u_b',

$$E_\mu[u_b'(1-y)] \geq u_b'\left(1 - E_\mu y\right) \geq u_b'(1-p) > \sigma u_b'(1-p).$$

$\square$

Now we are ready to prove Theorem 1.2.5.

Proof. By XPIRIC, $U_i^m(v)$ is uniformly continuous and monotonic w.r.t. v_i, at each $v \in [0,1]^2$, s.t. $U_i^m(v) > 0$, implying that the left and the right limit of the partial derivative of $U_i^m(v)$ w.r.t. v_i exist. Thus, the XPIC constraints can be written as:

$$\frac{\partial^+ U_s^m(v)}{\partial v_s} \leq -E_{\mu[v]}[u_s'(x - v_s)] \leq \frac{\partial^- U_s^m(v)}{\partial v_s} \leq 0,$$

$$\frac{\partial^- U_b^m(v)}{\partial v_b} \geq E_{\mu[v]}[u_b'(v_b - x)] \geq \frac{\partial^+ U_b^m(v)}{\partial v_b} \geq 0.$$

This is easily verified using standard arguments. A mechanisms $m = \{\mu[v] \mid v \in [0,1]^2\}$ is differentiable at $v \in [0,1]^2$ if and only if the incentive constraints hold at v

with equalities, i.e.,

$$\frac{\partial U_s^m(v)}{\partial v_s} = -E_{\mu[v]}[u_s'(x - v_s)],$$

$$\frac{\partial U_b^m(v)}{\partial v_b} = E_{\mu[v]}[u_b'(v_b - x)].$$

As in the case of risk neutrality, $U_i^m(v)$ is absolutely continuous. Hence, for each v_j, $\frac{\partial U_i^m(v)}{\partial v_i}$ exists almost everywhere, and $U_i^m(v)$ is the integral of its derivative w.r.t. v_i. By XPIRIC, this gives

$$\begin{aligned} U_s^m(v_s, v_b) &= \textstyle\int_{v_s}^{v_b} E_{\mu[\tau, v_b]}\left[u_s'(x - \tau)\right] d\tau, \\ U_b^m(v_s, v_b) &= \textstyle\int_{v_s}^{v_b} E_{\mu[\tau, v_s]}\left[u_b'(x - \tau)\right] d\tau. \end{aligned} \tag{1.6}$$

Now let m be simple, $m = (\varphi, \pi)$, and the allocation at $v = (0, 1)$ be Pareto optimal. Assume there exists an $\tilde{m}$ which *ex-post* dominates m. Assume first that $\tilde{m}$ is simple, $\tilde{m} = (\tilde{\varphi}, \tilde{\pi})$.

At $v = (0, 1)$ the allocation assigned by $\tilde{m}$ must be the same as the allocation under m, by Pareto optimality. Take the line $L_s(1) = \{(v_s, 1) \mid v_s \in [0, 1]\}$. By assumption, $U_s^m(v) \leq U_s^{\tilde{m}}(v), \forall\ v \in L_s(1)$, and since the seller's XPIC constraints for m and $\tilde{m}$ hold almost everywhere on $L_s(1)$ with equality, we have by representation (1.6), that $U_s^m(v) = U_s^{\tilde{m}}(v), \forall\ v \in L_s(1)$. Thus,

$$\frac{\partial U_s^m(v)}{\partial v_s} = \frac{\partial U_s^{\tilde{m}}(v)}{\partial v_s}, \forall\ v \in L_s(1).$$

Since $\frac{\partial U_s^m(v)}{\partial v_s} = -\varphi(v) u_s'(\pi(v) - v_s)$, we therefore have

$$\varphi(v) u_s(\pi(v) - v_s) = \tilde{\varphi}(v) u_s(\tilde{\pi}(v) - v_s) \text{ and}$$

$$\varphi(v) u_s'(\pi(v) - v_s) = \tilde{\varphi}(v) u_s'(\tilde{\pi}(v) - v_s), almost everywhere on L_s(1).$$

These imply that $\varphi(v) = \tilde{\varphi}(v), \pi(v) = \tilde{\pi}(v), almost everywhere L_s(1)$, so that $U_b^m(v) = U_b^{\tilde{m}}(v), almost everywhere on L_s(1)$.

Similarly, define $L_b(0) = \{(0, v_b) \mid v_b \in [0,1]\}$, and by an analogous argument we obtain $U_b^m(v) = U_b^{\tilde{m}}(v), \forall\ v \in L_b(0)$ and $U_s^m(v) = U_s^{\tilde{m}}(v), almost everywhere on L_b(0)$.

Now take for instance a $v = (0, v_b) \in L_b(0)$ s.t. $U_s^m(v) = U_s^{\tilde{m}}(v)$ and let $L_s(v_b) = \{(v_s, v_b) \mid v_s \in [0,1]\}$. As before, we obtain $U_s^m(v) = U_s^{\tilde{m}}(v), \forall\ v \in L_s(v_b)$. Thus, the set where $U_s^m(v) \neq U_s^{\tilde{m}}(v)$ has Lebesgue measure 0 and hence cannot be open. Similarly, the set where $U_b^m(v) \neq U_b^{\tilde{m}}(v)$ has Lebesgue measure 0, so that $\tilde{m}$ cannot *ex-post* dominate m.

Assume then that $\tilde{m}$ is not simple. Again, by Pareto optimality at $v = (0,1)$, $\tilde{m}$ has to be simple at $(0,1)$. Thus,

$$\bar{v} = \sup_{v_s} \inf_{v_b} \{v \mid \tilde{m} \text{ simple at } v\}$$

is well defined, and $\bar{v} \in [0,1]$. Moreover, by (1.6), $U_i^m(\bar{v}) = U_i^{\tilde{m}}(\bar{v}), i = s, b$, so that Lemma 1.4.1 applies, and $\tilde{m}$ cannot *ex-post* dominate m.

□

PROOF OF SUFFICIENCY OF PROPOSITION 1.2.7.

Proof. Consider $U_s(v, v_s')$. We show that for all all $v_s' \neq v_s$ the derivative of $U_s(v, v_s')$ w.r.t. v_s' is decreasing whenever $U_s(v, v_s') > 0$ (deviations that give negative expected utility cannot be profitable). We consider $v_s' > v_s$ (the case $v_s' < v_s$ is analogous). Thus compute

$$\frac{\partial U_s(v, v_s')}{\partial v_s'} = \varphi(v_s', v_b) \frac{\partial u_s(\pi(v_s', v_b), v_s)}{\partial p} \frac{\partial \pi(v_s', v_b)}{\partial v_s'} + \frac{\partial \varphi(v_s', v_b)}{\partial v_s'} u_s(\pi(v_s', v_b), v_s).$$

From the first order condition we can express

$$\frac{\partial \pi(v_s', v_b)}{\partial v_s'} = -\frac{\frac{\partial \varphi(v_s', v_b)}{\partial v_s'} u_s(\pi(v_s', v_b), v_s')}{\varphi(v_s', v_b) \frac{\partial u_s(\pi(v_s', v_b), v_s')}{\partial p}}.$$

Substituting this into the previous expression we get

$$\frac{\partial U_s\left(v, v_s'\right)}{\partial v_s'} = \frac{\partial \varphi\left(v_s', v_b\right)}{\partial v_s'}\left[u_s\left(\pi_s\left(v_s', v_b\right), v_s\right) - \frac{\frac{\partial u_s\left(\pi\left(v_s', v_b\right), v_s\right)}{\partial p} u_s\left(\pi\left(v_s', v_b\right), v_s'\right)}{\frac{\partial u_s\left(\pi\left(v_s', v_b\right), v_s'\right)}{\partial p}}\right]$$

Observe that $\frac{\partial u_s(\pi(v_s', v_b), v_s)}{\partial p} < \frac{\partial u_s(\pi(v_s', v_b), v_s')}{\partial p}$. Moreover by XPIC $\frac{\partial \varphi(v_s', v_b)}{\partial v_s'} < 0$. To see this, one can use the standard argument of writing down the XPIC constraints for two types of the seller and then expressing the derivative of φ as the limit of taking one of the two types toward the other. Thus, whenever $u_s\left(\pi\left(v_s', v_b\right), v_s\right) > 0$, $\frac{\partial U_s(v_s; v_s', v_b)}{\partial v_s'}$ is a decreasing function of v_s, implying that the local maximum of U_s is unique, and is also a global maximum. Similarly for U_b. □

In the proof of Theorem 1.3.2 we apply the following simple Lemma a few times.

Lemma 1.4.2. *Let a function* $g : [0,1]^2 \to [0,1]$ *have the property that*

$$g(v_1, v_2) = \frac{1}{v_2 - v_1}\int_{v_1}^{v_2} g(\tau, v_2) d\tau, \forall\ v_1, v_2 \in [0,1], v_1 < v_2. \tag{1.7}$$

Then $g(v_1, v_2) = g(v_1', v_2), \forall\ v_1, v_1', v_2 \in [0,1], v_1, v_1' < v_2$.

Proof. It is enough to prove that if a function $\bar{g} : [0,1] \to [0,1]$ has the property that $\bar{g}(x) = \frac{1}{x}\int_0^x \bar{g}(t)dt$, then $\bar{g}(x)$ must be a constant. We show that $\bar{g}$ is continuous and differentiable on $(0,1]$ and that it's derivative is 0 on $(0,1]$. We first show that $\bar{g}$ is continuous on $(0,1]$. Take an $\bar{\epsilon} > 0$, and for $\epsilon > 0$ take $x, x' \in [\bar{\epsilon}, 1], |x - x'| < \epsilon$. Since $\bar{g}$ is nonnegative and bounded by 1 we have

$$|\bar{g}(x) - \bar{g}(x')| = \left|\frac{1}{xx'}\left(\int_0^{x'} (x' - x)\bar{g}(t)dt + \int_0^{x'} x'\bar{g}(t)\right) dt\right| \leq \frac{2\epsilon}{\bar{\epsilon}^2},$$

implying that $\bar{g}$ is continuous on $[\bar{\epsilon}, 1], \forall\ \bar{\epsilon} > 0$, so that it is continuous on $(0,1]$. Now observe that on $(0,1]$, $\bar{g}$ is a product of two continuously differentiable functions, hence it is continuously differentiable. Since

$$x\bar{g}(x) = \int_0^x g(t)dt,$$

we can take derivatives to obtain $x\bar{g}'(x) = 0, \forall\, x \in (0,1]$, so that $\bar{g}'(x) = 0, \forall\, x \in (0,1]$, and the claim follows. $\square$

Proof of Theorem 1.3.2.

Proof. We prove the Theorem in a few steps. To understand the logic it is best to think of $\tilde{\varphi}(.)$ as a distribution, which induces a measure $\tilde{\mu}$. We know that $\tilde{\varphi}(.)$ is continuous if and only if $\tilde{\mu}$ is continuous with respect to the Lebesgue measure, and that in general, $\tilde{\mu}$ can be decomposed into $\tilde{\mu}_l + \tilde{\mu}_o$, where $\tilde{\mu}_l$ is continuous w.r.t. the Lebesgue measure, and $\tilde{\mu}_o$ is orthogonal w.r.t. the Lebesgue measure (i.e., the jumps in $\tilde{\varphi}(.)$). Moreover, $\tilde{\varphi}(.)$ is continuous if and only if $\varphi(v_s, v_b)$ is continuous in each of the two dimensions (i.e., v_s and v_b). In case 1, we treat the problem when $\varphi(v_s, v_b)$ is continuous. In case 2, we treat the general problem when $\varphi(v_s, v_b)$ can be discontinuous. If $\varphi(v_s, v_b)$ is not continuous in each dimension the set of discontinuities of $\varphi(v_s, v_b)$ could be very complex. Then, the fact that $\varphi(v_s, v_b)$ can be represented by $\tilde{\varphi}(.)$ implies that the discontinuities of $\varphi(v_s, v_b)$ have a very specific structure. That is, if for a fixed v_s, $\varphi(v_s, \tau)$ is discontinuous at some $\bar{\tau} \geq v_s$, then $\varphi(v_s', \tau)$ is discontinuous at $\bar{\tau}$ for all $v_1' < \bar{\tau}$ (step 2.1), and $\varphi(\tau, v_b)$ is discontinuous at $\bar{\tau}$ for all $v_b > \bar{\tau}$ (step 2.2).

Case 1. Let $\varphi(v_s, \tau)$ and $\varphi(\tau, v_b)$ be continuous in τ, for every $(v_s, v_b) \in [0,1]^2$.

We define $\phi(v_s, v_b, t) = \varphi(v_s, t) + \varphi(t, v_b) - \varphi(v_s, v_b)$, and we prove that $\phi(v_s, v_b, t) = 0, \forall\, t \in [v_s, v_b]$. Note that ϕ is continuous in each of its arguments, in particular it is continuous in t. We proceed as follows. In step 1.1 we show that there exists a $\bar{t} \in (v_s, v_b)$ s.t. $\phi(v_s, v_b, \bar{t}) = 0$. In step 1.2 we show that $\frac{\partial \phi(v_s, v_b, t)}{\partial t} = 0$ everywhere by showing that the derivative of $\phi(v_s, v_b, t)$ w.r.t. t from the left is equal to that derivative from the right everywhere (and both are equal to 0). From the definition of ϕ it is clear that its derivative from the left w.r.t. t will be equal to 0 if and only if the derivative from the left of $f(v_s, t)$ w.r.t. t is equal the derivative from the left of $f(t, v_b)$ w.r.t. t, which is precisely what we show in step 1.2. Similarly for the derivative from the right. Thus, ϕ is differentiable, its derivative is 0, and it is equal to 0 at some point by step 1.1 - then ϕ must be equal to 0 everywhere. While step 1.1 is straightforward, step 1.2 involves some calculus.

Step 1.1. There exists a $\bar{t} \in (v_s, v_b)$ s.t. $\phi(v_s, v_b, \bar{t}) = 0$.

Proof. Now (1.2) can be written as

$$0 = \frac{1}{v_b - v_s} \int_{v_s}^{v_b} \phi(v_s, v_b, \tau) d\tau.$$

By the mean value theorem (MVT), there exists a $\bar{t} \in (v_s, v_b)$, s.t. $\frac{1}{v_b - v_s} \int_{v_s}^{v_b} \phi(v_s, v_b, \tau) d\tau = \phi(v_s, v_b, \bar{t})$, which concludes the proof of step 1.1.

Step 1.2. $\phi(v_s, v_b, t)$ is differentiable in t and $\frac{\partial \phi(v_s, v_b, t)}{\partial t} = 0$, for all $t \in (v_s, v_b)$.

Proof. Denote by

$$\frac{\partial^+ \varphi(v_s, t)}{\partial t} = \lim_{\epsilon \to 0,\ \epsilon > 0} \frac{\varphi(v_s, t + \epsilon) - \varphi(v_s, t)}{\epsilon}$$

the derivative from the right of $\varphi(v_s, t)$ w.r.t.t. Similarly, let $\frac{\partial^- \varphi(v_s, t)}{\partial t}$ denote the derivative from the left. We will show that for every $t \in (v_s, v_b)$,

$$\frac{\partial^+ \phi(v_s, v_b, t)}{\partial t} = \frac{\partial^- \phi(v_s, v_b, t)}{\partial t} = 0.$$

We will do that by showing that $\frac{\partial^+ \varphi(v_s, t)}{\partial t} = -\frac{\partial^+ \varphi(t, v_b)}{\partial t}$ and $\frac{\partial^- \varphi(v_s, t)}{\partial t} = -\frac{\partial^- \varphi(t, v_b)}{\partial t}$, for all $t \in (v_s, v_b)$. Note that the left and the right-derivatives of $\varphi(v_s, t)$ and $\varphi(t, v_b)$ w.r.t.t exist for all t since φ is continuous and monotonic.

We first show that

$$\frac{\partial^+ \varphi(v_s, t)}{\partial t} = \frac{\partial \varphi^+(v_s', t)}{\partial t}, \forall\ v_s', v_s < t. \tag{1.8}$$

To see this, we write by definition,

$$\frac{\partial^+\varphi(v_s,t)}{\partial t} = \lim_{\epsilon\to 0,\ \epsilon>0} \frac{1}{\epsilon}\left(\varphi(v_s,t+\epsilon) - \varphi(v_1,t)\right).$$

We now use (1.2) and compute

$$\varphi(v_s,t+\epsilon) - \varphi(v_s,t) = \int_{v_s}^{t+\epsilon} \frac{\varphi(v_s,\tau)+\varphi(\tau,t+\epsilon)}{t+\epsilon-v_s}d\tau - \int_{v_s}^{t} \frac{\varphi(v_s,\tau)+\varphi(\tau,t)}{t-v_s}d\tau$$
$$= \int_{t}^{t+\epsilon} \frac{\varphi(v_s,\tau)+\varphi(\tau,t+\epsilon)}{t+\epsilon-v_s}d\tau + \int_{v_s}^{t} \frac{\varphi(v_s,\tau)+\varphi(\tau,t+\epsilon)}{t+\epsilon-v_s} - \frac{\varphi(v_s,\tau)+\varphi(\tau,t)}{t-v_s}d\tau$$
$$= \int_{t}^{t+\epsilon} \frac{\varphi(v_s,\tau)+\varphi(\tau,t+\epsilon)}{t+\epsilon-v_s}d\tau + \int_{v_s}^{t} \frac{-\epsilon(\varphi(v_s,\tau)+\varphi(\tau,t))}{(t+\epsilon-v_s)(t-v_s)} + \frac{\varphi(\tau,t+\epsilon)-\varphi(\tau,t)}{t+\epsilon-v_s}d\tau$$
$$= \int_{t}^{t+\epsilon} \frac{\varphi(v_s,\tau)+\varphi(\tau,t+\epsilon)}{t+\epsilon-v_s}d\tau - \frac{\epsilon\varphi(v_s,t)}{t+\epsilon-v_s} + \int_{v_s}^{t} \frac{\varphi(\tau,t+\epsilon)-\varphi(\tau,t)}{t+\epsilon-v_s}d\tau$$

From this last expression we can see that $\lim_{\epsilon\to 0,\epsilon>0} \frac{1}{\epsilon}\left(\varphi(v_s,t+\epsilon)-\varphi(v_s,t)\right) = \frac{1}{t+\epsilon-v_s}\int_{v_s}^{t} \frac{\partial^+\varphi(\tau,t)}{\partial v_b}d\tau$, since

$$\lim_{\epsilon\to 0,\epsilon>0} \frac{1}{\epsilon}\int_{t}^{t+\epsilon} \frac{\varphi(v_s,\tau)+\varphi(\tau,t+\epsilon)}{t+\epsilon-v_s}d\tau - \frac{\varphi(v_s,t)}{t+\epsilon-v_s} = 0,$$

by the MVT.

By Lemma 1.4.2 this implies that indeed (1.8) holds. Similarly, we obtain $\frac{\partial^+\varphi(t,v_b)}{\partial t} = \frac{\partial\varphi^+(t,v_b')}{\partial t}, \forall\ v_b', v_b > t$.

Now take a monotonic sequence $\epsilon_n, n = 1,...,\infty$, s.t. $\lim_{n\to\infty}\epsilon_n = 0$, and let $v_{b,n}' = t+\epsilon_n$. By above, for each n,

$$\lim_{l\to\infty, l\geq n} \frac{\varphi(t+\epsilon_l, v_{b,n}') - \varphi(t, v_{b,n}')}{\epsilon_l} = \frac{\partial^+\varphi(t,v_{b,n}')}{\partial t} = \frac{\partial^+\varphi(t,v_b)}{\partial t}.$$

Then, by the Cauchy diagonalization theorem,

$$\lim_{n\to\infty} \frac{\varphi(t+\epsilon_n, v_{b,n}') - \varphi(t, v_{b,n}')}{\epsilon_n} = \frac{\partial^+\varphi(t,v_b)}{\partial t}. \tag{1.9}$$

Next, since $\varphi(t,t) = 0$, and also applying (1.8), we have for ϵ_n sufficiently small (i.e.,

n large enough),

$$\varphi(t, v'_{b,n}) = \varphi(t, t + \epsilon_n) = \varphi(t,t) + \frac{\partial^+ \varphi(t,t)}{\partial v_b}\epsilon_n + O(\epsilon^2) = \frac{\partial^+ \varphi(v_s, t)}{\partial v_b}\epsilon_n + O(\epsilon_n^2).$$

Note that $\frac{\partial^+ \varphi(t,t)}{\partial v_b}$ is understood as $\lim_{v_b \to t, v_b > t} \frac{\partial^+ \varphi(t,v_b)}{\partial v_b}$. We insert this into (1.9), also noting that $\varphi(t+\epsilon_n, v'_{b,n}) = \varphi(t+\epsilon_n, t+\epsilon_n) = 0$, to obtain

$$\frac{\partial^+ \varphi(t, v_b)}{\partial t} = \lim_{n\to\infty} \frac{\varphi(t+\epsilon_n, v'_{b,n}) - \varphi(t, v'_{b,n})}{\epsilon_n} = \lim_{n\to\infty} \frac{-\frac{\partial^+ \varphi(v_s,t)}{\partial v_b}\epsilon_n + O(\epsilon_n^2)}{\epsilon_n} = -\frac{\partial^+ \varphi(v_s, t)}{\partial v_b}.$$

Thus we have shown that at every $t \in (v_s, v_b)$, $\frac{\partial^+ \varphi(t,v_b)}{\partial t} = -\frac{\partial^+ \varphi(v_s,t)}{\partial v_b}$, which implies that $\frac{\partial^+ \phi(v_s,v_b,t)}{\partial t}$ exists and is equal to 0. Similarly, we show that $\frac{\partial^- \phi(v_s,v_b,t)}{\partial t}$ exists and is equal to 0, which proves that $\phi(v_b, v_b, t)$ is differentiable w.r.t.t. This concludes the proof of step 1.2. and case 1.

Case 2. We complete the proof by showing that $\varphi(v_s, v_b)$ can only be discontinuous in a way which still admits a representation by some $\tilde{\varphi}(.)$. In particular, we show that there exists a step function $\underline{\varphi} : [0,1] \times [0,1] \to [0,1]$ s.t. $\varphi(v_s, v_b) - \underline{\varphi}(v_s, v_b)$ is continuous, and $\underline{\varphi}(v_s, v_b) = \underline{\tilde{\varphi}}(v_b) - \underline{\tilde{\varphi}}(v_s)$, for some step function $\underline{\tilde{\varphi}} : [0,1] \to [0,1]$. We proceed in 2 steps, both involve applying the Monotone Convergence Theorem (MCT), and some tedious calculus.

Step 2.1. If $\exists\ v_s \in [0,1]$, and $\bar{\tau} > v_s$ s.t. $\varphi(v_s, \bar{\tau}+) - \varphi(v_s, \bar{\tau}-) = \Delta_s(v_s, \bar{\tau}) > 0$, then $\varphi(v'_b, \bar{\tau}+) - \varphi(v'_s, \bar{\tau}-) = \Delta_s(v_s, \bar{\tau}) > 0$, $\forall\ v'_s < \bar{\tau}$.

Proof. We write

$$\varphi(v_s, \bar{\tau}+) = \lim_{\epsilon\to 0} \frac{1}{\bar{\tau}+\epsilon - v_s} \int_{v_s}^{\bar{\tau}+\epsilon} \varphi(v_s, \tau) + \varphi(\tau, \bar{\tau}+\epsilon) d\tau,$$

$$\varphi(v_s, \bar{\tau}-) = \lim_{\epsilon\to 0} \frac{1}{\bar{\tau}-\epsilon - v_s} \int_{v_s}^{\bar{\tau}-\epsilon} \varphi(v_s, \tau) + \varphi(\tau, \bar{\tau}-\epsilon) d\tau,$$

and since

$$\lim_{\epsilon\to 0}\frac{1}{\bar{\tau}+\epsilon-v_s}=\lim_{\epsilon\to 0}\frac{1}{\bar{\tau}-\epsilon-v_s}=\frac{1}{\bar{\tau}-v_s},$$

we have

$$\Delta_s(v_s,\bar{\tau})=\frac{1}{\bar{\tau}-v_s}\left[\lim_{\epsilon\to 0}\int_{\bar{\tau}-\epsilon}^{\bar{\tau}+\epsilon}\varphi(v_s,\tau)d\tau+\lim_{\epsilon\to 0}\int_{v_s}^{\bar{\tau}+\epsilon}\varphi(\tau,\bar{\tau}+\epsilon)d\tau-\int_{v_s}^{\bar{\tau}-\epsilon}\varphi(\tau,\bar{\tau}-\epsilon)d\tau\right]. \tag{1.10}$$

Now

$$\lim_{\epsilon\to 0}\int_{\bar{\tau}-\epsilon}^{\bar{\tau}+\epsilon}\varphi(v_s,\tau)d\tau=\lim_{\epsilon\to 0}\int_{v_s}^{1}1_{(\bar{\tau}-\epsilon,\bar{\tau}+\epsilon)}\varphi(v_s,\tau)d\tau=0,$$

by the (MCT). Similarly, we apply the (MCT) to the other part of (1.10), so that

$$\lim_{\epsilon\to 0}\int_{v_s}^{\bar{\tau}+\epsilon}\varphi(\tau,\bar{\tau}+\epsilon)d\tau-\int_{v_s}^{\bar{\tau}-\epsilon}\varphi(\tau,\bar{\tau}-\epsilon)d\tau=\lim_{\epsilon\to 0}\int_{v_s}^{1}1_{(v_s,\bar{\tau}+\epsilon)}\varphi(\tau,\bar{\tau}+\epsilon)-1_{(v_s,\bar{\tau}+\epsilon)}\varphi(\tau,\bar{\tau}+\epsilon)d\tau$$

$$=\int_{[v_s,\bar{\tau})}\varphi(\tau,\bar{\tau}+)-\varphi(\tau,\bar{\tau}-)d\tau.$$

Therefore,

$$\Delta_s(v_s,\bar{\tau})=\frac{1}{\bar{\tau}-v_s}\int_{[v_s,\bar{\tau})}\varphi(\tau,\bar{\tau}+)-\varphi(\tau,\bar{\tau}-)d\tau=\frac{1}{\bar{\tau}-v_s}\int_{[v_s,\bar{\tau})}\Delta_s(\tau,\bar{\tau})d\tau. \tag{1.11}$$

The claim now follows for $v_s<\bar{v_s}<\bar{\tau}$, by Lemma 1.4.2. This concludes the proof of step 2.1.

Step 2.2. If $\exists\ v_s\in[0,1]$, and $\bar{\tau}>v_s$ s.t. $\varphi(v_s,\bar{\tau}+)-\varphi(v_s,\bar{\tau}-)=\Delta>0$, then $\exists\ v_b>\bar{\tau}$ s.t. $\varphi(\bar{\tau}-,v_b)-\varphi(\bar{\tau}+,v_b)=\Delta$.

Proof. Since $\varphi(0,\tau)$ is bounded and monotonic, there exists a $\bar{v}_b$ s.t. $\varphi(0,\tau)$ is continuous for $\tau\in(\bar{\tau},\bar{v}_b]$. By step 2.1, $\varphi(v_s,\tau)$ is continuous for $\tau\in(\bar{\tau},\bar{v}_2],\forall\ v_s<\bar{v}_b$. We can proceed as in step 2.1 to obtain for each v_b,

$$\Delta_b(v_b,\bar{\tau})=\frac{1}{v_b-\bar{\tau}}\left[\lim_{\epsilon\to 0}\int_{\tau-\epsilon}^{v_b}\varphi(\bar{\tau}-\epsilon,\tau)d\tau-\int_{\tau+\epsilon}^{v_b}\varphi(\bar{\tau}+\epsilon,\tau)d\tau\right].$$

Next,

$$\lim_{\epsilon\to 0}\int_{\tau-\epsilon}^{v_b}\varphi(\bar\tau-\epsilon,\tau)d\tau-\int_{\tau+\epsilon}^{v_b}\varphi(\bar\tau+\epsilon,\tau)d\tau=\lim_{\epsilon\to 0}\int_{\tau+\epsilon}^{v_b}\varphi(\bar\tau-\epsilon,\tau)-\varphi(\bar\tau+\epsilon,\tau)d\tau+\int_{\tau-\epsilon}^{\tau+\epsilon}\varphi(\bar\tau-\epsilon,\tau)d\tau$$

$$=\lim_{\epsilon\to 0}\int_{\tau+\epsilon}^{v_b}\varphi(\bar\tau-\epsilon,\tau)-\varphi(\bar\tau+\epsilon,\tau)d\tau=\int_{(\tau,v_b]}\lim_{\epsilon\to 0}\varphi(\bar\tau-\epsilon,\tau)-\varphi(\bar\tau+\epsilon,\tau)d\tau,$$

where the second equality follows by MCT, and the third one by the bounded convergence theorem. Thus, for every v_b,

$$\Delta_b(v_b,\bar\tau)=\frac{1}{v_b-\bar\tau}\int_{(\tau,v_b]}\lim_{\epsilon\to 0}\varphi(\bar\tau-\epsilon,\tau)-\varphi(\bar\tau+\epsilon,\tau).$$

For each $k=1,...,\infty$, by continuity and monotonicity of $\varphi(\bar\tau+\frac{1}{k},\tau)$, and since $\varphi(\bar\tau+\frac{1}{k},\bar\tau+\frac{1}{k})=0$, there exists a $v_b^{(k)}>\bar\tau+\frac{1}{k}$, s.t. $\varphi(\bar\tau+\frac{1}{k},v_b^{(k)})<\frac{1}{k}$. On the other hand, $\varphi(\bar\tau-\frac{1}{k},v_b^{(k)})\geq\Delta$, so that

$$\Delta_b(v_b^{(k)},\bar\tau)>\Delta-\frac{1}{k},$$

which by step 2.1 implies that $\Delta_b(v_b,\bar\tau)\geq\Delta$. By a symmetric argument, it must be that $\Delta\geq\Delta_b(v_2,\bar\tau)$. This concludes the proof of step 2.2.

Now we wrap up the proof of the Theorem. Define $\underline{\tilde\varphi}(x)$ by the Lebesgue integral

$$\underline{\tilde\varphi}(x)=\int_0^x\Delta_b(0,y)dy.$$

By steps 2.1. and 2.2., $\varphi(v_s,v_b)-\big(\underline{\tilde\varphi}(v_b)-\underline{\tilde\varphi}(v_s)\big)$ is continuous, and we apply step 1 to conclude the proof of the Theorem. $\square$

Bibliography

[1] BERGEMANN, D. AND S. MORRIS [2005]: “Ex-post Implementation: The Role of Large Type Spaces”, Cowles Foundation Discussion Paper No. 1519.

[2] CHUNG, K-S. AND J. ELY [2005]: ”Foundations of Dominant Strategy Mechanisms”, mimeo.

[3] COPIC, J. AND C. PONSATI [2006]: “Optimal Robust Bargaining”, mimeo, www.hss.caltech.edu/∼jernej.

[4] CREMER, J. AND R. P. MCLEAN [1985]: “Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent”, *Econometrica*, **53**, 345-362.

[5] CREMER, J. AND R. P. MCLEAN [1988]: “Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions”, *Econometrica*, **56**, 1247-1257.

[6] HAGERTY, K. M. AND W. P. ROGERSON [1987]: “Robust Trading Mechanisms”, *Journal of Economic Theory*, **42**, 94-107.

[7] HOLMSTROM, B. AND R. B. MYERSON [1983]: “Efficient and Durable Decision Rules with Incomplete Information”, *Econometrica*, **51**, 1799-1819.

[8] JEHIEL, P., M. MEYER-TER-VEHN, B. MOLDOVANU AND W. R. ZAME [2005]: ”The Limits of Ex-Post Implementation”, UCLA Economics Working Papers 839.

[9] LEDYARD, J. O. [1978]: “ Incentive Compatibility and Incomplete Information”, *Journal of Economic Theory*, **18**, 171-189.

[10] MAILATH, G., AND A. POSTLEWAITE [1990]: "Asymmetric-information bargaining problems with many agents", *Review of Economic Studies*, **57**, 351-368.

[11] MCAFEE, P. R. AND P. RENY [1992]: "Correlated Information and Mechanism Design", *Econometrica*, **60**, 395-422.

[12] MYERSON, R. B., M. A. SATTERTHWAITE [1983]: "Efficient Mechanism for Bilateral Trading", *Journal of Economic Theory*, **29**, 265-281.

[13] WILSON, R. [1987]: "Game-Theoretic Analyses of Trading Processes," in *Advances in Economic Theory:Fifth World Congress*, ed. T. Bewley, Cambridge: Cambridge University Press, 33-70.

Chapter 2

Optimal Robust Bargaining Games

The work presented in this chapter has been done jointly with Clara Ponsatí.

2.1 Introduction

Bargaining between two impatient traders is a fundamental problem of economics. Since Rubinstein's [1982] result on the alternating-offers game with perfect information, many economists have been concerned with providing a similarly effective and tractable framework for settings with imperfect or incomplete information where agents' reservation values are private information. By the Myerson and Satterthwaite [1983] impossibility theorem, the outcomes are then necessarily inefficient, which is different from the perfect-information setup.

Most of the incomplete-information bargaining literature has focused on characterizing Bayesian equilibria of different versions of the alternating-offers game, with risk-neutral agents. This approach has three shortcomings. Such games have many Bayesian equilibria, which are hard to characterize.[1] In Bayesian approaches, agents are assumed to have precise knowledge of the distribution of each others' reservation values, or types. The effect of risk aversion on equilibrium contracts may matter, and in many relevant situations agents have different levels of risk aversion. Examples of such situations are wage bargaining between a worker and a firm, a large block-trade between a market maker and an investor, a real-estate trade, and bilateral peace

[1] See Cramton [1984], Cho [1990], and Ausubel and Denekere [1992].

talks. In the absence of adequate mechanism-design machinery, such contracts have been studied in the literature, but generally without imposing incentive constraints.

In this paper, we study robust equilibria for bargaining games in environments where agents are impatient and can be risk averse. We assume that the players' utility functions are common knowledge but that their reservation values are private. By robustness we mean that both the equilibrium concept and its efficiency are robust to traders' beliefs. We thus require an equilibrium of such a game to be an *ex-post* equilibrium, and that its outcome be no worse, in the Paretian sense, than any *ex-post* equilibrium outcome of any bargaining game. We call this equilibrium efficiency requirement *ex-post* constrained efficiency.

We define Mediated Bargaining Game as a continuous-time double auction in which the Mediator prevents traders from seeing each other's bids until the time of agreement.[2] The agreed price is then made public, and the trade takes place. The key feature of Mediated Bargaining Game is that the information flow between the agents is minimized, so that agents recognize the surplus only upon agreement, when the game is over. The main result of our analysis is that Mediated Bargaining Game is the optimal robust bargaining game.

We characterize regular equilibria of Mediated Bargaining Game. A regular equilibrium is an *ex-post* Nash equilibrium in undominated and type-monotone strategies. We show that the set of outcomes of Mediated Equilibria under risk neutrality is dense in the set of outcomes of *ex-post* constrained-efficient equilibria of *all* dynamic bargaining games. Due to incomplete information, delay arises endogenously as a part of every equilibrium, and delay causes the inefficiency. All equilibrium outcomes of Mediated Game are *ex-post* individually rational to both traders. Under risk neutrality all Mediated Equilibria have simple closed-form expressions, but this is true only for special cases of risk aversion. Nonetheless, when agents are risk averse or discount the future at different rates, regular equilibria are constrained efficient. For a rich set of environments with risk aversion, there exists a unique linear equilibrium with

[2] A version of Mediated Bargaining Game where the set of possible prices is finite was first proposed by Jarque, Ponsatí, and Sákovics [2003]. In our model the prices are not restricted.

a simple closed form, suitable for embedding in other models.

Under risk aversion there is risk sharing in equilibrium, and if an agent becomes more risk averse, the outcomes become closer to full efficiency. The effect on equilibrium outcomes is the same if agents are more impatient or if they are more risk averse, i.e., an external observer cannot distinguish between a more impatient and a more risk-averse agent just by observing an equilibrium outcome. Under perfect information, this is also a feature of the subgame-perfect Nash equilibrium in the Rubinstein alternating-offers game.

The Mediated Bargaining Game is a decentralized game form implementing *ex-post* individually rational, incentive compatible, and constrained-efficient direct revelation mechanisms. Ledyard [1978] proves that if a game has an *ex-post* Nash equilibrium then the corresponding direct-revelation mechanism is *ex-post* incentive compatible.[3] *Ex-post* constrained efficiency of *ex-post* equilibria is equivalent to the *ex-post* constrained efficiency of an *ex-post* individually rational and incentive compatible direct revelation mechanism. Hence, for each regular equilibrium, there is a direct revelation mechanism that is *ex-post* constrained efficient under *ex-post* incentive compatibility and individual rationality. Under risk neutrality, such mechanisms can be represented as probability distributions over posted prices (Theorem 1.3.1). Since Mediated Bargaining Game implements *all* constrained-efficient mechanisms and nothing else, it is *the* optimal robust bargaining game.

As a model, direct revelation mechanisms are not equivalent to the indirect Mediated Bargaining Game. First, to construct a specific *ex-post* individually-rational and incentive-compatible mechanism, the designer has to know the agents' utility functions, but neither rationality nor preferences have to be common knowledge among the agents. In Mediated Bargaining Game, the designer does not need to know anything about the agents, and just lets them play the game, but the preferences, rationality, and the equilibrium have to be common knowledge between the two traders. The second important difference is that Mediated Bargaining Game is free of a commitment

[3] A regular equilibrium *is not* a dominant-strategy equilibrium of Mediated Bargaining Game, so that the relation to direct revelation mechanisms is not simply a consequence of the revelation principle. In fact, Mediated Bargaining Game has no equilibria in dominant strategies.

problems that affect the direct revelation mechanism. Generically, direct revelation mechanisms prescribe lotteries at the *ex-post* stage, and the agents can invert each lottery to figure out the type of the opposing player that they are facing. The assumption in a direct revelation mechanism is that agents can commit to *ex-post* lotteries even though they can obtain something better once they know each other's types. But *ex-post* individual rationality requires that they cannot commit to *ex-post* payments, which will make them better off at the *interim* stage. Thus, *ex-post* individual rationality is inconsistent with commitment to *ex-post* lotteries. Such criticism does not apply to regular equilibria because they specify deterministic trade every time it occurs, and time cannot be reversed to ameliorate the inefficiency resulting from delays.

As a final remark, we show that the set of regular equilibria coincides with the set of separating perfect Bayesian equilibria of Mediated Bargaining Game. Perfect Bayesian equilibria are outcome equivalent to Bayesian equilibria because out-of-equilibrium deviations cannot affect the updating of beliefs. Thus, imposing a weaker equilibrium notion and additional sequential rationality on the players does not change the set of equilibrium outcomes of Mediated Bargaining Game, so that this set is robust to weaker equilibrium notions.

Mediated Bargaining Game is the optimal robust bargaining game, and is observed in practice. One can interpret the Mediator as an order book that is closed. Several electronic exchanges, e.g. Nasdaq, Frankfurt, Stockholm, and others, allow for *hidden orders* which are put in the book but are not observable by other traders. The justification is that hidden orders are supposed to enhance efficiency, and to our knowledge there is no theoretical foundation in the existing literature. Albeit a very stylized model of exchange, our analysis of Mediated Bargaining Game provides a strong theoretical support for that claim. Mediation is also widely used in conflict resolution, and practitioners point to the fact that effective mediation requires restricting direct information flows between the two parties.[4] That is precisely the

[4]For example, Francesc Vendrell-the UN envoy to Central America, Namibia, Timor and Afganistan-says: "I prefer to negotiate separately with each party, rather than with both parties talking face to face." (El País 30/12/01). See also Dunlop [1984].

defining feature of Mediated Bargaining Game.

In Section 2.2 we review the literature. In Section 2.3 we give definitions of robust bargaining games and *ex-post* constrained efficiency. In Section 2.4 we review direct mechanisms under risk neutrality. In Section 2.5 we analyze Mediated Bargaining Game under risk neutrality. In Section 2.6 we extend the results to risk aversion, and prove the existence of the unique linear regular equilibrium. In Section 2.7, we show that Bayesian equilibria of Mediated Bargaining Game are *ex-post* equilibria.

2.2 Related literature

Our work is related to the literature on robust mechanism design, see Hurwicz [1972], Ledyard [1978], D'Aspremont and Gerard-Varet [1979], Neeman [2004], Chung and Ely [2003], Bergemann and Morris [2004] and Jehiel et al. [2005]. The recent literature studies *ex-post* implementation in common value problems, where *ex-post* implementation does not imply dominant-strategy direct mechanisms, which is different from our setting. Čopič and Ponsatí [2005] characterize *ex-post* individually-rational, incentive-compatible and constrained-efficient mechanisms for bilateral trade with risk aversion and generalize the results of Hagerty and Rogerson [1987]. The latter establish payoff equivalence to distributions over posted prices for a subclass of *ex-post* individually-rational and incentive-compatible mechanisms under risk neutrality.

Our work is also related to the literature on non-cooperative bargaining under incomplete information (see Ausubel, Cramton, and Denekere [2002] for a survey). The closest are Cramton [1992] and Wang [2000]. Cramton [1992] extends the continuous-time game of Admati and Perry [1987] to two-sided uncertainty and constructs a separating equilibrium where trade occurs, with delay, whenever gains from trade exist. In the game of Wang [2000], each side chooses exactly one time at which they are willing to concede to the opponent. In equilibrium types are revealed by the choice of time, and the price is a linear combination of types. In that game, there exists a class of outcome-equivalent separating *ex-post* equilibria. The outcome coincides with that of the linear regular equilibrium under risk neutrality. In Example 2.5.8 we compare the efficiency (in *ex-ante* terms) of regular equilibria with the equilibrium in Wang [2000] and Cramton [1992]. We show that regular equilibria can dominate both of these, although the latter is not robust.

Comparison to cooperative bargaining is also relevant. The allocation in the linear regular equilibrium coincides with the Nash bargaining solution only when agents have the same risk aversion. Still, in asymmetric environments the risk sharing in the Nash bargaining solution goes in the same direction as in our model. But the approach of cooperative bargaining theory is nevertheless quite different. There is no incomplete information, incentives are not modeled explicitly, and outcomes are assumed to be efficient, so that there is no obvious way to model delay.

2.3 Dynamic bargaining games and robustness

The problem, preferences, and information structure. Two traders, a seller and a buyer $i = s, b$, bargain over the price $p \in [0, 1]$ of an indivisible good. The seller's cost of producing the good v_s, and the buyer's valuation of the good v_b are private information. We denote $v = (v_s, v_b)$. We assume that it is common knowledge that v is distributed according to *some* continuous G with a continuos density g, and support $supp(g) = [0, 1]^2$. We stress that common knowledge of the specific G is not necessary-it is necessary that the support of types is $[0, 1]^2$ and that is common knowledge.

We assume that the agents are risk neutral, and they discount the future exponentially. When an agreement to trade at price p is reached on date $t \geq 0$, the seller's payoff upon trading at price p at t is $u_s(v_s, p, t) = e^{-t}(p - v_s)$ and the buyer's is $u_b(v_b, p, t) = e^{-t}(v_b - p)$. In Section 2.6 we relax these assumptions.

Robust equilibrium and efficiency requirements. Given some dynamic bargaining game form Γ, we impose that the equilibrium and efficiency notions be robust. Therefore, strategies and outcomes must be independent of beliefs, implying that the equilibrium be an *ex-post* equilibrium. We define the *ex-post* equilibrium and the robustness notion for a general dynamic bargaining game Γ, so that our definitions are a bit loose. In this section we say nothing about the existence of such games. In Section 5 we construct a dynamic bargaining game with robust equilibria that are optimal.

A dynamic bargaining game Γ is in our setup defined by the sets of traders' strategies, contingent on their type, the set of histories for each player, given past play of the game, and the outcome function, mapping strategy profiles into outcomes (i.e., terminal histories). Let $\mathcal{H}_{i,t}$ be the set of possible histories at time t for player i. A strategy of player i is a mapping from his type v_i, time t, and history $h_i(t) \in \mathcal{H}_{i,t}$ into price bids,

$$p_i : [0,1] \times [0,\infty) \times \mathcal{H}_{i,t} \longrightarrow [0,1] \ , \ i = s, b.$$

Each outcome is specified by a time $\tau(p_s, p_b, v_s, v_b)$ and a price at which trade occurs at that time, $\bar{p}(p_s, p_b, v_s, v_b)$. If the trade never happens that is equivalent to $\tau(p_s, p_b, v_s, v_b) = \infty$.[5]

We say that strategies (p_s^*, p_b^*) constitute an *ex-post Nash equilibrium* (XPEQ) if they are mutual best responses for each pair of types (v_s, v_b). More precisely, given the equilibrium strategy of the buyer, p_b^*, the seller's strategy p_s^* satisfies:

$$p_s^* = \arg\max_{p_s} e^{-\tau(p_s, p_b^*, v_s, v_b)} \left(\bar{p}(p_s, p_b^*, v_s, v_b) - v_s\right), \forall \ v_s, v_b.$$

Denote $p^* = (p_s^*, p_b^*)$, and by $U_i(v; p^*)$ the equilibrium payoff to agent i, given strategies p^* and types v.

Note that every XPEQ is also a Bayes-Nash equilibrium, but not necessarily the converse.

In a dynamic bargaining game Γ, it can never be an XPEQ for a trader to trade at a price which gives him a negative utility, so that every XPEQ outcome must be individually rational to both agents.

[5]For an extensive discussion of when games in continuous time are well defined see Simon and Stinchcombe [1989]. For a discussion on admissible strategies and sensible outcomes in bargaining games with continuous time see Sákovics [1993].

The robust efficiency notion we impose is *ex-post* constrained efficiency,[6] which in the present context says the following. Take a dynamic bargaining game Γ and an XPEQ $(p^*; \Gamma)$. We say that this XPEQ is *ex-post* constrained efficient (XPCE) if there does not exist another pair $(\tilde{p}^*; \Gamma')$, Γ' a dynamic bargaining game and $\tilde{p}^*$ an XPEQ of Γ', such that

$$U_i(v; p^*) \leq U_i(v; \tilde{p}^*), \forall\ v \in [0,1]^2, i = s, b, \text{ and}$$

$$U_i(v; p^*) < U_i(v; \tilde{p}^*), \forall\ v \in V^{open} \subset [0,1]^2, \text{ for at least one } i.$$

One could impose additional sequential rationality requirements by requiring that the XPEQ be a subgame-perfect Bayes-Nash equilibrium (PBE) of Γ, for every prior G. However, such requirement would not have any bite-all XPEQ satisfy this property, which is easy to verify. In general PBE need not be robust.

A stronger requirement is that for an XPEQ profile p^*, XPCE is satisfied at every t, given what is common knowledge at t. Denote by $\mathcal{B}_0$ the initial bargaining problem as specified at the beginning of this section (i.e., agents' types, utility functions, the initial information structure, etc.). Common knowledge at t defines the continuation bargaining game $\Gamma_{|t}$ and also the *continuation bargaining problem*, $\mathcal{B}_{|t}$. Also denote by $p^*_{|t}$ the continuation of the play of the traders' strategies from time t on, given the profile p^*.[7]

We say p^*, an XPEQ of Γ satisfies *dynamic ex-post constrained efficiency* (DXPCE),

[6]This notion is related to *ex-post* incentive efficiency of direct revelation mechanisms, due to Holmstrom and Myerson [1983]. The difference is that *ex-post* incentive efficiency means *ex-post* constrained optimality of a mechanism given Bayesian incentive compatibility. Thus, it does not employ individual rationality, and incentive compatibility is imposed at the *interim* while XPEQ is equivalent to XPIC of the direct revelation mechanism. We use a different name in order to keep the distinction clear.

[7]The play of the agents until time t defines the histories for each player, which determine the information sets for each player. These describe what is common knowledge, and what is private information to each trader. Private information to each trader will in this case be *rich*-different types might have different knowledge, and might also attach different knowledge to different types of the other player, and so on. However, if one insists on the *ex-post* equilibrium, what matters is the payoff-relevant type of each trader. Moreover, in the present private-values environment, the fact that there is no common implies that *ex-post* equilibrium is the weakest possible equilibrium concept. See Bergemann and Morris [2005] for a comprehensive discussion of *ex-post* implementation on rich type spaces.

if $p^*_{|t}$ in $\Gamma_{|t}$ is XPCE for the problem $\mathcal{B}_{|t}$ for all $t \geq 0$.

DXPCE can be interpreted as a strong renegotiation-proofness requirement: at no time would the traders want to shift to playing another *ex-post* equilibrium of some other bargaining game.

To determine the bounds on possible XPCE allocations (and consequently on DXPCE allocations) we can consider direct revelation mechanisms. By the revelation principle and Ledyard [1978], for each dynamic bargaining game Γ, and each *ex-post* equilibrium (p^*, Γ), there exists a direct revelation mechanism (mechanism) m, which is *ex-post* incentive compatible, and such that $U_i(v; p^*) = U_i^m(v), \forall\ v \in [0,1], i = s, b$, where $U_i^m(v)$ is the payoff to agent i in mechanism m, under truthful reporting.

2.4 Direct revelation mechanisms

In this section we briefly review the results on *ex-post* incentive compatible (XPIC), *ex-post* individually rational (XPIR), and *ex-post* constrained-efficient (XPCE) mechanisms (under XPIR and XPIC). Note that XPCE is imposed at the *ex-post* stage, and should not confuse this with the *ex-ante* or *interim* notions of optimality. Note also that *ex-post* incentive compatibility is equivalent to equilibrium in dominant strategies and should not confuse that with the *interim* incentive compatibility which is equivalent to Bayesian equilibrium. We also remark that XPCE is weaker than either the *ex-ante* or *interim* constrained-efficiency notions (XPCE is necessary for either of these two). XPCE is the only notion among the three which is robust. For a more detailed discussion of the issues reviewed here see chapter 1.

We first note the well known fact that under risk neutrality, each XPIRIC mechanism can be represented by a pair of functions $(\pi, \delta) : [0,1]^2 \to [0,1]^2$, where $\pi(v)$ is the price and $\delta(v)$ is the probability of trade at that price; with complementary probability no trade occurs, and the agents obtain 0 utility. We remark that in an XPEQ (p^*, Γ), $\delta(v)$ corresponds to the shrinking of the surplus due to discounting, so that $\delta(v) = e^{-\tau(p^*(v))}$. Denote by $U_s^{\pi,\delta}(v_s', v_b; v_s) = \delta(v_s', v_b)(\pi(v_s', v_b) - v_s)$ the payoff to the seller under a mechanism (π, δ) when the reported types are (v_s', v_b) and seller's true type is v_s. Similarly for the buyer, $U_b^{\pi,\delta}(v_s', v_b; v_b) = \delta(v_s', v_b)(v_b - \pi(v_s', v_b))$. Also denote $U_i^{\pi,\delta}(v) = U_i^{\pi,\delta}(v_i, v_j; v_i), i, j \in \{s, b\}, i \neq j$. XPIC and XPIR of a mechanism $m = (\pi, \delta)$ are now formulated as follows:

$$U_s^{\pi,\delta}(v_i', v_j; v_i) \leq U_i^{\pi,\delta}(v), \forall\ v_i, v_i', v_j, i, j \in \{s, b\}, i \neq j; \text{ (XPIC)}$$

$$\delta(v) > 0 \Rightarrow v_s \leq \pi(v) \leq v_b, \forall\ v. \text{ (XPIR)}$$

The *ex-post* constrained efficiency under individual rationality (XPCE) of mechanisms is formulated similarly as the XPCE of a dynamic game. An XPIRIC mechanism (π, δ) satisfies XPCE if

$$\nexists\ (\pi', \delta'),\ \text{XPIRIC and s.t.} U_i^{\pi,\delta}(v) \leq U_i^{\pi',\delta'}(v), \forall\ v, i = s, b,\ \text{and}$$

$$U_i^{\pi,\delta}(v) < U_i^{\pi',\delta'}(v), \forall\ v \in V_0^{open} \subset [0, 1]^2,\ \text{for at least one } i.$$

We recall Theorem 1.3.1 from chapter 1.

Theorem 2.4.1. *A mechanism (π, δ) is XPCE if and only if there exists a probability distribution F_p, $supp(F_p) \subset [0, 1]$, such that $\delta(v) = F_p(v_b) - F_p(v_s)$ and $\pi(v) = E_{F_p}[p \mid v_s \leq p \leq v_b]$. Here $E_{F_p}[. \mid .]$ denotes the conditional expectation w.r.t. F_p.*

2.5 Mediated Bargaining Game

In this section we introduce Mediated Bargaining Game (MBG), we define regular XPEQ of MBG, and we show that all of these are XPCE. Thus, MBG is an optimal robust bargaining game.

The game. MBG is a dynamic double auction in continuous time, which can heuristically be described as a game with a mediator. The mediator is a dummy player whose only role is to receive bids, keep them secret while they are incompatible, and to announce the agreement as soon as it is reached. When the mediator announces that an agreement has been reached, trade takes place at the agreed price, and the game ends. In MBG, the mediator imposes a restriction on the agents' updating of beliefs. In particular, the agents can only update through the passing of time, and observable history is for each player at time t completely specified by t.

In MBG, a strategy of player i is a function $p_i(.)$, mapping i's types and time into bids, $p_i : [0,1] \times [0,\infty) \to [0,1]$, $i = s, b$. In order for outcomes to be well defined, it is necessary to require that p_i be left continuous w.r.t. time t, for each v_i (see Simon and Stinchcombe [1989]).

Outcomes of MBG are given by terminal histories. Given a profile of strategies p and a draw of valuations v, MBG ends at $t^* < \infty$ if $t^* = \min\{t \in [0,\infty) \mid p_s(t,v_s) \leq p_b(t,v_b)\}$. At t^* traders trade at a pre-specified $\bar{p} \in [p_s(t,v_s), p_b(t,v_b)]$, e.g. $\bar{p} = \frac{1}{2}\left(p_s(t,v_s) + p_b(t,v_b)\right)$.

We consider XPEQ of MBG. We restrict attention to *regular* XPEQ of MBG, which we define next.

First, observe that for each XPEQ profile p, a profile p' constructed by adding a *standstill interval* $[0,T)$, i.e., $p'_i\left(v_i, t+T\right) = p_i\left(v_i, t\right)$, is an XPEQ as well, for every $T < \infty$. That is, as the opponent does not concede any positive amount until T, no concession prior to T is useful. For $T < \infty$, such strategy profiles p' are weakly dominated. We say that an XPEQ is *undominated* if it does not have a standstill interval. We can similarly define undominated profiles under Bayesian Equilibrium (BE) and perfect Bayesian Equilibrium (PBE) concepts (see Section 6 for precise definitions).

A *Regular ex-post equilibrium* (REQ) is an undominated XPEQ such that p_i are differentiable w.r.t. t and v_i, strictly type monotone, p_s is strictly decreasing w.r.t. t, p_b is strictly increasing w.r.t. t, and such that if $v_b > v_v$, then $\exists\ t < \infty$ such that $p^*_s(v_s,t) = p^*_s(v_s,t)$.

In the Appendix we show that every Bayesian equilibrium (BE) of MBG has to be weakly type monotone (see Proposition 2.7.2). Since every XPEQ is a BE, this implies that all XPEQ have to be weakly type monotone. In the rest of this section we characterize the REQ of MBG and show that they exist. We also show that the set of outcomes of REQ is dense in the set of outcomes of XPEQ of MBG, so that the restriction to REQ is made purely for analytical convenience. First, we show that if a REQ of MBG satisfies XPCE, then it satisfies DXPCE.

Proposition 2.5.1. *Suppose p^* is an REQ profile of MBG. Then p^* satisfies DXPCE.*

Proof. Take a $v \in [0,1]^2$ and $t < t^*(v;p^*)$. Then at t, it is common knowledge only that $v \in [0,1]^2$. Take for instance the seller, who at t knows that v_b is such that $p_b^*(v_b,t) < p_s^*(v_s,t)$. Thus, the seller knows that $v_b \in [0, \bar{v}_b(v_s))$, by type monotonicity of the buyer's strategy, where $p_b^*(\bar{v}_b(v_s),t) = p_s^*(v_s,t)$. The seller also knows that the buyer's type $v_b = 0$ only knows at t that $v_s \in [0,1]$. Similarly, the buyer at t knows that seller may be of a type $v_s = 1$ which at t only knows that $v_b \in [0,1]$. Thus, at t it is common knowledge only that $v \in [0,1]^2$. Hence, the bargaining problem $\mathcal{B}_t$ at t is the same as at time 0, and since the profile p^* satisfies XPCE, it satisfies XPCE at each t. □

Next, we characterize the REQ of MBG.

Proposition 2.5.2. *Let p be a strategy profile such that both p_s and p_b are both strictly type monotone and differentiable w.r.t. t and and types, and such that $p_s(v_s,t)$ and $p_b(v_b,t)$ cross at $t < \infty$ iff $v_s < v_b$. Then, p is a REQ if and only if*

1. *$p_i(v_i,t)$, $i = s,b$ satisfy the first-order conditions*

$$\begin{aligned} (p_s(v_s,t) - v_s) &= \tfrac{\partial p_b(v_b,t)}{\partial t}, \\ (v_b - p_b(v_b,t)) &= -\tfrac{\partial p_s(v_s,t)}{\partial t}; \end{aligned} \tag{2.1}$$

   $\forall\ v,t,\ \ s.t. p_s(v_s,t) = p_b(v_b,t);$

2. *$p_s(0,0) = p_b(1,0)$.*

Proof. Let p^* be a REQ profile, and take a $v \in [0,1]^2$. We have to verify that if strategies p^* are differentiable, strictly time-monotone, and are best replies, then they satisfy the above first-order condition. We do that for the seller, a mirror argument works for the buyer. In an XPEQ, it is clear that if a pair of agents with types v agree at time t, then they agree with equality, i.e.,

$$p_s^*(v_s, t) = p_b^*(v_b, t), \tag{2.2}$$

Otherwise either one of the agents could profitably deviate against the given type of the opponent-to obtain all of the difference between the proposed prices. From equation (2.2), we can then define by the implicit function theorem, $v_s = v_s(v_b, t)$, and we have $\frac{\partial p_s}{\partial v_s}\frac{\partial v_s}{\partial v_b} = \frac{\partial p_b}{\partial v_b}$.

Next, given p_b^*, again by (2.2), the seller maximizes

$$\max_{t \in [0,\infty)} e^{-t}\left(p_b^*(v_b, t) - v_s\right),$$

which implies the first-order condition (FOC). It is also easy to check that the second derivative of the objective function is negative so that the FOC is indeed necessary and sufficient. The condition $p_s(0,0) = p_b(1,0)$ follows from (2.2) and strict type monotonicity. □

Theorem 2.5.3. *If a strategy profile p^* is a REQ of MBG then the corresponding mechanism is a lottery F_p over posted prices, with a continuous density f_p, and supp$(F_p) = [0,1]$. For the converse, take a lottery F_p, cont. density f_p, supp$(F_p) = [0,1]$. Then there exists a unique p^* which is a REQ of MBG p^*, and such that F_p is the mechanism corresponding to p^*.*

Proof. We first show that (2.1) is equivalent to XPIC. We will focus on the seller, the proof for the buyer is identical. Let p^* be a REQ profile, that is differentiable strictly type-monotone profile satisfying (2.1). Define for each v, $\tilde{t}(v) = \min\{t \mid p_b^*(v_b, t) = p_s^*(v_s, t)\}$, where $\min\{\} = \infty$. Since p^* is a REQ profile, $\tilde{t}$ is well defined, and it is differentiable, by the Implicit Function Theorem. Now let $\pi(v) = p_s^*(v_s, \tilde{t}(v)) =$

$p_s^*(v_s, \tilde{t}(v))$ so that taking the derivative w.r.t. v_s we obtain $\frac{\partial \pi(v)}{\partial v_s} = \frac{\partial p_b}{\partial t}\frac{\partial \tilde{t}}{\partial v_s}$. Therefore,

$$\frac{\partial p_b}{\partial t} = \frac{1}{\frac{\partial \tilde{t}}{\partial v_s}} \frac{\partial \pi(v)}{\partial v_s}.$$

Defining $\delta(v) = e^{-\tilde{t}(v)}$, substituting this and the expression for $\frac{\partial p_b}{\partial t}$ into (2.1), and multiplying by $e^{-\tilde{t}(v)}$ we obtain

$$\delta(v)\frac{\partial \pi(v)}{\partial v_s} = -\frac{\partial \delta(v)}{\partial v_s}(\pi(v) - v_s).$$

This is precisely the necessary and sufficient FOC for XPIRIC mechanisms given in Section 4, when π and δ are both differentiable. Since in a REQ $p_s^*(0,0) = p_b^*(1,0)$, this implies that $\tilde{t}(0,1) = 0$, so that (π, δ) must be a XPCE mechanism, so it is representable by some probability distribution F_p. The other properties of F_p follow immediately. For the converse, if F_p is a continuously differentiable distribution with $supp(F_p) = [0,1]$, then we can construct the equivalent representation (π, δ). Now we can do the above substitutions in the other direction, and thus construct a unique pair of strategies p^* satisfying (2.1), so that p^* is a REQ. Thus, the solutions to (2.1) exist, and implement precisely all the allocations that are implementable by differentiable XPCE mechanisms. □

Corollary 2.5.4. *The REQ equilibria of MBG are XPCE.*

Proof. Take a REQ of MBG, $(p^*; MBG)$. Suppose there existed a dynamic bargaining game Γ and an XPEQ profile $(\tilde{p}^*; \Gamma)$, dominating $(p^*; MBG)$. Now let m be the mechanism corresponding to $(p^*; MBG)$, and let $\tilde{m}$ be the mechanism corresponding to $(\tilde{p}^*; \Gamma)$. Since $(\tilde{p}^*; \Gamma)$ dominated $(p^*; MBG)$, it must be that $\tilde{m}$ dominates m, which is a contradiction by Theorem 1.3.1 and Proposition 2.5.3. □

Theorem 2.5.3 implies that the set of outcomes of REQ is dense in the set of outcomes of XPEQ of MBG. To see this take an XPEQ of MBG, and the associated XPIRIC mechanism, which is representable as a distribution F_p over posted prices by Theorem 1.3.1. Then there exists a sequence of continuously differentiable distributions converging to F_p pointwise (on $[0,1]$), and the outcomes of the mechanisms converge pointwise (in the type space) to the outcomes under F_p. By Theorem 2.5.3, for each continuously-differentiable distribution over posted prices there is an REQ of MBG implementing the allocation of utilities under the distribution F_p. See also Example 2.5.8.

In the next example we show that there is a unique REQ which is linear in agents' types and the allocation is consistent with the Nash solution for all draws of types (and thus with the limit of the allocations in the Rubinstein bargaining game, as the time between the offers goes to 0). Note that in the present setup with incomplete information, the delay occurs almost surely (i.e., except when $v_s = 0$ and $v_b = 1$). We will show in the next section that a unique linear REQ exists under more general circumstances.

Example 2.5.5. There is a unique Nash-solution consistent REQ. It is given by the following type-contingent strategy profile:

$$\begin{aligned} p_s\left(v_s, t\right) &= \min\left\{1, v_s + \frac{e^{-t}}{2}\right\}, \\ p_b\left(v_b, t\right) &= \max\left\{0, v_b - \frac{e^{-t}}{2}\right\}. \end{aligned}$$

The Nash solution prescribes $\pi(v) = \frac{v_b+v_b}{2}$. Taking a uniform distribution over posted prices in $[0,1]$ yields the mechanism $\pi(v) = \frac{v_b+vs}{2}$, $\delta(v) = \max\{v_b - v_s, 0\}$. Checking that (2.1) holds is a straightforward computation. It is also easy to check that no other positive density over $[0,1]$ can sustain $\pi(v) = \frac{v_b+v_b}{2}$.

We remark that our model admits an interpretation as the limit of a game of alternating moves à la Rubinstein [1982], when the length of the period goes to zero and proposals are submitted to the Mediator. Example 2.5.5 describes the unique REQ profile consistent with such interpretation, since agreement at $\frac{v_b+v_b}{2}$ prevails uniquely at subgames where types have been revealed (See Binmore, Rubinstein, and Wolinsky [1986]). However, note that this linear equilibrium outcome only coincides with the Nash solution when agents are risk neutral or they have the same risk aversion, see Proposition 2.6.1 in section 2.6 and the subsequent comment.

Example 2.5.6. In this example we construct two REQ in non-linear strategies. In the first one the strategies can be explicitly computed. The second one is symmetric, but the strategies cannot be computed in closed form. Take a lottery over posted prices given by a pdf $f_p(x) = 2x, x \in [0,1]$. Note that f_p is differentiable and strictly positive, so that the corresponding strategies of MBG will satisfy all the conditions for a REQ. To construct the strategies proceed as follows. First,

$$\delta(v) = \int_{v_s}^{v_b} f_p(\tau)d\tau = v_b^2 - v_s^2, \text{ and}$$

$$\pi(v) = E_{f_p}[p \mid p \in [v_s, v_b]] = \frac{1}{v_b^2 - v_s^2}\int_{v_s}^{v_b} \tau f_p(\tau)d\tau = \frac{2}{3}\frac{v_b^2 + v_s v_b + v_s^2}{v_b + v_s}.$$

Since $\delta(v) = e^{-t}$, where t is the time of agreement between types v_s and v_b, we get $\tilde{v}_b(v_s, t) = \sqrt{e^{-t} + v_s^2}$, where $\tilde{v}_b(v_s, t)$ is the type of buyer who agrees with the seller v_s at time t. Noting that $p_s(v_s, t) = \pi\left(v_s, \tilde{v}_b(v_s, t)\right)$ we obtain

$$p_s(v_s, t) = \frac{2\left(e^{-t} + 2v_s^2 + v_s\sqrt{e^{-t} + v_s^2}\right)}{3\left(v_s + \sqrt{e^{-t} + v_s^2}\right)}.$$

Similarly, we could compute the strategy of the buyer.

For the second example consider $f_p(x) = 6x(1-x)$. Then $\pi(v) = \frac{2(v_b^3 - v_s^3) - \frac{3}{2}(v_b^4 - v_s^4)}{3(v_b^2 - v_s^2) - 2(v_b^3 - v_s^3)}$ and $\tau(v) = -\ln\delta(v)$, where $\delta(v) = F(v_b) - F(v_s) = (3v_b^2 - 2v_b^3) - (3v_s^2 - 2v_s^3)$. Thus, the strategy of the buyer is $p_b(v_s, t) = \pi(v_b, \chi(v_b, t))$ where $\chi(v_b, t))$ solves $3v_b^2 - 2v_b^3 - e^{-t} = 3\chi^2(v_b, t) - 2\chi^3(v_b, t)$, and similarly for the seller.

Next, we provide a simple example of an XPEQ which is not a REQ.

Example 2.5.7. Let $p^* \in [0,1]$, and consider the following strategies of the traders. The seller's types $v_s \leq p^*$ commit to always demanding p^*, and the types $v_s > p^*$ commit to always demanding 1. Similarly, the buyer's types $v_b \geq p^*$ always bid p^*, and $v_b < p^*$ always bid 0. It is trivial to check that this is an XPEQ of MBG, and it is clearly not a REQ. The direct-revelation mechanism corresponding to this XPEQ is a degenerate distribution F_p (by virtue of Theorem 1.3.1) with point mass at $p^* \in [0,1]$.

Using this logic, and the representation of Theorem 1.3.1, the reader can construct more contrived examples at will. That is, take some distribution F_p which is not continuous w.r.t.the Lebesque measure, and there exists an XPEQ (which is not a REQ) of MBG, such that the direct mechanism corresponding to that XPEQ is the given F_p.

Finally, we present a standard example of welfare analysis in terms of *ex-ante* constrained efficiency. We again stress that XPCE is necessary for *ex-ante* constrained efficiency. Thus, it is enough to look for optimal mechanisms within the class of probability distributions over posted prices. Moreover, under risk-neutrality, the *ex-ante* optimal mechanism is a deterministic posted price (i.e., a point-mass at the *ex-ante* optimal posted price). By the previous example, the corresponding XPEQ is not a REQ. (In contrast, under risk aversion the *ex-ante* optimal XPEQ is generically a REQ, see Example 2.6.3 of Section 2.6, and Čopič and Ponsatí [2005].)

Example 2.5.8. Let v_b and v_s be iid, uniform on $[0,1]$. For simplicity we find the *ex-ante* constrained-efficient mechanism that maximizes the sum of expected utilities. In this case it is quite obvious that the only candidate is by symmetry a posted price $p^* = \frac{1}{2}$ (i.e., a degenerate distribution over posted prices with a point-mass at $\frac{1}{2}$). The welfare under this mechanism is

$$\int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^1 (\frac{1}{2} - v_s) + (v_b - \frac{1}{2}) dv_b dv_s = \frac{1}{8}.$$

From the previous example we know that there is an XPEQ of MBG corresponding to this mechanism, but this XPEQ is not an REQ. On the other hand, it is straightforward that for each continuously differentiable F_p with full support, there is an REQ of MBG which implements that F_p. Therefore, there exists a sequence of REQ, approximating the outcome under $p^* = \frac{1}{2}$ (pointwise in the type space)-take for instance $f^n = k_n x^n (1-x)^n, n = 1, 2, ...,$ and k_n is chosen so that f_n integrates to 1. Thus, the outcome under p^* is a limit point of the set of REQ outcomes. The welfare under the linear REQ is $\frac{1}{12}$, and the welfare under the REQ corresponding to

$f_p(x) = 6(1-x)x$ is $\frac{1}{10}$. For comparison, in this same environment, Cramton [1992] computes a symmetric stationary separating PBE where expected benefits equal to $\frac{3}{32}$, so that the *ex-ante* ranking of welfare in these equilibria is $\frac{1}{8} > \frac{1}{10} > \frac{3}{32} > \frac{1}{12}$ (the optimal robust XPEQ $\succ$ symmetric non-linear REQ of Example 2.5.6 $\succ$ Cramton's PBE $\succ$ the linear REQ). Note that the PBE in Cramton [1992] is not an XPEQ (thus it is not robust). Also note that when agents are risk averse, the linear REQ is *ex-ante* more efficient than the most efficient posted price-see the continuation of this example, Example 2.6.3 of Section 2.6.

2.6 Risk aversion and unequal discount rates.

In this section, we discuss MBG in a slightly richer model. The agents' static preferences display constant relative-risk aversion (CRRA), i.e., $u_s(v_s, p) = (p - v_s)^{\gamma_s}, u_b(v_b, p) = (v_b - p)^{\gamma_b}$, where $\gamma_i \in (0, 1], i = s, b$. The agents are allowed to discount the future differently, so that time preference of i is given by $\rho_i \geq 1$- agent i discounts according to $e^{-\rho_i t}$, $i = s, b$. The restriction that $\rho_i \geq 1$ is without loss of generality since all that matters are relative rates of discounting. Parameters γ and ρ are common knowledge. We will show that risk aversion and time preference act as substitutes, so that behaviorally, an agent that is more impatient *acts as if he were more risk averse.* In particular, this is true in a static direct mechanism, so that the mechanism has to be adjusted for risk aversion *and impatience*-even though the game is static. The point is that it matters that a direct mechanism is a reduced form of a dynamic game.

We limit ourselves to the present setup mostly for the sake of tractability and also because the present case exhausts the environments where MBG admits REQ in linear strategies.[8] It is worth noting that in a dynamic game, for agents to display preferences that are consistent in the intertemporal sense, we have to restrict the agents instantaneous utility functions to display constant relative-risk aversion

[8] One could consider a richer model, where each agent has some concave utility function and some discounting criterion.

(CRRA) (See Fishburn and Rubinstein [1982]). Nonetheless, redoing the present exercise under other behavioral assumptions may be interesting.

We first derive the FOC for a REQ in this environment (the argument is identical to the argument in the Proof of Proposition 2.5.2 above). So let p^* be a REQ profile. Now, given p_b^*, the seller considers the problem

$$\max_{t\in[0,\infty)} e^{-\rho_s t}\left(p_b^*(v_b,t)-v_s\right)^{\gamma_s},$$

which yields the first-order condition (similarly for the buyer)

$$(p_b^*(v_b,t)-v_s)=\frac{\gamma_s}{\rho_s}\frac{\partial p_b^*(v_b,t)}{\partial t}. \tag{2.3}$$

Observe that it is impossible to distinguish the first-order condition for agents that are risk averse from the first-order condition for impatient agents. In particular, in the direct mechanism, we need to consider $\gamma_i' = \frac{\gamma_i}{\rho_i}$ as the risk-aversion parameter of agent i.

Consider first the case when $\rho_s = \rho_b = \rho$, so that $\gamma_i' = \frac{\gamma_i}{\rho}$, $\delta_v = e^{-\rho\tilde{t}(v)}$, and $\pi(v) = p_b^*(v_b,\tilde{t}(v)) = p_s^*(v_s,\tilde{t}(v))$, where again $\tilde{t}(v) = \min\{t \mid p_b^*(v_b,t) = p_s^*(v_s,t)$, to obtain

$$(\pi(v)-v_s)\frac{\partial\delta(v)}{\partial v_s} = -\gamma_s'\delta(v)\frac{\partial\pi(v)}{\partial v_s}. \tag{2.4}$$

By differentiability of the profile p^* both δ and π are differentiable, and (2.4) is precisely the XPIRIC condition for differentiable mechanisms when γ' are the risk aversion parameters, which is easy to check along the lines of Section 4.

When $\gamma_i' \neq 1$ for at least one i there is no representation of XPIRIC mechanisms in terms of distributions over posted prices as in Theorem 1.3.1. Still, for each mechanism $m = (\delta,\pi)$, satisfying (2.4), we can construct by the above substitutions exactly one strategy profile p^* satisfying the necessary and sufficient conditions (2.3) for a REQ. Thus, the analog to Theorem 2.5.3 holds. For a more detailed treatment of XPIRIC mechanisms under risk aversion see Čopič and Ponsatí [2005], where we also prove that the mechanisms described by equation (2.4) are XPCE. [9]

Since $\delta(v) = e^{-\rho\tilde{t}(v)}$, each trader perceives the deterministic trade at price $\pi(v)$ and time $\tilde{t}(v)$ exactly the same as instantaneous trade at price $\pi(v)$ with probability $\delta(v)$. The price is distorted due to risk sharing, and the probability may be affected by impatience as well. Notice that we could also reparametrize time to $\tau = \rho t$, and under this new time-scale there would be no distortion of perceived probability (i.e., agents getting older faster or being more impatient is formally equivalent).

Similarly, when $\rho_s \neq \rho_b$ what matters is the $\gamma_i' = \frac{\gamma_i}{\rho_i}, i = s, b$, and Equation 2.4 still describes the XPIRIC condition for the direct mechanisms. Therefore, the difference of the relative impatience also has an effect on pricing, as well as on the probability of trade. Again, even in the static setup of direct mechanisms we have to take into account the impatience, and not only the risk aversion of the agents. Behaviorally, more impatient agents act as if they were more risk averse. Now there does not exist a rescaling of time units that would work for both traders. See also Example 2.6.2 at the end of this section.

For the rest of this section we limit ourselves to the unique mechanism (and REQ of MBG) where pricing is linear in agents' types. We remark that in environments where agents' risk attitudes are not CRRA or they do not discount the future exponentially, no linear pricing mechanism exists (see Čopič and Ponsatí [2005]).

Proposition 2.6.1. *Given the environment described by* (γ, ρ)*, there exists a unique solution* (δ, π) *to (2.4) such that* $\delta(v)$ *is linear in* v *and* $\delta(0, 1) = 1$*. More precisely,*

$$\pi(v) = \frac{\sqrt{\gamma_s'}}{\sqrt{\gamma_s'} + \sqrt{\gamma_b'}} v_b + \frac{\sqrt{\gamma_b'}}{\sqrt{\gamma_s'} + \sqrt{\gamma_b'}} v_s, \delta(v) = (v_b - v_s)^{\sqrt{\gamma_s' \gamma_b'}}, v_b \geq v_s. \tag{2.5}$$

Proof. Let $\pi(v) = \alpha v_s + (1 - \alpha) v_b$, and insert this into (2.4). This gives

$$\frac{\partial \log \delta(v)}{\partial v_s} = -\frac{\gamma_s' \alpha}{1 - \alpha} \frac{1}{v_b - v_s},$$

[9] The main problem is that in a non-linear environment there are XPIRIC mechanisms that cannot be represented as binary lotteries (since an agent is no longer indifferent between the lottery and its mean, both on and off the equilibrium path). In Čopič and Ponsatí [2005] we prove that the mechanisms that are binary lotteries are XPCE.

$$\frac{\partial \log \delta(v)}{\partial v_b} = \frac{\gamma_b'(1-\alpha)}{\alpha} \frac{1}{v_b - v_s}.$$

By integrating the first equation we obtain $\log \delta(v) = \frac{\gamma_s' \alpha}{1-\alpha} \log(v_b - v_s) + K_s(v_b)$, and from the second we obtain $\log \delta(v) = \frac{\gamma_b'(1-\alpha)}{\alpha} \log(v_b - v_s) + K_b(v_s)$, where $K_s(v_b)$ and $K_b(v_s)$ are integration constants. But then it must be that $K_s = K_b = const.$ (determined from $\delta(0,1) = 1$) and $\frac{\gamma_b'(1-\alpha)}{\alpha} = \frac{\gamma_s' \alpha}{1-\alpha}$. Therefore α is uniquely determined.

$\square$

Observe that pricing under the linear mechanism is different from the Nash-solution pricing, which is $\frac{\gamma_s'}{\gamma_s' + \gamma_b'} v_b + \frac{\gamma_b'}{\gamma_s' + \gamma_b'} v_s$, and is still attainable as the limit of SPNE of the Rubinstein alternating offers game, when v_s and v_b are known. This difference is not surprising since the contraction independence property that is required in the (generalized) Nash solutions, and implied by equilibrium conditions in the Rubinstein game, is clearly not equivalent to the incentive constraints. However, it is remarkable that risk sharing goes in the same direction for both allocations: the more risk averse agent obtains less surplus. Note also that in both models (the present one and the Rubinstein alternating-offers model) risk sharing and impatience have an effect on pricing which goes in the same direction: the more impatient agent gets less, and the more risk-averse agent gets less. We conclude this section with two examples.

Example 2.6.2. Let $\frac{\gamma_s}{\rho_s} = \gamma_s' = \frac{\gamma_b}{\rho_b} = \gamma_b' = \gamma'$, so that the unique linear mechanism is given by $\pi(v) = \frac{1}{2}(v_1 + v_2), \delta(v) = (v_2 - v_1)^{\gamma'}$. The strategies of the agents in the appropriate REQ of MBG are given by

$$p_s^*(v_s, t) = v_s + \frac{e^{-\gamma' t}}{2}, p_b^*(v_b, t) = v_b - \frac{e^{-\gamma' t}}{2}.$$

Thus, there is a two-parametric family of environments where the utility outcome is invariant, and it is the same regardless of whether each agent is impatient or risk averse. We remark that while in the static direct mechanism sense we have to adjust for dynamic time preference, in the dynamic setting we have to adjust for the static risk aversion, even though at each time the outcome is deterministic.

In the last example we show that when agents are risk averse the *ex-ante* optimal (under a utilitarian social welfare function) XPEQ of MBG is a REQ. Computing the *ex-ante* optimal mechanism is a bit complicated (and can in general only be done numerically), and we refer an interested reader to Čopič and Ponsatí [2005].

Example 2.6.3. Let $\gamma_s = \gamma_b = \gamma, \gamma \in (0,1]$ and $\rho_s = \rho_b = 1$, and as before, let the social welfare be given by $u_s + u_b$. Also, let v_b and v_s be iid, uniform on $[0,1]$. Then, by symmetry, the most *ex-ante* efficient posted price is $p^* = \frac{1}{2}$. The *ex-ante* social welfare under p^*, as a function of risk aversion γ is

$$W^p(\gamma) = \frac{1}{2(\gamma+1)} \left(\frac{1}{2}\right)^{\gamma+1},$$

and the social welfare under the linear REQ is

$$\frac{1}{2(\gamma+1)(2\gamma+1)} \left(\frac{1}{2}\right)^{\gamma}.$$

These two expressions are equal when $\gamma = \frac{1}{2}$. For more risk-averse traders (i.e. $\gamma < \frac{1}{2}$) social gains are higher under the linear REQ. In fact, the linear REQ approaches *ex-post* efficiency as the traders' risk aversion goes to infinity.

2.7 Separating PBE are REQ

We now show that every separating perfect Bayesian equilibrium (PBE) of MBG must be a REQ. We note that in MBG, the off-equilibrium deviations are unobservable so that the set of outcomes of PBE and the set of outcomes of BE coincide.

Beliefs. Recall from Section 2 that it is common knowledge that types are drawn from some distribution with support $[0,1]^2$. In this section, we assume that the specific pdf G is also common knowledge. Agent i updates her beliefs over the distribution of the opponent's types over time. As described in Section 4, the histories depend only on t. Thus, given a strategy profile p, the beliefs of a player about the opponent are updated only as a function of time. We denote by $G_j\left(v_j|v_i,t;p\right)$ the distribution of the belief of agent i of type v_i about agent j at time t, conditional on no agreement until time t. By $g_j(v_j|v_i,t;p)$ we denote the density of G_j, whenever it exists. Finally, we denote by $H_j(v_i,t;p)$ the mass of types of player j with whom agent i has agreed with by time t. We will economize the notation and omit parameters v_i and p whenever that is unambiguous. Note that if the strategies of both players are differentiable with respect to both parameters, and these partial derivatives are non-zero, the beliefs will be differentiable with respect to time.

Bayes and perfect Bayes-Nash equilibrium. Denote by $EU_i(v_i;p,G)$ the expected payoff of player i of type v_i, when agents play according to strategy profile p and types are distributed according to G. Let $G_j(v_i)$ denote the conditional distribution of j's types. Thus,

$$EU_i(v_i;p,G) = \int_0^1 u_i(\bar{p}i(p,v_i,v_j),v_i)\, e^{-\tau(p,v_i,v_j)} dG_j(v_i),$$

or alternatively

$$EU_i(v_i;p,G) = \int_{t\in[0,\infty)} u_i(p_i(v_i,t),v_i)\, e^{-t} dH_j(v_i,t),$$

where both of these integrals have to be understood as Lebesgue integrals.

Denote by Π_i the set of strategies for player i, and by $\Pi = \Pi_s \times \Pi_b$ the set of strategy profiles. A strategy profile $p = (p_i,p_j) \in \Pi$ constitutes a *Bayes-Nash equilibrium* if and only if

$$EU_i(v_i;p,G) \geq EU_i(v_i;p_i',p_j,G),\ \forall\ p_i' \in \Pi_i,$$

$\forall\ \ v_i \in [0,1]$, $i = s,b$, $j \neq i$.

A careful definition of the PBE in our setting requires specifying agents' expected utility in every subgame, which in our setup means at every time t. Let $EU_i(v_i,t;p,G)$ denote the expected payoff to player i of type v_i in the subgame starting at t, when agents play strategies p (note that p, v_i, and t also specify the history observed by agent i):

$$EU_i\left(v_i,t;p,G\right)=\int\limits_{\tau\in[t,\infty)} u_i\left(p_i\left(v_i,\tau\right),v_i\right)e^{-\tau}dH_j\left(v_i,\tau\right)$$

A strategy profile $p\in\Pi$ constitutes a *perfect Bayesian equilibrium* if

$$\begin{aligned} EU_i\left(v_i,t;p,G\right) &\geq EU_i\left(v_i,t;p_i',p_j,G\right), \\ \forall\ p_i' &\in \Pi_i \text{ s.t.} p_i'\left(v_i,t'\right)=p_i\left(v_i,t'\right) \text{ for all } t'\leq t, \end{aligned}$$

for all $t\geq 0$, for all $v_i\in[0,1]$, $i=s,b$, $j\neq i$. As we noted earlier, BE and PBE are outcome equivalent in MBG.

We impose the following conditions and restrict attention to BE in strongly regular strategies. Note that a BE in regular strategies is a BE of MBG which we show in Corollary 2.7.7, in the Appendix.

R We say that a strategy is *strongly regular* if $\frac{\partial p_i(v_i,t)}{\partial v_i}$ is continuous $\forall\ t\in[0,\infty)$ and $\lim_{t\to\infty}p_i\left(v_i,t\right)$ is a left-continuous function of v_i, for all $v_i\in[0,1]$.

SEP We say that a strongly regular strategy is *separating* if $\frac{\partial p_i(v_i,t)}{\partial v_i}\neq 0,\forall\ \ t\in[0,\infty)$ and $\forall\ v_i\in(0,1)$.

The regularity condition R imposes a pattern of behavior that rules out dramatic changes when types change only marginally, which is a natural requirement since types and dates take values in a continuum. The second part of R is roughly an indifference breaking rule: if an agent of some type is at the horizon indifferent between two concessions to the opponent, she will concede more. In Lemma 2.7.3 in the Appendix we show that this condition is enough to assure the continuity of the demands with respect to types at the time horizon and that in a regular equilibrium the agents' bids asymptotically approach the reservation values.

With the main theorem of this section we wrap up our paper.

Theorem 2.7.1. *All strongly regular and separating PBE (and thus BE) of MBG are REQ.*

The sketch of the proof goes roughly as follows. First, we show that a differential first-order condition for a strongly regular BE is well defined. Then we show that the strategies resulting from this first-order condition must be belief independent so that a BE is an XPEQ. The intuition behind this is that a separating equilibrium is fully revealing, i.e. for each proposal and each date the seller will know exactly the valuation of the opponent with whom she agrees at that proposal and date. Thus, once the agreement occurs the agents know each other's types, and since this is common knowledge *ex-ante*, they must play best-replies against the strategy of each type of the other player. For details see the Appendix.

Appendix

Proposition 2.7.2. WEAK TYPE MONOTONICITY: *In every BE,* $\frac{\partial p_i(v_i,t)}{\partial v_i} \geq 0$, *for all times* $t \in (0,\infty)$ *and types* $v_i \in [0,1]$, *which satisfy the condition that* $H_j(v_i,t)$ *is strictly increasing at* t.

Proof. Fix the buyer strategy at some $p_b(.,.)$. Denote by $H_b(v_s,t;p_s)$ the mass of buyer's types with whom v_s enters in agreement until time t if she plays the strategy $p_s(.,.)$. Observe that at any t, s.t.$\exists$ v_b with $p_s(v_s,t) = p_b(v_b,t)$, $H_b(v_s,t;p_s)$ is strictly increasing if and only if $p_s(v_s,.)$ is strictly decreasing or $p_b(v_b,.)$ is strictly increasing in t. This follows from continuity of $p_s(.,.)$ and $p_b(.,.)$ w.r.t. v. Moreover, $H_b(v_s,t;p_s)$ has a jump at t if and only if $\exists$ v_b', v_b'' s.t.$p_s(v_s,t) = p_b(v_b,t)$ for all $v_b \in (v_b'', v_b')$.

We have to show that $p_s(v_s,t) \geq p_s(v_s',t)$ for any $v_s \geq v_s'$ and any t s.t.$H_b(v_s,t;p_s)$ is strictly increasing at t (at any v_s, where the condition in the statement of the lemma is satisfied, $H_b(v_s,t;p_s)$ is strictly increasing, and it can have a jump).

We proceed by contradiction. Assume there are $v_s > \hat{v}_s$ and $\hat{t}$ s.t.$p_s(v_s,\hat{t}) < p_s(\hat{v}_s,\hat{t})$ and $H_b(v_s,\hat{t};p_s)$ is strictly increasing at $\hat{t}$. Denote by

$$t_0 = \inf\left\{t | H_b(v_s,t;p_s) > 0, t < \hat{t}, \text{ and } p_s(v_s,t) < p_s(\hat{v}_s,t) \text{ for all } \tau \in (t,t')\right\},$$

$$t_1 = \min\left\{t | t > \hat{t}, p_s(v_s,t) = p_s(\hat{v}_s,t)\right\}.$$

In other words, t_0 is the largest time until which the demands of v_s and $\hat{v}_s$ are monotonic, and t_1 is the first time after t_0 at which these demands are equal. First, by continuity of $p_s(v_s,.)$ and $p_s(\hat{v}_s,.)$ it is clear that $t_0<\hat{t}<t_1$. Moreover, $t_1<\infty$ since by the previous lemma, $\lim_{t\to\infty} p_s(v_s,t)=v_s>\hat{v}_s=\lim_{t\to\infty} p_s(\hat{v}_s,t)$, hence, by continuity there exists a $\bar{t}<\infty$ s.t.$p_s(v_s,t)>p_s(\hat{v}_s,t)$ for all $t\geq\bar{t}$. Since $H_b(v_s,\hat{t};p_s)$ is strictly increasing at $\hat{t}$, it is also clear that $H_b(v_s,t_0;p_s)<H_b(v_s,t_1;p_s)$.

If some seller type bids lower at time t she will have agreed with a larger mass of the buyer's types. In other words, $p_s(v_s,t)\leq p_s(\hat{v}_s,t)\Rightarrow H_b(v_s,t;p_s)\geq H_b(\hat{v}_s,t;p_s)$ for all t and all v_s and $\hat{v}_s$, which follows from the monotonicity of $p_s(.,.)$ and $p_b(.,.)$ w.r.t. t. Applying this twice at t_0 and t_1, we get $H_b(v_s,t_0;p_s)=H_b(\hat{v}_s,t_0;p_s)$ and that $H_b(v_s,t_1;p_s)=H_b(\hat{v}_s,t_1;p_s)$. By construction, we have $p_s(v_s,t)<p_s(\hat{v}_s,t)$ for all $t\in(t_0,t_1)$. This implies that $H_b(v_s,t;p_s)\geq H_b(\hat{v}_s,t;p_s)$ for all $t\in(t_0,t_1)$.

In equilibrium, $p_s(v_s,.)$ is the optimal strategy for type v_s, and $p_s(\hat{v}_s,.)$ is optimal for type $\hat{v}_s$ on the interval (t_0,t_1). In particular (from now on we omit subindexes and write $p_s(v_s,t)=p(t)$, $p_s(\hat{v}_s,t)=\hat{p}(t)$, $H_b(v_s,t;p_s)=H(t)$ and $H_b(\hat{v}_s,t;p_s)=\hat{H}(t)$)

$$\int_{t_0}^{t_1} e^{-t}(p(t)-v_s)\,dH(t)\geq\int_{t_0}^{t_1} e^{-t}(\hat{p}(t)-v_s)\,dH(t) \tag{2.6}$$

and

$$\int_{t_0}^{t_1} e^{-t}(\hat{p}(t)-\hat{v}_s)\,d\hat{H}(t)\geq\int_{t_0}^{t_1} e^{-t}(p(t)-\hat{v}_s)\,d\hat{H}(t)\,. \tag{2.7}$$

Subtracting these two inequalities, we obtain

$$\int_{t_0}^{t_1} e^{-t}dH(t)\leq\int_{t_0}^{t_1} e^{-t}d\hat{H}(t)\,.$$

Integrate by parts to get $\int_{t_0}^{t_1} e^{-t}dH(t)=H(t_1)\,e^{-t_1}-H(t_0)\,e^{-t_0}+\int_{t_0}^{t_1} e^{-t}H(t)\,dt$, and similarly for the right hand side. Now we use $H(t_0)=\hat{H}(t_0)$ and $H(t_1)=\hat{H}(t_1)$, to obtain

$$\int_{t_0}^{t_1} e^{-t}H(t)\,dt\leq\int_{t_0}^{t_1} e_b^{-t}\hat{H}(t)\,dt.$$

But since $H(t)\geq\hat{H}(t)$ for all $t\in(t_0,t_1)$ the last inequality implies that it must in fact be $H(t) = \hat{H}(t)$ for almost all $t \in (t_0, t_1)$. Now take for example (2.6), and rewrite it as

$$\int_{t_0}^{t_1} e^{-t} (p(t) - \hat{p}(t)) \, dH(t) \geq 0.$$

But $\hat{p}(t) > p(t)$ for $t \in (t_0, t_1)$, which implies that

$$\int_{t_0}^{t_1} e^{-t} (p(t) - \hat{p}(t)) \, dH(t) < 0,$$

which is a contradiction. □

Lemma 2.7.3. TOTAL CONCESSION AT INFINITY: *In a regular PBE,* $\lim_{t\to\infty} p_i(v_i, t) = v_i$ *for all* $v_i \in [0, 1]$.

Proof. Denote $P_i(v_i) = \lim_{t\to\infty} p_i(v_i, t)$. The proof is divided into three steps. In step 1 we show that $P_s(1) = 1$ (which holds trivially) and the continuity at 1 imply that $P_s(0) = 0$. In step 2 we show that $P_s(.)$ is a continuous function, hence it attains all values in the interval $[0, 1]$. Finally, in step 3 we show that the claim is true for the seller. An analogous proof works for the buyer.

Step 1: $P_s(0) = 0$. Suppose this isn't the case, i.e. $P_s(0) = K > 0$ in equilibrium. Denote by $p_s(0, t)$ such equilibrium strategy for the seller, and by $p_b(v_b, t)$ the equilibrium strategy of the buyer of type v_b. By individual rationality we have that $P_b(0) = 0$. Also by individual rationality, we have that $P_b(v_b)$ is bounded above, i.e. $P_b(v_b) \leq v_b$. Since $P_b(v_b) \geq 0$, these imply that $P_b(v_b)$ is continuous at point $v_b = 0$. From continuity of P_b around $v_b = 0$ we get that there is a positive mass of types $v_b \in [0, 1]$ for which $P_b(v_b) < K$. But then the seller of type 0 could improve her expected payoff by playing p_s until some large time t', and then lowering her demand to 0, according to some strategy p'_s. To see this, notice that p_s and p_b are continuous and for all v_b, $p_s(0, t)$ is non-increasing and $p_b(v_b, t)$ is non-decreasing in t. Thus the support of $g_b(v_b|t)$ is shrinking as time elapses. When t is very large, the support of $g_b(v_b|t)$ will be very close to the *ex-post* belief when no agreement has been reached. Hence t' is given as the moment when the expected continuation payoff of playing p_s, conditional on $v_b \geq K$, is lower than the expected continuation payoff of playing p'_s, conditional on $v_b > 0$. This establishes the contradiction. The same argument shows that $P_s(v_s)$ is continuous in a neighbourhood of the point $v_s = 0$.

Step 2. Assume thus that $P_s(v_s)$ is discontinuous at $\bar{v}_s$, i.e., $P_s(\bar{v}_s) = \hat{l}$ and $\lim_{v_s \searrow \bar{v}_s} = \bar{l}$, where $\bar{l} > \hat{l}$. Then there must exist an $\bar{v}_b$ s.t.$P_b(\bar{v}_b) = \bar{l}$, and $\lim_{v_b \nearrow \bar{v}_b} P_b(v_b) = \hat{l}$ (same argument as in step 1, left continuity of P_s and right continuity of P_b). Take a $\hat{v}_s > \bar{v}_s$. By continuity of p_s in t, there exists an M_s s.t.$p_s(\bar{v}_s, t) - \hat{l} < \varepsilon$ for all $t \geq M_s$. Also, notice that $p_s(\hat{v}_s, t) \geq \bar{l}$. Now fix $\varepsilon = \frac{\bar{l} - \hat{l}}{4} > 0$ and take a $t \geq M_s$. Then at t, $p_s(\bar{v}_s, t) < \hat{l} + \varepsilon$ while $p_s(\hat{v}_s, t) \geq \bar{l}$ for all $\hat{v}_s > \bar{v}_s$, contradicting the continuity of p_s in v_s. This proves that $P_s(v_s)$ has to be right-continuous. By assumption, $P_s(v_s)$ is left-continuous,[10] hence it is continuous. In step 1 we proved that $P_s(1) = 1$ and $P_s(0) = 0$, so by Rolle's theorem it attains all values between 0 and 1.

Step 3: $P_s(v_s) = v_s$ for all $v_s \in [0, 1]$. Take an $v_s \in (0, 1)$. By steps 1 and 2, P_s takes all the values in the interval $[0, 1]$ and is continuous (thus measurable), strictly positive on $(0, 1]$. Thus we can define the measure μ_s

$$\mu_s(V) = \int_V P_s(v)\, dm(v) \text{ for each mesaurable } V \subset [0, 1],$$

where $m(.)$ denotes the usual Lebesgue measure. By strict positivity, continuity, and boundedness of $P_s(v_s)$, μ_s is an equivalent measure to m. Now suppose that $P_s(v_s) > v_s$. By equivalence of μ_s to m there exists a positive mass of types v_b s.t.$p_b(v_b) \in (v_s, P_s(v_s))$. To see this define $B = \{v_b | p_b(v_b) \in (v_s, P_s(v_s))\}$. Since μ_s and m are equivalent, $m(B) > 0$. Now repeat the same argument as in step 1 to get a contradiction. Hence indeed $P_s(v_s) = v_s$.

□

Recall that the entry time $t_i^E(v_i)$ is the first time when v_i could agree with some type of player j, $t_i^E(v_i) = \min\{t \mid \tilde{v}_j(v_i, t) \neq \emptyset\}$. It is easy to see that at $t_i^E(v_i)$ the

[10] Type $\bar{v}_s$ is at $t = \infty$ indifferent between demanding $\hat{l}$ and $\bar{l}$; the former does not improve her probability of reaching an agreement since the mass of opposing types with bids between $\bar{l}$ and $\hat{l}$ is 0. However, by an argument similar to the proof of step 1, we can argue, that she does not lose anything by bidding $\hat{l}$, which gives us left continuity of P_s. Left continuity of P_s is thus essentially an assumption on how agents resolve their indifference at the horizon.

demand of type v_i must be compatible exactly with that of the weakest type of the opponent.

Lemma 2.7.4. Initial proposal and entry time: *In an undominated regular BE* $p_s(v_s, t_s^E(v_s)) = p_b(1, t_s^E(v_s))$ *and* $p_b(v_b, t_b^E(v_b)) = p_s(0, t_b^E(v_b))$*, for all* $v_i \in [0,1]$.

Proof. Denote by $\gamma_i(v_i)$ the starting point of the bids of type v_i: $\gamma_i(v_i) = \lim_{t \searrow 0} p_i(v_i, t)$. We will prove that $\gamma_s(0) = \gamma_b(1)$, which proves the Lemma. In an undominated BE the type $v_s = 0$ at time 0 demands a share that will give her a positive probability of agreement in an infinitesimal amount of time. On the other hand, it cannot be that at $t = 0$, the seller $v_s = 0$ bids a price which meets the bid of some buyer of type $v_b^0 < 1$-meaning that $\gamma_s(0) = \gamma_b(v_b^0)$. The reason is that type $v_s = 0$ could profitably deviate by starting with a bid that meets type $v_b = 1$ and then in an infinitesimal time lower her bid to $\gamma_b(v_b^0)$. By type monotonicity of buyer's strategy, such deviation would be profitable. □

We remark that in each undominated strongly regular BE $t_i^E(v_i) < \infty$ if and only if $v_s < 1, v_b > 0$. Otherwise the strategy of v_i would be strictly dominated.

We now write down the dynamic-optimization problem. In equilibrium, agents maximize payoffs, given the type-contingent strategies of the other player. Thus, agents are picking optimal functions $p_i(v_i, \cdot)$, $i = s, b$, determining how bids change over time.

An important step in the proof of proposition 2.7.5 below is to show that for every $(v_i, t) \in [0,1] \times [t_i^E(v_i), \infty)$,$\tilde{v}_j(v_i, t)$ is a function (and not a correspondence), defined by

$$p_j(\tilde{v}_j(v_i, t), t) = p_i(v_i, t). \tag{2.8}$$

This is a consequence of the assumption that the opponent plays a strictly type-monotone strategy, and the implicit function theorem.

Proposition 2.7.5. Optimization Program: *If the strategy of agent j is strongly regular and separating, then the best reply of agent i of type v_i solves the following* *optimization program*

$$Max_{p_i(v_i,\cdot)\in\Pi_i} \int\limits_{[t_i^E(v_i),\infty)} e^{-t} u_i(p_i(v_i,t),s_i) g_j(\tilde{v}_j(v_i,t)) \frac{\partial \tilde{v}_j(v_i,t)}{\partial t} dt,$$

$$\textit{s.t. (2.8) and } t_i^E(v_i) \textit{ defined by } \tilde{v}_b\left(v_s, t_s^E(v_s)\right) = 1 \textit{ or } \tilde{v}_s\left(v_b, t_b^E(v_b)\right) = 0.$$

Proof. Consider the seller and fix her type to be v_s. When entering into negotiations at $t_s^E(v_s)$, she decides her optimal concession plan $p_s(v_s,t)$, $t > t_s^E(v_s)$, in order to maximize her expected discounted future payoff. Denote by $H_b(t)$ the probability of seller v_s reaching agreement up to time t (we omit the parameter v_s in $H_b(t;v_s)$). The seller is solving the following program

$$Max_{p_s(v_s,\cdot)\in\Pi_s} \int\limits_{[t_s^E(v_s),\infty)} e^{-t}(p_s(v_s,t) - v_s) dH_b(t).$$

But the possibility of reaching an agreement at some $t > t_s^E(v_s)$ is exactly the possibility that the seller's bid will at t meet the bid of some type of the buyer. For each $t \geq t_s^E(v_s)$, recall that $\tilde{v}_b(v_s,t)$ is the type of buyer with whom v_s reaches agreement at moment t. Thus $\tilde{v}_b(v_s,t)$ is implicitly defined from the relation

$$p_b(\tilde{v}_b(v_s,t), t) = p_s(v_s,t).$$

Type monotonicity implies that at every instant there will be at most one type reaching agreement with each type of the other agent. Thus, by definition of t_i^E $\tilde{v}_b\left(v_s, t_s^E(v_s)\right) = 1$, and by Lemma 2.7.3 $\lim_{t\to\infty} \tilde{v}_b(v_s,t) = v_s$. Taking the derivative with respect to t, we can express

$$\frac{\partial \tilde{v}_b(v_s,t)}{\partial t} = \frac{\frac{\partial p_s(v_s,t)}{\partial t}}{\frac{\partial p_b(\tilde{v}_b(v_s,t),t)}{\partial v_b}}.$$

By assumption, $\frac{\partial p_i}{\partial t}$ are both finite, $\frac{\partial p_s}{\partial t} \leq 0$ and $\frac{\partial p_b}{\partial t} \geq 0$. Hence type monotonicity, and the implicit function theorem imply that, for each $t \geq t_i^E(s_i)$, $\tilde{v}_b(v_s,t)$ is a well-defined differentiable function of time, with $0 \leq \left|\frac{\partial \tilde{v}_b(v_s,t)}{\partial t}\right| < \infty$. In other words, at every $t \geq t_s^E(v_s)$ there exists exactly one type $\tilde{v}_b(v_s,t)$ of the buyer, with whom v_s would reach agreement at that moment. These facts have two consequences. First, the probability of reaching an agreement by t, $H_b(t)$, has no mass points because the distribution of types of the buyer has no mass points. Second, the marginal increase in $H_b(t)$, $dH_b(t)$, is equal to the marginal increase of the mass of buyer's types that the seller would agree with by moment t. Also, the seller knows that before $t_s^E(v_s)$ her bids were unrealistic, so she cannot update her beliefs until that moment. Since $\tilde{v}_b$ is differentiable with respect to time, the beliefs are updated continuously and differentiably from $t_s^E(v_s)$ on. In other words, we have established that at $t_s^E(v_s)$ the belief of the seller is exactly $G_b(v_b)$, and at every moment $dH_b(t) = dG_b(\tilde{v}_b(v_s,t)) = g_b(\tilde{v}_b(v_s,t))\frac{\partial \tilde{v}_b(v_s,t)}{\partial t}dt$. This completes the proof for the seller. The case of the buyer is analogous. $\square$

The optimization problem stated in Proposition 2.7.5 can be best approached as a problem where i is choosing two unknown functions $p_i(v_i,\cdot)$ and $\tilde{v}_j(v_i,.)$ which are bound by the constraint (2.8), where $p_j(\cdot,\cdot)$ is a given and fixed function (the strategies of all possible types of agent j). A good reference for the calculus of variations is Elsgolts [1973].

The optimality condition at the lower boundary of optimization is given by definition of $t_i^E(v_i)$-implicitly written as $\tilde{v}_b(v_s, t_s^E(v_s)) = 1$ or $\tilde{v}_s(v_b, t_b^E(v_b)) = 0$. In the following lemma we provide the first-order condition for the optimization program of agent i, for $t > t_i^E(v_i)$. To save on cumbersome notation we omit several unambiguous arguments in the functions.

Lemma 2.7.6. First-order condition: *In a regular and separating BE, strategies $p_i(v_i,.)$, $i = s,b$, satisfy the following first-order conditions*

$$\begin{aligned} (p_s - v_s) &= \left(\frac{\partial p_b(\tilde{v}_b,t)}{\partial \tilde{v}_b}\frac{d\tilde{v}_b}{dt} - \frac{\partial p_s}{\partial t}\right), \forall\ v_s \in [0,1], \forall\ t > t_s^E(v_s); \\ (v_b - p_b) &= -\left(\frac{\partial p_s(\tilde{v}_s,t)}{\partial \tilde{v}_s}\frac{d\tilde{v}_s}{dt} - \frac{\partial p_b}{\partial t}\right), \forall\ v_b \in [0,1], \forall\ t > t_b^E(v_b). \end{aligned} \tag{2.9}$$

Proof. We fix v_i and economize the notation to write $\tilde{v}_j\,(v_i,t)=\tilde{v}_j$ and $\frac{\partial\tilde{v}_j(v_i,t)}{\partial t}=\dot{\tilde{v}}_j$. We write the Hamiltonian

$$\begin{aligned} H_i(t) &= e^{-t}u_i(p_i(v_i,t),v_i)g_j(\tilde{v}_j)\;\dot{\tilde{v}}_j \\ &\quad -\mu(t)\left(p_j(\tilde{v}_j,t)-p_i(v_i,t)\right), \end{aligned}$$

and compute the Euler conditions for the unknown functions

$$\begin{aligned} \frac{\partial H_i}{\partial\tilde{v}_j} &= e^{-t}u_i(p_i(v_i,t),v_i)g_j'(\tilde{v}_j)\;\dot{\tilde{v}}_j-\mu\frac{\partial p_j(\tilde{v}_j,t)}{\partial\tilde{v}_j}, \\ \frac{d}{dt}\frac{\partial H_i}{\partial\,\dot{\tilde{v}}_j} &= e^{-t}u_i(p_i(v_i,t),v_i)g_j'(\tilde{v}_j)\;\dot{\tilde{v}}_j \\ &\quad +e^{-t}\frac{\partial u_i(p_i(v_i,t),v_i)}{\partial p}\frac{\partial p_i(v_i,t)}{\partial t}g_j(\tilde{v}_j) \\ &\quad -e^{-t}u_i(p_i(v_i,t),v_i)g_j(\tilde{v}_j), \\ \frac{\partial H_s}{\partial p_i} &= e^{-t}\frac{\partial u_i(p_i(v_i,t),v_i)}{\partial p}g_j(\tilde{v}_j)\;\dot{\tilde{v}}_j+\mu, \\ \frac{\partial H_s}{\partial\,\dot{p}_i} &= 0. \end{aligned}$$

Whence we have the two Euler equations

$$\begin{aligned} &-\mu\tfrac{\partial p_j(\tilde{v}_j,t)}{\partial\tilde{v}_j}-e^{-t}\tfrac{\partial u_i(p_i(v_i,t),v_i)}{\partial p}\tfrac{\partial p_i(v_i,t)}{\partial t}g_j(\tilde{v}_j)+e^{-t}u_i(p_i(v_i,t),v_i)g_j(\tilde{v}_j)=0, \\ &e^{-t}\tfrac{\partial u_i(p_i(v_i,t),v_i)}{\partial p}g_j(\tilde{v}_j)\;\dot{\tilde{v}}_j+\mu=0. \end{aligned}$$

From the second Euler equation we can eliminate μ and the density g_j also disappears from the first to obtain the condition

$$\begin{aligned} u_i\left(p_i,v_i\right) &= \frac{\partial u_i\left(p_i,v_i\right)}{\partial p_i}\left(\frac{\partial p_j(\tilde{v}_j,t)}{\partial\tilde{v}_j}\frac{d\tilde{v}_j}{dt}-\frac{\partial p_i}{\partial t}\right), \\ \text{for } t &\geq t_E\left(v_i\right), i=s,b. \end{aligned}$$

□

Lemma 2.7.6 has two important implications. The first is that a best response to a strictly type-monotone strategy is strictly type monotone.

Corollary 2.7.7. *If $p_i(.,.)$ is a best response to a regular strategy $p_j(.,.)$, such that $\frac{\partial p_j}{\partial v_j} > 0$, then $\frac{\partial p_i}{\partial v_i} > 0$.*

Proof. Let $i = s$ and assume that $\frac{\partial \tilde{v}_b}{\partial t} = 0$. By assumption, $\frac{\partial p_b(\tilde{v}_b, t)}{\partial v_b} > 0$ and $\frac{\partial p_s}{\partial t} \leq 0$, so that $\frac{\partial \tilde{v}_b}{\partial t} = 0$ contradicts equation (2.9). □

The second implication of Lemma 2.7.6 is that the equilibrium strategies must be independent of players' beliefs. Hence the following is immediate (see for instance Ledyard [1978]).

Corollary 2.7.8. *A PBE in regular and strictly type monotone strategies must be an ex-post equilibrium.*

Combining Corollary 2.7.8 with Lemma 2.7.3 yields Theorem 2.7.1.

Bibliography

[1] Admati A.R. and M. Perry [1987]: "Strategic Delay in Bargaining", *Review of Economic Studies*, **54**, 345-364.

[2] Ausubel, L. M., P. Cramton, and R. J. Denekere [2002]: "Bargaining with Incomplete Information", in *Handbook of Game Theory,* volume 3 , edited by R. J. Aumann and S. Hart, Amsterdam: Elsevier Science B. V.

[3] Ausubel, L. M., and R. J. Denekere [1992]: "Durable Goods Monopoly with Incomplete Information", *Review of Economic Studies*, **59**, 795-812.

[4] Bergemann, D. and S. Morris [2004]: "Robust Mechanism Design", mimeo.

[5] Binmore, K., A. Rubinstein, and A. Wolinsky [1986]: "The Nash Bargaining solution in economic modelling" *Rand Journal of Economics*, **17**, 176-188.

[6] Cho, I-K. [1990]: "Uncertainty and Delay in Bargaining", *Review of Economic Studies*, **57**, 575-596.

[7] Čopič, J. and C. Ponsatí [2005]: "Robust Bilateral Trade under Risk Aversion", mimeo.

[8] Čopič, J. [2005]: "Robust mechanisms for private value problems", mimeo, www.hss.caltech.edu/∼jernej.

[9] Cramton, P. [1984]: "Bargaining with Incomplete Information: An infinite horizon model with continuous uncertainty", *Review of Economic Studies*, **51**, 579-593.

[10] Cramton, P. [1992]: "Strategic Delay in Bargaining with two-sided uncertainty", *Review of Economic Studies*, **59**, 205-225.

[11] Chung, K.S. and J. Ely [2003]: "Ex-post Incentive Compatible Mechanism Design", mimeo.

[12] d'Aspremont, C. and L.A. Gerard-Varet [1979]: "Incentives and Incomplete Information", *Journal of Public Economics*, **11**, 24-45.

[13] Dunlop, J.T. [1984]: *Dispute Resolution*, Auburn House, Dover MA.

[14] Elsgolts, L. [1973] *Differential Equations and the Calculus of Variations,* Mir Publishers, Moscow.

[15] Fishburn, P.C. and A. Rubinstein, A. [1982]: "Time Preference", *International Economic Review*, **23**, 677-694.

[16] Hagerty, K. M. and W. P. Rogerson [1987]: "Robust Trading Mechanisms", *Journal of Economic Theory*, **42**, 94-107.

[17] Holmstrom, B. and R. B. Myerson [1983]: "Efficient and Durable Decision Rules with Incomplete Information", *Econometrica*, **51**, 1799-1819.

[18] Hurwicz, L. [1972]: "On Informationally Decentralized Systems", in *Decision and Organization; a volume in honor of Jacob Marschak*, ed. R. Radner and C. B. McGuire, 297-336. Amsterdam, North Holland Publishing Company.

[19] Jarque, X., C. Ponsatí and J. Sákovics [2003]: "Mediation: Incomplete Information Barganing with Filtered Communication", *Journal of Mathematical Economics*, **39**, 803-830.

[20] Jehiel, P., M. Meyer-ter-Vehn, B. Moldovanu, and W. R. Zame [2005]: "The limits of ex-post implementation", mimeo.

[21] Ledyard, J. O. [1978]: "Incentive Compatibility and Incomplete Information", *Journal of Economic Theory*, **18**, 171-189.

[22] Myerson, R. and M. A. Satterthwaite [1983]: "Efficient Mechanism for Bilateral Trading", *Journal of Economic Theory*, **29** , 265-281.

[23] Neeman, Z. [2004]: "The Relevance of Private Information in Mechanism Design", *Journal of Economic Theory* **117**. 5577.

[24] Rubinstein, A. [1982]: "Perfect Equilibrium in a Bargaining Model", *Econometrica*, **50**, 97-109.

[25] Rudin, W. *Real And Complex Analysis,* McGraw-Hill, Singapore, 1987.

[26] Sàkovics, J. [1993]: "Delay in Barganing Games with Complete Information" *Journal of Economic Theory*, **59**, 78-95.

[27] Simon, L. and M. Stinchcombe [1989]: "Extensive Form Games in Continuous Time. Part I: Pure Strategies", *Econometrica*, **57**, 1171-1214.

[28] Wang, R. [2000]: "Separating equilibria in a continuous-time bargaining model with two-sided uncertainty", *International Journal of Game Theory*, **29**, 229-240.

Chapter 3

Awareness as an Equilibrium Notion

The work presented in this chapter has been done jointly with Andrea Galeotti.

3.1 Introduction

Awareness defines the way that individuals perceive the world. In this paper, we address normal-form games, where agents may have limited awareness of strategies available to them and to others. In such world, the agents then need to conjecture the strategies that the others are aware of, conjecture the others' conjectures, and so on. We call such infinite regress an *awareness architecture*. While the modeler may be unsure about the agents' awareness architectures, he may know something about them. We describe this by the set of possible awareness architectures. Given an awareness architecture, each player adapts his play accordingly. However, in equilibrium, neither the agents' awareness architecture nor their play can change.

In Section 3.2 we define awareness as an equilibrium notion in the following way. Awareness equilibrium is a situation where the play of the game is, from the perspective of each agent, consistent with rationality of all players, and consistent with the awareness architectures. Awareness architectures need to be consistent (an agent cannot make a conjecture about a fact he is not aware of) and they have to be in the set of possible awareness architectures. The play of the game should be consistent with the awareness architectures, and with each player best responding (within the set of actions he is aware of) to the others' actions. We then provide conditions on the set of possible awareness architectures under which an equilibrium exists.

Although our approach to awareness is quite different, it is related to the epistemic models of unawareness, in particular those concerned with multi-person settings. After the seminal contribution of Dekkel et al. [1998], showing that standard state-space representation precludes non-trivial forms of unawareness, this literature has focused on providing general state-space models which are able to overcome this negative result. Recent papers on this are Heifetz et al [2005], Li [2004], Modica and Rustichini [1999]. In these models, general state-space representations of unawareness are the core; in specific situations, agents' unawareness structure is a primitive that models agents' states of minds. Feinberg [2004] provides a an approach that is more similar to ours-in particular, he also models agents' awareness as an awareness architecture. In his approach, agents awareness might change after having observed an outcome, so that we can think of this as a dynamic approach. There are two key differences between these approaches and ours. First, in our approach, the agents' awareness cannot change in equilibrium, which can be thought of as a steady state of situations when it can. Second, through the specification of the set of possible awareness architectures, we model the modeler's knowledge about the agents' awareness and conjectures.

Revealed awareness is conceptually the core of our approach and is consistent with other notions of equilibrium. Awareness equilibrium is close to the equilibrium models where off the equilibrium deviations are only conjectured but never actually observed (see Rubinstein and Wolinsky [1991], Fudenberg and Levine [1993], and Battigalli and Guaitolli [1998]). That is a very natural weakening of standard notions of equilibrium, and awareness equilibrium weakens it further by not requiring that the model itself be common knowledge. As all equilibrium models, awareness equilibrium can be thought of as characterizing steady states of dynamic processes-in these case, processes where agents adjust their actions and awareness.

In Section 3.3 we provide a simple class of sets of possible awareness architectures for which the awareness equilibrium exists. Each player may be aware of a fixed number of strategies of every player, which is a parameter of the model. We interpret the level of awareness of agents, i.e. how many actions they are aware of, as the agents' cognitive bound. We then study the equilibrium outcomes for different levels of cognitive bounds. This gives a specific method to measure awareness from observed outcomes. We demonstrate that generically the set of action-awareness equilibrium outcomes is a much larger superset of the set of Nash equilibria. As the agents' cognitive bound increases, the former set shrinks towards the latter. Nevertheless, there do not exist restrictions on games which would make these two sets coincide for a given cognitive bound while allowing for the games to have an arbitrary size of action sets. We conclude in Section 3.4.

3.2 Awareness equilibrium in normal-form games

In this section we provide a formal definition of the awareness equilibrium in normal-form games where actions might not be common knowledge. We consider the simplest case where we allow awareness to be defined only with respect to players' actions. The set of agents is common knowledge. Each agent may be aware only of some actions and the corresponding outcomes. The mapping from outcomes into payoffs is common knowledge, so that if an agent is aware of a profile of actions, he is aware of the corresponding payoffs to all agents. Here we focus on games with complete information, so that the awareness equilibrium builds on Nash equilibrium in the sense that agents play and conjecture best responses (given their awareness). We present this section in terms of 2-player finite games to make the section easier to read.

Let $\mathbf{N} = \{1, 2\}$ be the set of agents, let $\mathbf{A} = \mathbf{A}_1 \times \mathbf{A}_2$ be the set of action profiles, where $\mathbf{A}_n$ is finite for each $n \in \{1, 2\}$, $\mathbf{A}_1 = \{\underline{1}, ..., \underline{K}\}$ and $\mathbf{A}_2 = \{\bar{1}, ..., \bar{K}\}$, $a = (\underline{a}, \bar{a})$ is a typical element of $\mathbf{A}$, and $\sigma = (\underline{\sigma}, \bar{\sigma})$ is a mixed strategy profile. The set of pure-strategy outcomes corresponds to $\mathbf{A}$, and denote by $\Delta(\mathbf{A})$ the set of mixed-strategy outcomes, i.e. corresponding to lotteries over pure strategies. Payoffs over pure-strategy outcomes are represented by a mapping $u : \mathbf{A} \rightarrow \Re^2$, $u(a) =$

$(u_1(a), u_2(a)), \forall\, a \in \mathbf{A}$. Payoffs associated to a mixed strategy profile σ are $U(\sigma) = (U_1(\sigma), U_2(\sigma)) = E_\sigma[u(a)]$.

If $\mathbf{N}, \mathbf{A}, \mathbf{u}$ were all common knowledge then this would be a standard game. Players' awareness restricts the set of actions of each player, and players make conjectures about others' awareness and others' conjectures and so on, which we call an awareness architecture. Denote by $A^{(n)} = A_1^{(n)} \times A_2^{(n)} \subset \mathcal{A}$ the action-awareness of player n, so that $A_1^{(n)}$ are actions of player 1 that player n is aware of, $n \in \{1, 2\}$. A player is aware of $u(a) = (u_1(a), u_2(a))$ if and only if he is aware of a. A first-order conjecture of agent n about m's awareness $A^{(m)}$ is denoted by $A^{(n,m)} = A_1^{(n,m)} \times A_2^{(n,m)}$, and so on. Define the awareness architecture of agent n by $c_n = (A^{(n)}, A^{(n,m)}, A^{(n,m,n)}, ...)$, $n, m \in \{1, 2\}$. The set of all possible awareness architectures for player n is $C_n \subset \{0, 1\}^A \times \{0, 1\}^A \times ...$ and the space of possible awareness architectures is $\mathbf{C} = C_1 \times C_2$. A normal form game with action awareness is $\mathcal{U} = (\mathbf{N}, \mathbf{A}, \mathbf{u}, \mathbf{C})$.

For a finite sequence k of length x let (k, n) be a sequence of length $x + 1$, such that $(k, n)_i = k_i$ for each $i \leq x$ and $(k, n)_{i+1} = n$.

Definition 3.2.1. The *Action-awareness equilibrium* (AAE) is an outcome $\sigma \in \Delta(\mathbf{A})$ and awareness architectures $(c_1, c_2) \in C_1 \times C_2$ such that

AAE1 $A_l^{(k,m)} \subset A_l^k$, and if $k = (n, ..., m)$, then $A^{(k,m)} = A^k$, $\forall\, l \in \mathbf{N}$, for all $k \in \mathbf{N}^x, \forall\, x < \infty$, $n, m \in \mathbf{N}$.

AAE2 $a \in A^k, \forall\, k \in \mathbf{N}^x, \forall\, x < \infty$, $\forall\, a \in \mathbf{supp}(\sigma)$.

AAE3 $\underline{\sigma} = \arg\max_{\underline{\sigma}' \in \Delta(A_1^k)} U_1(\underline{\sigma}', \bar{\sigma})$, $\bar{\sigma} = \arg\max_{\bar{\sigma}' \in \Delta(A_2^k)} U_2(\underline{\sigma}, \bar{\sigma}'), \forall\, k \in \mathbf{N}^x, \forall\, x < \infty$.

An AAE is a situation where the agents' perception of the world is internally consistent (AAE1), consistent with the outcome (AAE2), and consistent with the aspects that are common knowledge (AAE3). The requirement AAE1 is that agents cannot reason about facts that they are not aware of. For example, if player 1 is not aware of action $\underline{a}$, then he cannot conjecture that player 2 is aware of that action.

This is very different from knowledge, where a player may not know a fact, but is allowed to make conjectures about this fact. AAE1 also requires that if 1 is aware of some actions, he cannot conjecture otherwise about himself. AAE2 requires that in equilibrium the players are aware of the action profile that is realized, and correctly conjecture that others are aware of that action profile, and so on at all orders of conjectures. AAE3 requires that the action profile that obtains is consistent with agents' optimization, at every order of conjectures.[1]

$\mathcal{U}$ is a different and more complex object than the standard game $\Gamma = (\mathbf{N}, \mathbf{A}, \mathbf{u})$. Nonetheless there is a relationship between Nash equilibria of Γ and AAE of $\mathcal{U}$. Nash equilibria of Γ are AAE of $\mathcal{U}$ that are not sensitive to the details of the specification of $\mathbf{C}$. Observe that in general, by virtue of AAE1, for every $\mathbf{C}$, we can restrict attention to $\mathbf{C}_E \subset \mathbf{C}$, such that AAE1 holds for every element of $\mathbf{C}_E$. We from now on restrict attention to architecture spaces $\mathbf{C}$ such that $\mathbf{C}_E \neq \emptyset$.[2]

Proposition 3.2.2. *Given $\Gamma = (\mathbf{N}, \mathbf{A}, \mathbf{u})$, the profile σ is a Nash equilibrium of Γ if and only if it is supportable in AAE for every $\mathcal{U} = (\mathbf{N}, \mathbf{A}, \mathbf{u}, \mathbf{C})$, such that there exist $(c_1, c_2) \in \mathbf{C}_E$ with $\mathbf{supp}(\sigma) \subset \cap_{x<\infty} \cap_{k \in N^x} A^k$.*

Proof. We provide the proof for pure strategies, the proof for mixed strategies is analogous. Let $(\underline{a}, \bar{b})$ be a Nash equilibrium of Γ and suppose $\exists$ $(c_1, c_2) \in \mathbf{C}_E$ such that $(\underline{a}, \bar{b}) \in \cap_{x<\infty} \cap_{k \in N^x} A^k$. Since $(\underline{a}, \bar{b})$ is a Nash equilibrium there are no profitable deviations to either of the two players even if their action sets are restricted, so that AAE3 is satisfied. AAE1 and AAE2 are satisfied by assumption. For the converse, $(\underline{a}, \bar{b})$ is supportable on the architecture space $\mathbf{C}$, where $A^{(1)} = A^{(2)} = A$, in which case players must be playing a Nash equilibrium by AAE3. $\square$

The above proposition states that a Nash-equilibrium profile is the only one for which players can make any conjectures that are internally consistent (in the sense of AAE1), and consistent with the given profile (in the sense of AAE2), and such

[1] This could be restated into saying that the agents must conjecture that at every order of awareness, each player is playing a best reply to the actions of the other players.

[2] It is very easy to provide examples of $\mathbf{C}$ such that $\mathbf{C}_E = \emptyset$. For instance, that is true if $A^{(n)} \cap A^{(n,m)} = \emptyset$.

conjectures along with the action profile constitute an AAE. That is, for a strategy profile that is not a Nash equilibrium in Γ we can find an architecture space $\mathbf{C}$ such that even if this strategy profile satisfies AAE2 (is feasible) it does not constitute a part of an AAE.

We now turn to the question of existence of AAE. Existence of AAE may depend on the specification of $\mathbf{C}$. There are two very different situations at the opposite extremes of the possible specifications of $\mathbf{C}$. The first situation is one where $C_n = \{0,1\}^A \times \{0,1\}^A \times ....$ This corresponds to the case where an omniscient outside observer sees the game, but has no indication on the agents' awareness of the game. In this case, existence is not an issue, since for instance the outcomes associated with Nash equilibria of Γ will be supported in AAE of $\mathcal{U}$. However, in this case, every outcome will be supportable in AAE-simply take $A^k = \{\underline{a}, \bar{b}\}, \forall\, k$. At the other extreme is the situation where $C_n = \{0,1\}^{\{\underline{a},\bar{b}\}} \times \{0,1\}^{\{\underline{a},\bar{b}\}} \times ..., n \in \mathbf{N}$, then the outcome corresponding to $\{\underline{a}, \bar{b}\}$ is the unique outcome supportable in AAE, but this is a very restrictive case where agents' awareness is trivial. The interesting cases are somewhere in between, where some restriction on $\mathbf{C}$ is exogenously specified. For example, an experimenter *tells* each player something about A, in which case $C_n \subset \{0,1\}^A \times \{0,1\}^A \times ...$, where $A^{(n)}$ has to equal to what player n was told. The following example illustrates that in such situation, an AAE may fail to exist.

Example 3.2.3. Let Γ be described by the following normal-form representation.

$_1\backslash^2$	$\bar{1}$	$\bar{2}$
$\underline{1}$	$6,4$	$8,7$
$\underline{2}$	$5,9$	$\mathbf{10,10}$

Observe that Γ has a unique pure-strategy Nash equilibrium, $(\underline{2}, \bar{2})$. If $\mathbf{C}$ is such that $A^{(1)} = \{\underline{1}, \underline{2}, \bar{1}\}$ and $A^{(2)} = \{\underline{1}, \bar{1}, \bar{2}\}$, then no AAE exists. The reason is that regardless of $A^{(2,1)}$, player 2 would always play $\bar{2}$, which would violate AAE2.
In contrast, if $\mathbf{C}$ is such that $A^{(1)} = \{\underline{1}, \bar{1}, \bar{2}\}$ and $A^{(2)} = \{\underline{1}, \underline{2}, \bar{1}\}$, then $\{\underline{1}, \bar{1}\}$ can be supported in an awareness equilibrium. An awareness architecture that supports it is $A^{(12)} = A^{(21)} = A^k = \{\underline{1}, \bar{1}\}, \forall\, k$, s.t. $k \in N^x, x \geq 3, k = (n, m, ...), n \neq m$. Note

that not all awareness architectures will be equilibrium architectures, for instance if $A^{(12)} = A^{(1)}$, no such AAE will exist.
Finally, we remark that if $\mathbf{C}$ is such that $A^{(1)} = A^{(2)} = A$ then the only AAE outcome is the Nash-equilibrium outcome of Γ.

It is natural to ask what sets of possible awareness architectures will give existence of AAE.

Proposition 3.2.4. *Given* $\mathcal{U} = (\mathbf{N}, \mathbf{A}, \mathbf{u}, \mathbf{C})$*, an AAE exists, if and only if there exists* $(c_1, c_2) \in \mathbf{C}_E$*, and* $\exists\ \sigma$, $\mathbf{supp}(\sigma) \subset \cap_{x=1,2,\ldots} \cap_{k \in \mathbf{N}^x} A^{(k)}$*, with* $\underline{\sigma} = \arg\max_{\underline{\sigma}' \in \Delta(A^k)} U_1(\underline{\sigma}', \bar{\sigma})$*, for* $k \in \{(1), (21)\}$ *and* $\bar{\sigma} = \arg\max_{\bar{\sigma}' \in \Delta(A^k)} U_2(\underline{\sigma}, \bar{\sigma}')$*, for* $k \in \{(2), (12)\}$*.*

Proof. The only if part follows from the fact that if such (c_1, c_2) did not exist, then for every outcome satisfying AAE2, there would be a player $n \in \mathbf{N}$, such that either n would deviate given $A^{(n)}$, or $m \neq n$ would deviate under n's conjecture $A^{(n,m)}$. In either of these cases, AAE3 is violated.
The if part follows from AAE1. If n does not have a profitable deviation under $A^{(n)}$ and under $A^{(m,n)}$, then he does not have a profitable deviation under S, $\forall\ S \subset A^{(n)}$ nor under P, $\forall\ P \subset A^{(m,n)}$, so that the claim follows by AAE1. $\square$

Proposition 3.2.4 shows that generically, a restriction on $\mathbf{C}$ for which AAE will exist, will be much stronger than just requiring that there exist an outcome in the intersection of all the players' conjectures. There must exist such outcome, which is also consistent with players' own optimization, and the first-order conjecture that the other agent optimizes.

Proposition 3.2.4 also illustrates that admissible restrictions on $\mathbf{C}$ in general depend on the specification of Γ. As we noted earlier, one class of such restrictions on $\mathbf{C}$ is to impose the awareness of the agents.[3] Another possibility is to restrict the number of actions that a player can be aware of, but not which actions these are. Such

[3]One could also consider the possibility that the experimenter also tells the players what he told the other player, and possibly lies about that, but this might not be enough to control the players' conjectures if they see a reason not to trust the experimenter. This consideration does not apply to their awareness, since if the experimenter tells them something he is sure that they are aware of it.

restriction has the interpretation that the agents have bounded cognitive abilities, but the experimenter does not control their awareness directly. For example, the subjects in an experiment might be confronted with a very large normal-form game, only a fraction of which fits on the computer screen. This approach has the advantage that the existence of AAE will not be in question, regardless of the specification of Γ. In the next section we study how many outcomes may never be supported in AAE under such restrictions on the architecture space.

3.3 A model of cognitive bounds

In this section we focus on pure strategies simply because the question of cardinality of the set of outcomes that may never be supported in AAE is much more straightforward to define. Let ℓ be the number of actions of each agent that agent i is aware of, and assume that the number of actions that agents are aware of is common knowledge.

Definition 3.3.1. Fix an $\ell \geq 1$. An ℓ-Action-awareness equilibrium, ℓ-AAE, is an AAE where $|A_l^{(n)}| = \ell$, and this is common knowledge.

By Proposition 3.2.2, a Nash-equilibrium profile would be an ℓ-AAE whenever the corresponding Nash-equilibrium action profile a is in the awareness sets of both players. This observation only holds for Nash-equilibrium profiles of Γ. This motivates the next definition.

Definition 3.3.2. Fix an $\ell \geq 1$ and let $a^* = (\underline{a}^*, \bar{a}^*)$ be a pure-strategy Nash equilibrium of the game $\Gamma = (\mathbf{N}, \mathbf{A}, \mathbf{u})$. An ℓ^*-Action-awareness equilibrium, ℓ^*-AAE, is an ℓ-AAE where $a^* \in A^{(n)}, n = 1, 2$.

That is an ℓ^*-AAE, is an ℓ-AAE which is an equilibrium even if some Nash-equilibrium profile is in the awareness sets of both players.

We are interested in comparing how the sets of ℓ-AAE and ℓ^*-AAE change as we vary ℓ. Such comparison is useful for providing a measure of how strengthening the restriction on the architecture space strengthens the equilibrium notion. This comparative-statics approach also provides a method of estimating the cognitive bound ℓ of the agents. In the absence of other considerations, if a certain outcome is observed, then ℓ has to be low enough, in order to support that outcome as an ℓ-AAE. The following result simplifies our analysis.

Lemma 3.3.3. *A profile of actions a is an ℓ-AAE if and only if it is an ℓ-AAE with $A^k = A^{k'}, \forall\ k \in N^x, \forall\ x < \infty$.*

Proof. The if part is trivial and we omit it. The only if part is as follows. Let a be an ℓ-AAE outcome under *some* awareness architecture, different from those specified in the claim. This implies that there at least $\ell - 1$ deviations by each player to which a is a best reply. But then a can be supportable also with some awareness architecture where all agents are aware of the same actions and make the correct conjectures about others. □

We will further restrict our analysis to generic games. Namely, it is possible to construct non-generic and non-trivial games where there is a unique Nash equilibrium in the game $(\mathbf{N}, \mathbf{A}, \mathbf{u})$, but for every $\ell < K$ every outcome can be supported in an ℓ^*-AAE. We illustrate this with the next example.

Example 3.3.4. For each K there exists a $\Gamma = (\mathbf{N}, \mathbf{A}, \mathbf{u}), |A_n| = K, \forall\ n \in \{1, 2\}$, such that the following holds. Γ has a unique pure-strategy Nash equilibrium, let $(\underline{1}, \bar{1})$ be the unique NE. Then for each $\ell, 2 \leq \ell < k$, every outcome is sustainable as an l^*-AAE.

To see this, consider the following game. To define u, take first the matrix for the row player, $u_1(\underline{p}, \bar{q}), 1 \leq p, q \leq K$. Let $u_1(\underline{1}, \bar{q}) = 1, u_1(\underline{q}, \bar{1}) = 0, q = 1, ..., K$. For each column $p = 2, ..., K$, assign a 1 in precisely one unassigned location in such a way that the assigned 1's do not lie in only one row. This can obviously be done. Let $u_1(\underline{p}, \bar{q}) = 0$ for all the other locations. Take player 2 and do exactly the same, but also take care so that $(u_1(\underline{p}, \bar{q}), u_2(\underline{p}, \bar{q})) \neq (1, 1)$ for $(\underline{p}, \bar{q}) \neq (\underline{1}, \bar{1})$. Since the 1s assigned to columns of player 1 are not in the same row, such assignment is possible (reader can easily verify that). See Figure 2 for an example of such a game.

$_1\backslash^2$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\underline{1}$	$1,1$	$1,0$	$1,0$
$\underline{2}$	$0,1$	$1,0$	$0,1$
$\underline{3}$	$0,1$	$0,1$	$1,0$

Figure 2

Now we have to show that u has the desired properties. Clearly, the profile $s = (\underline{1}, \bar{1})$ is a pure-strategy NE of Γ. To show that this is the unique pure-strategy NE of Γ observe that for every $(p, q) \in \{1, ..., K\}^2, (p, q) \neq (1, 1)$, at least one player gets a 0. Suppose (wlog) it is the row player 1. Then, by construction there is another column q' such that $u_1(\underline{p}, \bar{q}') = 1$ so that 1 would want to deviate.

To show that every outcome is an ℓ^*-AAE, for all $\ell < K$, observe first that if an outcome is an ℓ^*-AAE, $\ell > 2$ then it must be an $(\ell-1)^*$-AAE (reduce the supporting awareness set of each player by one action). Thus, it is enough to show the claim for $\ell = K - 1$. So take an outcome $(\underline{p}, \bar{q}) \in \{1, ..., k\}^2, (p, q) \neq (1, 1)$, and suppose that $u_1(\underline{p}, \bar{q}) = 0$ (if it is 1, then there is no deviation for player 1 anyway). This is not column 1, since there player 1 gets 1. By construction there are $K - 2$ other rows in column p such that player 1 gets 0 in those rows, and taking those $K - 2$ rows and row p also includes row 1. Similarly for player 2, so that we have constructed the awareness sets which include action 1 for both players, and no player has a profitable deviation from the profile $(\underline{p}, \bar{q})$.

We say that a game is *generic* if the following *no-indifference* condition holds. Let Γ be a $K \times K$ game. We say that Γ satisfies the no-indifference condition if $u_1(\underline{p}, \bar{q}) \neq u_1(\underline{p}', \bar{q}), \forall\ p \neq p', \forall\ q$ (and similarly for player 2).

Theorem 3.3.5. *Let Γ be a generic $K \times K$ game. Denote by $e_\ell(\Gamma)$ the number of distinct ℓ-AAE outcomes of Γ, for each $\ell \in \{1, ..., K\}$. Then $K^2 - 2(\ell - 1)K \leq e_\ell(\Gamma) \leq K^2 - (\ell - 1)K$, $\forall\ \ell \in \{1, ..., K\}$.*

Proof. See the Appendix. □

As is evident from the proof, the bounds in Theorem 3.3.5 are tight. Theorem 3.3.5 shows that as ℓ increases, in a generic game the set of ℓ-AAE shrinks, which is not too surprising. In particular, when ℓ converges to K, the set of ℓ-AAE generically converges to the set of Nash equilibria of Γ. However, it is a simple corollary that when ℓ is substantially smaller than K, the set of ℓ-AAE is strictly larger than the set of Nash equilibria.

Corollary 3.3.6. *Let $K \geq 3$, and let Γ be a generic $K \times K$ game, then the set of ℓ-AAE outcomes of Γ is a strict superset of the set of pure-strategy Nash-equilibrium outcomes of Γ, $\forall\ \ell \leq \frac{K}{2}$.*

Proof. A generic $K \times K$ game can have at most K pure-strategy Nash-equilibrium outcomes, and the claim follows. □

Theorem 3.3.5 says nothing about the bounds on the number of ℓ-AAE relative to the number of pure-strategy Nash equilibria. When pure-strategy Nash equilibria exist, the lower bound on the number of ℓ-AAE may be in some cases improved, since every Nash equilibrium is also an ℓ-AAE, for all ℓ. Nevertheless, the main point of Theorem 3.3.5 is that it is impossible to provide general conditions on game forms which would assure that the set of ℓ-AAE equals the set of Nash equilibria, without tying ℓ to K.

In contrast, the number of ℓ^*-AAE is linked to the number of pure-strategy Nash equilibria. The lower bound on the number ℓ^*-AAE is always equal to the number of Nash equilibria, regardless of how large ℓ is. In the following result we denote by $floor[x]$ the largest integer that is smaller than $x \in \Re$, and by $mod[y, r]$ the leftover from integer division of an integer y with integer r.

Theorem 3.3.7. *Let Γ be a generic $K \times K$ game. Denote by $e_N(\Gamma)$ the number of pure-strategy Nash equilibria of Γ and by $e_{\ell^*}(\Gamma)$ the number of ℓ^*-AAE of Γ. Then $e_N(\Gamma) \leq e_{\ell^*}(\Gamma) \leq floor[\frac{K-1}{K-\ell+1}](K-\ell+1)^2 + (mod[K-1, K-\ell+1])^2 + 1$.*

Proof. See the Appendix. □

The upper bound as stated in Theorem 3.3.7 is independent of the number of Nash equilibria. However, if a generic game has a unique Nash equilibrium this imposes additional structure on the game, and the upper bound may never be attained. We illustrate this with an example of potential games.[4] Potential games are a very natural class to consider since a subgame of a potential game is also a potential game, and every potential game has at least one pure-strategy Nash equilibrium. Many commonly studied games are potential games, e.g., prisoners' dilemma, congestion games, or Cournot games with quasilinear demand.

Example 3.3.8. Let Γ be the following 4×4 potential game with $e_N(\Gamma) = 2$, given by the following matrix P.

P	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$
$\underline{1}$	10	0	2	6
$\underline{2}$	1	3	5	2
$\underline{3}$	2	4	6	3
$\underline{4}$	5	8	7	9

Figure 4

Let $\ell = 2$, so that by Theorem 3.3.7 the upper bound on $e_{\ell^*}(\Gamma) = 10$. Clearly, the 9 right-lower-corner outcomes of Γ along with the left-upper-corner Nash equilibrium constitute the set of 2^*-AAE of Γ, so that the upper bound is tight in this case. Also note that it is easy to extend the example to general potential games with at least 2 Nash equilibria and different ℓ.

Consider now the potential game Γ with a unique pure-strategy Nash equilibrium, given by the matrix $\tilde{P}$ below. The unique Nash equilibrium of Γ is the profile $(\underline{1}, \bar{1})$.

[4]A game Γ is an *ordinal potential game* if there exists a potential function $P : \mathbf{A} \to \Re$ which represents Γ in the following way: $u_1(\underline{p}, \bar{q}) - u_1(\underline{p}', \bar{q}) > 0 \iff P(\underline{p}, \bar{q}) - P(\underline{p}', \bar{q}) > 0$, and $u_2(\underline{p}, \bar{q}) - u_2(\underline{p}, \bar{q}') > 0 \iff P(\underline{p}, \bar{q}) - P(\underline{p}, \bar{q}') > 0, \forall\, \underline{p}, \underline{p}' \in \mathbf{A}_1, \forall\, \underline{q}, \underline{q}' \in \mathbf{A}_2$. A profile $(\underline{p}, \bar{q})$ is a pure-strategy Nash equilibrium of Γ if and only if $P(\underline{p}, \bar{q}) \geq \max\{P(\underline{p}', \bar{q}); p' = 1, ..., K\} \cup \{P(\underline{p}, \bar{q}'); q' = 1, ..., K\}$. In particular the maximum of all elements of matrix P is a pure-strategy Nash equilibrium of Γ.

P	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$
$\underline{1}$	10	9	2	6
$\underline{2}$	1	3	5	2
$\underline{3}$	2	4	6	3
$\underline{4}$	9	8	7	4

Figure 5

Now observe that the path of best replies from each profile $(\underline{p}, \bar{q})$ eventually ends up in $(\underline{1}, \bar{1})$. This is in fact a property of potential games with a unique Nash equilibrium. At some point such path enters either column or row 1, suppose that the path enters column 1 in row p (in $\tilde{P}$, $p = 3$). But this implies that $P(\underline{p}, \bar{1}) > P(\underline{p}, \bar{q}), q > 1$, so that no element in row p can be sustainable as an ℓ^*-AAE outcome, which means that the upper bound may never be attained in a potential game with a unique Nash equilibrium. Nonetheless, the additional structure imposed by uniqueness of Nash equilibrium eliminates only one additional row (or column), so that in a large game this effect is negligible.

3.4 Conclusion

Here we presented a model of awareness equilibrium for normal-form games, when the set of players is common knowledge. A natural extension, allowing us to model Bayesian games, is to model situations where agents may be unaware of other agents. The awareness of agents can then be represented as an infinite collection of directed networks. One question that one may ask in that model is whether in equilibrium an agent may be unaware of the moves of nature. Our conjecture is that whenever moves of nature are payoff relevant to the agent in equilibrium, the agent has to be aware of the moves of nature.

Appendix

Proof. (Theorem 3.3.5) By Lemma 3.3.3 we can focus on ℓ-AAE with the property that agents are aware of the same actions and make correct conjectures. Let player 1 be the row player. Given an ordered set S, Denote by $S_{(r)}$ the r th order statistic of S.

Step 1. A profile $(\underline{p}, \bar{q})$ is supportable as an ℓ-AAE if and only if

$$u_1(\underline{p}, \bar{q}) \geq \{u_1(\underline{1}, \bar{q}), ..., u_1(\underline{K}, \bar{q})\}_{(\ell)} \text{ and } u_2(\underline{p}, \bar{q}) \geq \{u_1(\underline{p}, \bar{1}), ..., u_2(\underline{p}, \bar{K})\}_{(\ell)}. \quad (3.1)$$

For illustration, suppose first that $\ell = 2$. Under genericity, the claim is that a strategy profile $(\underline{p}, \bar{q})$ is then sustainable as a2-AAE if and only if

$$u_1(\underline{p}, \bar{q}) > \min_{p' \in \{1,...,K\}} u_1(\underline{p}', \bar{q}) \text{ and}$$

$$u_2(\underline{p}, \bar{q}) > \min_{q' \in \{1,...,K\}} u_2(\underline{p}, \bar{q}').$$

To see the only if part, suppose that $u_1(\underline{p}, \bar{q}) = \min_{p' \in \{1,...,K\}} u_1(\underline{p}', \bar{q})$. By genericity of Γ it is therefore $u_1(\underline{p}, \bar{q}) < u_1(\underline{p}', \bar{q}), \forall\ p' \neq p$. This implies that regardless of what other row p' comprises $A^{(1)}$, player 1 will at the profile $(\underline{p}, \bar{q})$ deviate to $\underline{p}'$.

To see the if part suppose that a profile $(\underline{p}, \bar{q})$ satisfies the above condition. There exist a $p' \neq p$ and a $q' \neq q$ such that $u_1(\underline{p}, \bar{q}) > u_1(\underline{p}', \bar{q})$ and $u_1(\underline{p}, \bar{q}) > u_1(\underline{p}, \bar{q}')$. Let $A^{(1)} = A^{(2)} = \{\underline{p}, \underline{p}', \bar{q}, \bar{q}'\}$, and $(\underline{p}, \bar{q})$ is a 2-AAE outcome supported by such awareness structure. Similarly, we prove the claim for general ℓ. Note that we do not need genericity in this step. End of Step 1.

By genericity of Γ, there exists a strict ordering of 1's payoffs in each column, and a strict ordering of 2's payoffs in each row.

Step 2. $e_\ell(\Gamma) \geq K^2 - 2(\ell - 1)K$.

Fix an $\ell \in \{1, ..., K\}$. By Step 1, we will minimize the number of outcomes that can be supported under ℓ-AAE by "optimally" assigning the $\ell - 1$ lowest payoffs to player 1 in each column and $\ell - 1$ lowest payoffs to player 2 in each row. An allocation which minimizes the number of outcomes supportable as ℓ-AAE is one where all these payoffs are allocated to different profiles. Since there are K columns, $\ell - 1$ worse payoffs to 1 in each column, K rows, and $\ell - 1$ worse payoffs to 2 in each row, there are in total at most $2K(\ell-1)$ action profiles that can be eliminated. This gives the desired lower bound on $e_\ell(\Gamma)$.

Step 3. $e_\ell(\Gamma) \leq K^2 - (\ell-1)K$.

Fix $\ell \in \{1, ..., K\}$. By Step 1, we will maximize the number of outcomes by allocating the $\ell - 1$ lowest elements of each row and each column in a way which takes the least space in the game matrix. That is achieved for instance by having every outcome which is the worst payoff in a given row for the column player to also be the worst payoff in the given column for the row player. Since there are K rows and columns and there are by genericity $\ell - 1$ strictly worst payoff in each, we can thus eliminate at least $K(\ell - 1)$ outcomes, which gives the desired upper bound on $e_\ell(\Gamma)$. □

Proof. (Theorem 3.3.7) The lower bound is a consequence of the following simple Lemma.

Lemma 3.4.1. $e_N(\Gamma) = e_{\ell^*}(\Gamma)$ *if and only if the following condition holds. For every profile* $(\underline{p}, \bar{q})$ *and every Nash-equilibrium profile* $(\underline{p}^*, \bar{q}^*)$*, either* $u_1(\underline{p}, \bar{q}) \leq u_1(\underline{p}^*, \bar{q})$ *or* $u_2(\underline{p}, \bar{q}) \leq u_2(\underline{p}, \bar{q}^*)$*.*

Proof. The if part is obvious:regardless of what Nash equilibrium is taken along with a strategy profile $(\underline{p}, \bar{q})$, one of the players has incentives to deviate (also by genericity) to the Nash-equilibrium strategy.

To see the only if part, take a profile $(\underline{p}, \bar{q})$ and suppose there exists a Nash equilibrium $(\underline{p}^*, \bar{q}^*) \neq (\underline{p}, \bar{q})$ such that the above condition does not hold. Take $A^{(1)} = A^{(2)} = \{\underline{p}, \underline{p}^*, \bar{q}, \bar{q}^*\}$ and it is clear that $(\underline{p}, \bar{q})$ is an ℓ^*-AAE profile for $\ell = 2$. □

The upper bound is constructed via a "geometric" argument. Fix an $\ell, 2 \leq \ell < K$, and we show by induction on ℓ and K that $e_{\ell^*}(\Gamma) \leq floor[\frac{K-1}{K-\ell+1}](K - \ell + 1)^2 + (mod[K - 1, K - \ell + 1])^2 + 1$. Consider first a Γ, such that $e_N(\Gamma) = 1$, and assume without loss of generality that $(\underline{1}, \bar{1})$ is the Nash-equilibrium profile.

Suppose first that $\ell = 2$. Then we can for every K do the following. By genericity of Γ all the outcomes in the row 1 and column 1 cannot be sustained as ℓ^*-AAE. Also, without loss of generality, we make a construction where as many ℓ^*-AAE profiles as possible are concentrated in the lower right hand corner of the game bimatrix. Consider a profile $(\underline{K}, \bar{K})$. This profile can be supported if in row K there is 1 outcome which is worse for player 1, and in column K there is 1 worse outcome for player 2. Moreover, $(\underline{1}, \bar{K})$ and $(\underline{K}, \bar{1})$ have to be worse for the corresponding player (since $\{\underline{1}, \bar{1}\} \subset A^{(i)}$ by definition of ℓ^*-AAE). By the same logic, all other outcomes in Kth row and Kth column can be sustained. Similarly, in all the rows $K-1, ..., 2$ the first outcome cannot be sustained but all the others can. The same applies to the columns.

Now let $2 < \ell < K$. Exactly as before, the outcomes in rows $\ell, ..., K$ and columns $\ell, ..., K$ are sustainable. If $2\ell - K - 1 \leq 1$ then all the outcomes in rows $2, ..., \ell - 1$ and columns $2, ..., \ell - 1$ can also be sustained as ℓ^*-AAE by making them higher than $\ell - 1$ outcomes in the succeeding rows and columns. In the first row and column only the Nash equilibrium is sustainable.

If $2\ell - K - 1 > 1$, then consider the game Γ' obtained by taking the first $\ell - 1$ rows and $\ell - 1$ columns of Γ and let $\ell' = 2\ell - K - 1$. Now, the outcomes of Γ' that are sustainable as ℓ^*-AAE of Γ must be sustainable as $(\ell')^*$-AAE of Γ', so that $e_{\ell^*}(\Gamma) \leq (K - \ell + 1)^2 + e_{(\ell')^*}(\Gamma')$. The claim now follows from induction.

Note that the assumption that Γ has a unique Nash equilibrium was made only for convenience, since if there are more Nash equilibria, we can first rearrange the players' actions so that all of those lie on the diagonal. $\square$

Bibliography

[1] Battigalli P. and P. Guaitoli [1988]: "Conjectural Equilibrium", mimeo.

[2] Dekel B., B. Lipman, and A. Rustichini [1998]: "Standard state-space models preclude unawareness", *Econometrica*, **66**, 159-173.

[3] Fagin R. and J. Y. Halpern [1988]: "Belief, awareness and limited reasoning", *Artificial Intelligence*, **34**, 39-76.

[4] Feinberg Y. [2005]: "Games with Incomplete Awareness", mimeo. Stanford Research Paper 1894.

[5] Fudenberg D. and D. K. Levine [1993]: "Self-Confirming Equilibrium", *Econometrica*, **61**, 523-545.

[6] Halpern J. Y. [2001]: "Alternative semantics for unawareness", *Games and Economic Behavior*, **37**, 321-339.

[7] Heifetz A., M. Meier, and B. Schipper [2004]: "Interactive Unawareness", *Journal of Economic Theory*, forthcoming.

[8] Kawamura E. [2004]: "Competitive Equilibrium with Unawareness in Economies with Production", *Journal of Economic Theory*, forthcoming.

[9] Li J. [2003]: "A generalized correspondence model with unawareness", mimeo., University of Wisconsin-Madison.

[10] Modica S. and A. Rustichini [1999]: "Unawareness and partitional information structures", *Games and Economic Behavior*, **27**, 265-298.

[11] RUBINSTEIN A. AND A. WOLINSKY [1990]: "Rationalizable Conjectural Equilibrium Between Nash and Rationalizability", mimeo.