The Evolutionary Role of Toughness in Bargaining

Aviad Heifetz* Ella Segev†

July 2001

Abstract

The experimental evidence on the “endowment effect” (Kahneman et al. 1990) and the “self serving bias” in negotiations (Babcok and Loewenstein 1997) suggests that individuals enter a tough state of mind when they have to make a stand vis-a-vis somebody else. In this work we show how a toughness bias in bargaining may indeed be evolutionary viable. When the inherent toughness of the bargainer is observed by the opponent, this opponent will adjust his behavior accordingly, in a way which may enhance the *actual* payoff of the biased bargainer. Suppose, then, that a population consists initially of individuals with different inherent degrees of toughness or softness. They are often matched at random to bargain, and biases which are objectively more successful tend to appear more frequently in the society. We show how under various bargaining protocols with asymmetric information, the population will consist, asymptotically, of individuals with some moderate degree of toughness.

JEL Classification Code.: C78

Keywords: bargaining, asymmetric information, endowment effect, self-serving bias, toughness, evolution of preferences, payoff-monotonic dynamics

---

[*] The Division of Humanities and Social Sciences, California Institute of Technology and The School of Economics, Tel Aviv University. heifetz@post.tau.ac.il. Support from the Hammer fund for Economic cooperation is gratefully acknowledged.

[†] The Faculty of Management, Tel Aviv University, segeve@post.tau.ac.il

1 Introduction

In the course of bargaining, rational players manipulate their offers so as to end up with eventual gains. With asymmetric information, inefficiencies typically arise, in the form of the failure to strike mutually profitable agreements in some of the cases in which they are feasible, or in the form of long, costly delays before an agreement is reached (Myerson and Satterthwaite 1983).

However, such inefficiencies do not emerge solely from rational strategic behavior under asymmetric information. Casual observation suggests that *emotions* play a role in the course of bargaining in a large variety of contexts – from daily disputes over small issues to international conflicts between nations.

Babcok and Loewenstein (1997) survey the indications for a "self-serving" or "toughness" bias that appears in laboratory experiments of negotiations. They describe how subjects' predictions about the judge's decision in an actual tort case were biased in favor of the side – defendant or plaintiff – they were asked to play. These biases were diminished when the subjects were made aware that they might be biased, and asked to list the weaknesses in their side's arguments. In negotiations for an "out-of-court settlement", such subjects had a much higher rate of success in reaching an agreement relative to individuals who did not go through this process of considering the opponent's view. A sharp increase in the settlement rates was also achieved when the subjects were asked to form their opinion about the fair outcome or the judge's decision *before* they were told the side they will represent in the bargaining phase.

Toughness biases are well documented even in non-strategic settings. Experiments in which potential buyers and sellers have to post a reserve price for an object (Kahneman et al. 1990) suggest that most subjects tend to make a tough stand (particularly, sellers exhibit an "endowment effect") even when the market is almost competitive and a bias is self-harming, resulting in the loss of opportunities for profitable trade. These results were duplicated for a large variety of objects – from coffee mugs to fishing licenses.

In this study we propose an evolutionary perspective on the role of toughness in bargaining. We show how a toughness bias may be evolutionary viable. When the inherent toughness of the bargainer is observed by the opponent, this opponent will adjust his behavior accordingly, in a way which may enhance the *actual* payoff of the biased bargainer. We find that for various bargaining procedures under asymmetric information, this positive effect more than compensates for the losses due to the departure of the biased bargainer from cold-blooded, optimizing behavior. Assume, therefore, an evolutionary process in which individuals are often matched at random to bargain, and biases which are (objectively) more successful on average tend to appear more frequently in the society. We show how under various bargaining protocols, the distribution of biases will converge towards some monomorphic population of moderately tough individuals, whatever is the initial distribution of biases in the society[1].

---

[1] Provided only that the support of the initial distribution is an interval of toughness and possibly softness levels.

These findings shed light on the above empirical observations. It suggests that individuals may be unconsciously "programmed" to be in a tough mood when they have to make a stand, *on top of any rational tough posture they may consciously and strategically like to exhibit.* In particular, it suggests that an endowment effect (or more generally a discrepancy between the willingness to pay for some good and the willingness to accept money in exchange for it) may appear in experiments when the subjects are framed into this tough mood by the need to post a price, which contrasts with the usual competitive setting in which a price is already posted and agents are price-takers. Indeed, in experiments in which subjects are asked to *hypothesize* their behavior had they been given an object and required to ask a price for it (Loewenstein and Adler 1995), to make valuation decisions *on behalf of somebody else* (Marshall et al. 1987), or to compare alternative compensation schemes – with or without the object – for *their own private benefit* (Arlen et al. 2001), the endowment effect diminishes or disappears. Presumably, such settings do not trigger the appearance of a tough mood, because they are different enough from a strategic confrontation in which such a mood may ultimately confer an advantage.

Furthermore, a survey of experiments on the endowment effect (Horowitz and McConnel 2000) suggests that the gap between willingness to pay and willingness to accept is highest for public and non-market goods, next highest for ordinary and private goods, and lowest for various forms of money. In this hierarchy, the uncertainty of each subject about the others' monetary valuations for the good is clearly decreasing as the good becomes more accessible and easy to substitute. This is aligned with our finding, that the absolute size of the emerging toughness bias increases monotonically with the range and variance of potential valuations that each bargainer considers as possible for the other to have.

In a trade setting, the bias determines the shift in the perceived valuation of the object in comparison with the object's true value for the trader. A seller with toughness bias ("endowment effect") ε believes and acts on the premise that her valuation of the object is higher by ε than the minimal price for which she would in fact be willing to sell the object as a price-taker in a competitive environment. Similarly, a buyer with a toughness bias τ believes and acts as if his valuation of the object is lower by τ than the maximal price he would in fact be willing to pay for the object as a price-taker.

We study the emergence of biases in a variety of bargaining procedures – with one-time offers, delayed offers or sequential offers, by one or both traders, and with asymmetric information on one or both sides. In all of these models, we find that toughness biases emerge and take over the population under any payoff-monotonic selection dynamics. For example, with the Chatterjee-Samuelson (1983) linear strategies equilibrium of the sealed-bid double auction, a toughness bias of $\frac{1}{10}$ of the valuations' interval will take over the population in the long run.

The intuition for these results is straightforward. From an ex ante perspective, inducing a little toughness leads to the loss of trade opportunities with very small gains from trade, while it increases a little the objective profit from *all* of the remaining trade opportunities. Thus, unless there are extremely many trade opportunities with very small gains from trade (so that the distribution of valuations is very skewed), and unless the inefficiency at equilibrium, due to asymmetric information, increases extremely as the ex ante gains from trade decrease slightly, the first bit of toughness "pays off". On the other hand, exaggerated toughness will clearly hinder too many trade opportunities. The toughness bias which emerges in each setting has the optimal balance between these concurring considerations.

Obviously, real-life bargaining rarely follows any structured game-theoretic procedure or its prescribed equilibria. Thus, our main message is qualitative in nature: *Toughness emerges in bargaining schemes in which the bargainers ascribe a fair chance to strike fairly profitable deals,* relative to the overall foreseeable deals in the context at hand.[2] If, indeed, this is a feature of most incidences of real-life bargaining, the tendency to toughness documented today may be due, at least in part, to the strategic advantage that it conferred upon individuals during the evolution of mankind.

In the same evolutionary framework, Huck et al. (1997) studied the emergence of an endowment effect in an exchange setting of complete information, where the outcome is the Nash bargaining solution. They showed that the zero endowment effect as well as extreme endowment effects are weeded out in the limit. However, unlike in the current work, their process is not shown to converge towards any particular bias or distribution of biases.

More generally, there are by now several works which study the evolutionary stability of various biases – fairness concerns (Güth and Yaari 1992, Huck and Oechssler 1998), social mindedness (Fershtman and Weiss 1997, 1998), altruism (Bester and Güth 1998), spitefulness (Possajennikov 2000, Bolle 2000), concern with relative success (Koçkesen, Ok and Sethi 1998). Heifetz and Spiegel (2000) study the dynamic emergence of optimism.[3] The current work is the first to study the dynamic emergence of biases in a strategic setting with asymmetric information.

The paper is organized as follows. Section 2 elaborates further the assumptions of the model and its technical ingredients. Section 3 brings two examples of bargaining protocols and the toughness biases that evolve in them. Section 4 provides general sufficient conditions for the emergence of toughness. This is done, first, under some symmetry assumptions, which make it possible to pin down the emerging bias. Subsequently, we deal with more general bargaining procedures, for which we characterize when an inclination to toughness will evolve, even if the evolutionary dynamics may remain in flux, and not converge to one specific degree of toughness (or to a distribution of biases).

In Section 5 we list various other asymmetric information bargaining procedures and solve for the toughness biases that emerge in them. In section 6 we check the robustness of the assumption that the bias of each bargainer is observed and taken as given by her opponent. We exemplify how this assumption may be relaxed in two different ways without changing our qualitative conclusions. We first consider the possibility that the

---

[2] In other words, the bargaining setting is not too extreme, in the two senses alluded to in the previous paragraph.

[3] That paper also contains further references to the literature on the evolution of preferences.

biases are observed only in some of the interactions. In a second refinement, we assume that individuals may be inclined to make a (costly) effort to appear as if they are tougher (or softer) than they really are. Also in this case, we show that a positive degree of toughness emerges, with an even tougher corresponding "mask". We conclude in section 7 with indications to further directions for research. Technical details and proofs appear in appendices.

2 The Evolutionary Model

Let there be a large population of individuals. At each point in time, a pair of individuals is matched at random to bargain. One of the players assumes at random the role of the seller and the other that of the buyer, with equal probabilities for each of the roles.

The individuals differ from one another in the degree of their inherent toughness (or softness). A seller is inherently tough if when she has to bargain, she misperceives her valuation of the object to be higher than it really is. With a toughness bias of ε, she believes and acts as if her valuation is $S = s + \varepsilon$ when the objective worth of the object for her is s (that is, when s is the minimal price for which she would be willing to sell the object as a price-taker in a competitive environment). Similarly, we say that a buyer has a toughness bias τ if when he has to bargain, he misperceives his valuation of the object to be $B = b - \tau$ when the objective worth for him is actually b. At time t, the (possibly correlated) distributions of seller and buyer biases are $\mathcal{E}_t, \mathcal{T}_t$, with support in the close intervals $\mathsf{E} = [\underline{\varepsilon}, \bar{\varepsilon}]$, $\mathsf{T} = [\underline{\tau}, \overline{\tau}]$, respectively. The true valuations b and s are drawn at random from the respective distributions $\mathcal{B}$ and $\mathcal{S}$.

Before bargaining starts, the seller only observes the distribution $\mathcal{B}_\tau$ of perceived valuations of the buyer, which is the distribution $\mathcal{B}$ of true valuations b shifted to the left by τ. Similarly, the buyer only observes the distribution $\mathcal{S}_\varepsilon$ of perceived valuations of the seller – the true distribution $\mathcal{S}$ of valuations s shifted to the right by ε. Thus, the actual perceived valuations are not mutually known. However, we do assume that when a player is confronted with a tough opponent, she can assess his level of toughness even before the start of negotiations.[4] This is mainly because toughness as a line of character usually manifests itself by "body language"[5]. When the opponent is indeed biased in nature, the player can do nothing but take this bias as given.

We assume that even after an exchange takes place, the traders do not become aware of the discrepancy between their conceived and actual valuations: Typically, an object acquired in a bargaining process is not available in parallel in a price-taking environment, so the individual does not get the opportunity to compare her behavior in the two settings and realize her bias, and neither is there an obvious way for her to compute the objective present value of the object once she has it. Thus, the individual does not account or correct for her bias in future incidences of bargaining. The same operational conclusion

---

[4] We relax this assumption in section 6.

[5] or "national ethos" in international disputes, or "corporate culture" in a business setting, etc.

is sustained even if the traders realize the discrepancy ex post trade, but their tendency to toughness is emotional. In such a case, even though they regret the forgone gains from trade in case of a negotiation breakdown or an excessively long delay[6], they are unable to avoid their inclination to toughness when they have to bargain again. This interpretation of the emotions as a commitment device is in the spirit of Frank (1988).

Depending on the bargaining protocol, the players then engage in a perfect Bayesian equilibrium. Denote by $\Pi_s(S, B; \varepsilon, \tau)$ and $\Pi_b(S, B; \varepsilon, \tau)$ the *true, objective* payoffs for the seller and for the buyer in the bargaining equilibrium when they have perceived valuations S, B and biases ε, τ, respectively[7]. Then the average true ex ante payoffs to a seller and buyer when endowed with the biases ε, τ are

$$\begin{aligned} f_s(\varepsilon, \tau) &\equiv \int\int \Pi_s(S, B; \varepsilon, \tau) \mathrm{d}\mathcal{S}_\varepsilon(S) \mathrm{d}\mathcal{B}_\tau(B) \\ f_b(\tau, \varepsilon) &\equiv \int\int \Pi_b(S, B; \varepsilon, \tau) \mathrm{d}\mathcal{S}_\varepsilon(S) \mathrm{d}\mathcal{B}_\tau(B) \end{aligned} \tag{2.1}$$

Therefore, the average true ex ante payoffs at time t for a seller with bias ε and a buyer with bias τ are, respectively,

$$\begin{aligned} f_s(\varepsilon, \mathcal{T}_t) &\equiv \int f_s(\varepsilon, \tau) \mathrm{d}\mathcal{T}_t(\tau) \\ f_b(\tau, \mathcal{E}_t) &\equiv \int f_b(\varepsilon, \tau) \mathrm{d}\mathcal{E}_t(\varepsilon) \end{aligned} \tag{2.2}$$

We assume that the joint distribution of biases $(\mathcal{E}_t, \mathcal{T}_t)$ evolves according to some *regular payoff-monotonic dynamics,* in which biases with higher average true payoff proliferate at the expense of biases with lower average true payoff *(*the technical definition is given in appendix 1). To study the evolution of biases with such dynamics, we consider a metaphorical two-player game between the "genes" of the players, who contemplate what bias they should hard-wire in the minds of their carriers in order to maximize long term prosperity. In this game, the seller's gene selects a bias $\varepsilon \in \mathsf{E}$ for the seller, the buyer's gene selects a bias $\tau \in \mathsf{T}$ for the buyer, and the payoffs in the game are the average ex ante payoffs (2.1) to players with these biases.[8] As it turns out, for all of the bargaining procedures that we examine below, this game between the genes turns to be dominance-solvable: A unique pair of biases (ε^*, τ^*) survives the process of iterative elimination of strongly dominated biases. A theorem of Heifetz and Spiegel (2000) (detailed as well in appendix 1) then implies that the joint distribution of biases $(\mathcal{E}_t, \mathcal{T}_t)$ will converge in distribution towards a unit mass at (ε^*, τ^*) from whatever initial distribution $(\mathcal{E}_0, \mathcal{T}_0)$ with support $\mathsf{E} \times \mathsf{T}$.

---

[6] Beyond the inevitable inefficiency due to asymmetric information, as in Myerson and Satterthwaite (1983)

[7] and hence true valuations $s = S - \varepsilon$, $b = B + \tau$. The seller thus misperceives her payoff to be $\Pi_s(S, B; \varepsilon, \tau) - \varepsilon$, and the buyer misperceives his payoff to be $\Pi_b(S, B; \varepsilon, \tau) - \tau$.

[8] These payoffs are of course different from what the biased individuals expect them to be, but it is the objective succeess which determines prosperity.

3 The Emergence of Toughness: Two Examples

We now turn to see how toughness can emerge in bargaining with asymmetric information on either one or both sides.

3.1 Signaling by Delay

The first bargaining game we examine, with asymmetric information on one side, is originally due to Admati & Perry (1987) for the case of finitely many types and to Cramton (1992) for the case of a continuum of types. There is a seller with a known valuation S for the object she tries to sell, and a buyer whose privately-known valuation B is uniformly distributed in the interval $[\underline{B}, \overline{B}]$ containing S. If trade takes place at time t for the price p, the seller's payoff is $e^{-rt}(p - S)$, and the buyer's payoff is $e^{-rt}(B - p)$. The players alternate in making offers to each other, and the buyer is the first one to offer. The minimal time between offers is $t^0 = -\frac{1}{r}\log\delta \;\; (0 < \delta < 1)$, so δ is the discount factor for this minimal delay. Alternatively the buyer can announce immediately that there are no gains from trade (if $B \leq S$) and the game ends.

On the separating equilibrium path the buyer of type $B > S$ chooses the delay

$$\beta(B) = -\frac{\delta}{r}\log\left(\frac{B-S}{\overline{B}-S}\right), \tag{3.1}$$

and then offers to trade at the Rubinstein (1982) subgame-perfect price

$$p(B,S) = \frac{\delta B + S}{1+\delta} \tag{3.2}$$

which the seller accepts.[9]

---

[9] The buyer's optimization problem is finding a delay that maximizes his payoff

$$\max_{\Delta} e^{-r\Delta}(B - \frac{\delta \mathsf{b}(\Delta) + S}{1+\delta})$$

where $\mathsf{b}(\Delta)$ is the inverse of $\beta(B)$. The first order condition (multiplied by $(1+\delta)$) is thus

$$-re^{-r\beta(B)}\,(B-S) - e^{-r\beta(B)}\delta\frac{d\mathsf{b}(\beta(B))}{d\Delta} = 0$$

which yields the differential equation

$$\frac{d\mathsf{b}(\beta(B))}{d\Delta} = -\frac{r}{\delta}\,(B-S)\,,$$

or equivalently

$$\frac{d\beta(B)}{dB} = -\frac{\delta}{r\,(B-S)}$$

with the initial condition that the buyer type with largest gains from trade makes an offer immediately –

$$\beta(\overline{B}) = 0.$$

Integration then gives (3.1).

Let us normalize the true valuation of the seller to be $s = 0$, and assume the distribution $\mathcal{B}$ of the true valuations b of the buyer to be uniform in the interval $[0, 1]$.[10] When the seller has a toughness bias ε, she believes and acts as if her valuation is $S = \varepsilon$. Similarly, when the buyer has a toughness bias τ, he believes and acts as if his valuation is B whenever his true valuation is in fact $b = B + \tau$. The perceived valuations of the buyer are therefore uniformly distributed in the interval $[\underline{B}, \overline{B}] = [-\tau, 1 - \tau]$.

At equilibrium, when the seller meets a buyer with perceived valuation B, her *true* payoff is

$$\Pi_s(S, B; \varepsilon, \tau) = e^{-r\beta(B)}\left(p(B, \varepsilon) - 0\right) = \left(\frac{B - \varepsilon}{1 - \tau - \varepsilon}\right)^{\delta}\left(\frac{\delta B + \varepsilon}{1 + \delta}\right)$$

and the *true* payoff of the buyer is

$$\Pi_b(S, B; \varepsilon, \tau) = e^{-r\beta(B)}\left(B + \tau - p(B, \varepsilon)\right) = \left(\frac{B - \varepsilon}{1 - \tau - \varepsilon}\right)^{\delta}\left(\frac{B - \varepsilon}{1 + \delta} + \tau\right)$$

The true ex ante expected payoffs for traders with biases ε, τ are therefore

$$\begin{aligned}
f_s(\varepsilon, \tau) &= \int_{\varepsilon}^{1-\tau} \Pi_s(S, B; \varepsilon, \tau)\mathrm{d}B = \\
&= \int_{\varepsilon}^{1-\tau} \left(\frac{B - \varepsilon}{1 - \tau - \varepsilon}\right)^{\delta}\left(\frac{\delta B + \varepsilon}{1 + \delta}\right)\mathrm{d}B = \frac{\delta(1 - \tau - \varepsilon)^2}{(1 + \delta)(2 + \delta)} + \frac{\varepsilon(1 - \tau - \varepsilon)}{(1 + \delta)} \\
f_b(\tau, \varepsilon) &= \int_{\varepsilon}^{1-\tau} \Pi_b(S, B; \varepsilon, \tau)\mathrm{d}B = \\
&= \int_{\varepsilon}^{1-\tau} \left(\frac{B - \varepsilon}{1 - \tau - \varepsilon}\right)^{\delta}\left(\frac{B - \varepsilon}{1 + \delta} + \tau\right)\mathrm{d}B = \frac{(1 - \tau - \varepsilon)^2}{(1 + \delta)(2 + \delta)} + \frac{\tau(1 - \tau - \varepsilon)}{(1 + \delta)}
\end{aligned}$$

If prosperity is increasing in wealth, then these expected payoffs may be viewed as the fitness of the traders as a function of their biases. In this sense, $f_s(\varepsilon, \tau)$ and $f_b(\varepsilon, \tau)$ define a metaphorical game between the "gene" of the seller and the "gene" of the buyer, each of whom chooses which bias to "plant" in the mind of the corresponding trader in order to maximize long-term prosperity. The reaction functions in this game

$$\begin{aligned}
\varepsilon(\tau) &\equiv \arg\max_{\varepsilon} f_s(\varepsilon, \tau) = \frac{2 - \delta}{4}(1 - \tau) \\
\tau(\varepsilon) &\equiv \arg\max_{\tau} f_b(\tau, \varepsilon) = \frac{\delta}{2(1 + \delta)}(1 - \varepsilon)
\end{aligned}$$

have a slope which is smaller than 1 in absolute value. The game between the genes is thus dominance solvable[11], i.e. the unique Nash equilibrium biases

$$(\varepsilon^*, \tau^*) = \left(\frac{2 - \delta}{\delta + 4}, \frac{\delta}{\delta + 4}\right)$$

---

[10] The results are not altered if we assume a uniform distribution on a larger interval $[\underline{b}, 1]$ for some $\underline{b} < 0$.

[11] All the sufficient conditions for dominance solvability in Moulin (1984, thm. 4) are fulfilled.

are the only ones that survive the process of iterative elimination of strongly dominated strategies. As explained in section 2, these biases will therefore asymptotically take over the population under any payoff monotonic dynamics, whatever is the initial distribution of biases on the entire set of biases $\mathsf{E} \times \mathsf{T}$. To sum up, we have proved the following Proposition:

Proposition 1 *(The emergence of toughness in bargaining with signaling by delay). Given any full-support initial distribution* $(\mathcal{E}_0, \mathcal{T}_0)$ *of seller and buyer biases, as time goes by the distribution of biases* $(\mathcal{E}_t, \mathcal{T}_t)$ *will converge in distribution towards a unit mass on the pair of positive biases* $(\frac{2-\delta}{\delta+4}, \frac{\delta}{\delta+4})$ *under any regular, payoff-monotonic dynamics. In particular, as the minimal delay* t^0 *between offers vanishes (so that* $\delta \to 1$*), a moderate toughness bias of* $\frac{1}{5}$ *for both the seller and the buyer will emerge.*

3.2 A Sealed-Bid Double Auction

In this trade procedure with asymmetric information on both sides, the seller and the buyer have to submit simultaneously an ask price and a bid price, respectively. If the ask price is lower or equal to the bid price, trade takes place at the average of these prices. The seller's valuation and the buyer's valuation are drawn from uniform distributions on the intervals $[\underline{S}, \overline{S}]$, and $[\underline{B}, \overline{B}]$, respectively, with $\underline{B} \leq \underline{S} < \overline{B} \leq \overline{S}$. If there are sellers with valuations $S > \overline{B}$ or buyers with valuations $B < \underline{S}$, they do not show up for trade. We assume that the traders engage in the unique Bayesian equilibrium of this game in which each trader's bidding strategy is linear in his/her valuation (Chatterjee and Samuelson 1983). The use of linear strategies received some experimental support (Radner and Schotter 1989), and this equilibrium realizes the highest expected gains from trade possible in any Bayesian equilibrium with this information structure (Myerson and Satterthwaite 1983).

We assume that the distributions $\mathcal{S}$ and $\mathcal{B}$ of the true valuations of the seller and the buyer are uniform on the $[0, 1]$ interval.[12]

Proposition 2 *(Emergence of biases in a sealed-bid double auction). Given any full-support initial distribution* $(\mathcal{E}_0, \mathcal{T}_0)$ *of seller and buyer biases with highest values* $\bar{\varepsilon}, \bar{\tau} < \frac{5}{14}$*, as time goes by the distribution of biases* $(\mathcal{E}_t, \mathcal{T}_t)$ *will converge in distribution towards a unit mass on a pair of positive biases* $(\frac{1}{10}, \frac{1}{10})$ *under any regular, payoff-monotonic dynamics.*

The proof is in appendix 2.

---

[12] The results remain the same if the seller's true valuations are uniformly distributed on $[0, \overline{s}]$ for some $\overline{s} \geq 1$, and the buyer true valuations are uniformly distributed on $[\underline{b}, 1]$ for some $\underline{b} \leq 0$.

4 General Analysis

What are the general principles that lead to the emergence of toughness in the above (and subsequent) bargaining procedures? We first turn to answer this question when the bargaining game satisfies some symmetry conditions, in which case we can characterize the biases that emerge. In less symmetric settings, it may not be possible to determine whether one specific bias takes over the population in the long run, but we can still spell out the conditions under which the population will asymptotically consist exclusively of tough individuals.

4.1 Symmetric Bargaining Games

Assume that the true seller and buyer valuations are distributed in the $[0, 1]$ interval.[13] Thus, when the seller and buyer have biases ε, τ, respectively, they believe there is a potential for trade between them only when their perceived valuations are in the interval $[\varepsilon, 1 - \tau]$.

We first consider bargaining games with the following symmetry properties:

1. On average, the gains from trade are shared equally between the players.
2. Whatever are the biases ε, τ of the seller and buyer, the distribution of their perceived valuations and their equilibrium strategies are just a re-scaling from their distributions and biases in the interval $[0, 1]$ when they have no biases to the interval $[\varepsilon, 1 - \tau]$. In particular, the seller's distribution satisfies

   $$\frac{\mathcal{S}(cx)}{\mathcal{S}(cy)} = \frac{\mathcal{S}(x)}{\mathcal{S}(y)}$$

   for every $c \in [0, 1]$ and every $x, y \in [0, 1]$.
3. The distribution $\mathcal{S}$ of the seller's true valuations is a mirror image of the distribution $\mathcal{B}$ of the buyer's true valuations: $\mathcal{B}(x) = 1 - \mathcal{S}(1 - x)$.

The second property implies that for every $c \in [0, 1]$ and every $x, y \in [0, 1]$ we have

$$\frac{\mathcal{S}(cx)}{\mathcal{S}(x)} = \frac{\mathcal{S}(cy)}{\mathcal{S}(y)} \equiv g(c) \tag{4.1}$$

---

[13] More precisely, to account for the possibility of negative (softness) biases, we should assume that for some $a > 0$, the true seller and buyer valuations are distributed in the intervals and $[0, a]$ and $[-a, 1]$, respectively. The results below obtain also with this more general setting, at the cost of more complex notation. The same applies for the case in which the valuations are distributed in some other intervals $[\underline{s}, \overline{s}]$ and $[\underline{b}, \overline{b}]$ with $\underline{b} < \underline{s} < \overline{b} < \overline{s}$.

This means that

$$\mathcal{S}(x) = x^{\alpha} \tag{4.2}$$

for some $\alpha > 0$.[14]

Denote by $p(s,b)$ the probability of trade at equilibrium in the game with no biases, when the seller and buyer valuations are s and b. When the game involves delay and discounting, let $p(s,b)$ be the probability of trade between s and b multiplied by the discounting caused by the delay. Then

$$P = \int_0^1 \int_0^1 p(s,b) d\mathcal{S}(s)\, d\mathcal{B}(b)$$

is the average probability of trade, and

$$G = \int_0^1 \int_0^1 (b-s)\, p(s,b) d\mathcal{S}(s)\, d\mathcal{B}(b)$$

is the average gains from trade. By property 1,

$$U = \frac{G}{2}$$

is the average profit of each of the traders.

With biases ε and τ of the seller and the buyer, trade takes place with a positive probability only when the true valuation s of the seller is in fact smaller than $1-\tau-\varepsilon$,[15] and the true valuation b of the buyer is larger than $\tau+\varepsilon$.[16] By properties 2 and 3 the probability of trade becomes

$$\begin{aligned} & \int_0^1 \left(\int_0^1 p(s,b)\mathcal{S}(1-\tau-\varepsilon) d\mathcal{S}(s) \right) (1-\mathcal{B}(\tau+\varepsilon)) d\mathcal{B}(b) \\ = & \int_0^1 \left(\int_0^1 p(s,b)(1-\tau-\varepsilon)^{\alpha} d\mathcal{S}(s) \right) (1-(1-(1-(\tau+\varepsilon))^{\alpha}))\, d\mathcal{B}(b) = P(1-\tau-\varepsilon)^{2\alpha} \end{aligned}$$

and the perceived gains from trade become

$$\int_0^1 \left(\int_0^1 ((1-\tau-\varepsilon)\,(b-s))\, p(s,b)\mathcal{S}(1-\tau-\varepsilon) d\mathcal{S}(s) \right) (1-\mathcal{B}(\tau+\varepsilon)) d\mathcal{B}(b) = G(1-\tau-\varepsilon)^{2\alpha+1}$$

However, the true profit of the seller of each completed transaction is actually higher by ε than she perceives it to be, summing up on average to

$$f_s(\varepsilon,\tau) = U(1-\tau-\varepsilon)^{2\alpha+1} + P(1-\tau-\varepsilon)^{2\alpha}\varepsilon \tag{4.3}$$

---

[14] Equation (4.1) implies that $g(c_1 c_2) = g(c_1)g(c_2)$, whence $g(c) = c^{\alpha}$ for some α. Substituting $\mathcal{S}(1) = 1$ in (4.1) yields (4.2).

[15] That is, when her perceived valuation S is smaller than $1-\tau$, which is the maximal perceived valuation of the buyer.

[16] That is, when his perceived valuation B is higher than ε – the minimal perceived valuation of the seller.

Similarly, the true ex ante average profit of the buyer is

$$f_b(\tau,\varepsilon) = U(1-\tau-\varepsilon)^{2\alpha+1} + P(1-\tau-\varepsilon)^{2\alpha}\tau. \tag{4.4}$$

These payoff functions of the "genes game" are twice continuously differentiable and strictly concave in the gene's bias in the area defined by: $\varepsilon < \frac{2P-(2\alpha+1)U}{(2\alpha+1)(P-U)}(1-\tau)$ and $\tau < \frac{2P-(2\alpha+1)U}{(2\alpha+1)(P-U)}(1-\varepsilon)$, so in particular for

$$\varepsilon, \tau < \frac{2P-(2\alpha+1)\,U}{2P+(P-2U)\,(2\alpha+1)} \tag{4.5}$$

The reaction functions are

$$\varepsilon(\tau) \equiv \arg\max_{\varepsilon} f_s(\varepsilon,\tau) = \frac{P-(2\alpha+1)U}{(2\alpha+1)\,(P-U)}(1-\tau) \tag{4.6}$$

$$\tau(\varepsilon) \equiv \arg\max_{\tau} f_b(\tau,\varepsilon) = \frac{P-(2\alpha+1)U}{(2\alpha+1)\,(P-U)}(1-\varepsilon) \tag{4.7}$$

with the Nash equilibrium

$$\varepsilon^* = \tau^* = \frac{P-(2\alpha+1)\,U}{P+(2\alpha+1)\,(P-2U)} \tag{4.8}$$

Thus, if

$$\frac{P}{U} > 2\alpha+1 \tag{4.9}$$

then the slope of the reaction functions are smaller than 1 in absolute value, the genes game is dominance solvable, and hence the biases ε^*, τ^* take over the population under regular payoff-monotonic dynamics. Furthermore, under (4.9) the emerging biases $\varepsilon^* = \tau^* > 0$ are toughness biases.

The left-hand side of (4.9) is a measure for the probability of trade relative to the gains from trade. It gets its minimal value if trade takes place with a positive probability only when the gains from trade are the largest possible, i.e. $b-s=1$. In this extreme mechanism, the average gains from trade G are equal to the average probability of trade P, and so $\frac{P}{U} = \frac{P}{\frac{G}{2}} = 2$. The larger is the probability of trade $p(s,b)$ also when the gains from trade $b-s$ are small, the larger will become the ratio $\frac{P}{U}$. In the second example in the previous section (Chatterjee-Samuelson 1983) we have $\frac{P}{U}=4$.

The right-hand side of (4.9) increases as the density of the seller's valuations becomes skewed to the right and that of the buyer skewed to the left. With such an increase, the ex ante attractiveness of toughness decreases, since it implies the loss of more large-profit opportunities. With uniform distributions we have $\alpha = 1$ and hence $2\alpha+1=3$.

To sum up, we should expect to see the condition (4.9) fulfilled and toughness emerging in bargaining mechanisms which allow a fair chance for trade even when the gains from trade are moderate, and in which the traders' distributions are not extremely skewed towards large-profit potential trades.

A similar analysis can be carried out for the case in which only the buyer's perceived valuation is not known by the seller, while the seller's true valuation is 0, and her perceived valuation is known by the buyer. With the symmetry properties 1 and 2 above, similar results obtain, in which every occurrence of the term 2α in expressions (4.3)-(4.9) is replaced by α. This is applicable at the limit when $\delta \to 1$ in the first example of the previous section (signaling by delay). In that case we have $U = \frac{1}{6}$ and $P = \frac{1}{2}$, and so $\frac{P}{U} = 3 > \alpha + 1 = 2$.

4.2 Asymmetric Bargaining Games

When the traders are impatient ($\delta < 1$) in the first bargaining game of the previous section (signaling by delay), the gains from trade are not shared equally between the buyer and the seller.[17] In that game, as well as in other asymmetric bargaining protocols detailed in the next section, we are able to show the toughness biases that will emerge under payoff-monotonic evolutionary dynamics. However, it is not always the case that the "genes game" is dominance solvable in asymmetric bargaining mechanisms, and in those cases it is not generally possible to determine whether the evolutionary system converges.[18] It is therefore instructive to find out under what bargaining procedures it is the case that softness and slight toughness will be asymptotically weeded out by some positive degree of toughness, even if the evolution of biases does not converge. Under such conditions, we should expect to see only tough individuals in the long run, even though the proportions of toughness degrees in the population may never stabilize but remain in flux.

To this end, we shall check what conditions guarantee that a small toughness bias dominates the zero bias when most of the population is almost unbiased. Since these conditions will be derived from a marginal analysis, they imply that if the intervals $\mathsf{E} = [\underline{\varepsilon}, \bar{\varepsilon}]$, $\mathsf{T} = [\underline{\tau}, \overline{\tau}]$ of the seller and buyer biases are small enough and contain the zero bias, the distribution of biases will gradually assign more and more weight to positive, tough biases under payoff monotonic dynamics.

We will exemplify this marginal analysis for a seller whose (perceived) valuation $S = \varepsilon$ is known by the buyer, while the buyer is unbiased and his valuations b are distributed on the support $[0, 1]$, with the probability distribution $\mathcal{B}$. Denote by $p(b; \varepsilon)$ the probability of trade when the buyer has valuation b and the seller has a bias ε. When the bargaining protocol allows for delays, $p(b; \varepsilon)$ is the probability of trade multiplied by the discounting suffered by the seller until trade occurs. Denote also by $u(b; \varepsilon)$ the perceived profit of a seller with bias ε from the trade with a buyer with valuation b, if and when trade takes place. Assume that for $\mathcal{B}$ a.e. b, both p and u are differentiable with respect to ε. Then

---

[17] The proposer in that bargaining procedure – the buyer – gets $\frac{1}{1+\delta}$ of the gains from trade.

[18] See Huck et al. (1997) for such an example with complete information.

the actual expected profit of the seller is

$$f_s(\varepsilon, 0) = \int_{\varepsilon}^{1} (u(b;\varepsilon) + \varepsilon)\, p(b;\varepsilon) d\mathcal{B}(b)$$

We will look for the conditions under which $\frac{d}{d\varepsilon} f_s(0,0) > 0$. This will guarantee that an interval of positive biases of the seller dominate the zero bias, provided that the buyer is unbiased. Also, if the average fitness $f_s(\varepsilon, \mathcal{T})$ is continuously differentiable in both its arguments[19] (as is the case in all the examples we analyze), the same conclusion holds if $\mathcal{T}$ is sufficiently concentrated around the zero bias for the buyer.

Explicitly,

$$\frac{d}{d\varepsilon} f_s(0,0) = -u(0;0)p(0;0)\frac{d\mathcal{B}(0)}{d\varepsilon} + \int_0^1 \left(\left(\frac{d}{d\varepsilon} u(b;0) + 1 \right) p(b;0) + \left(\frac{d}{d\varepsilon} p(b;0) \right) u(b;0) \right) d\mathcal{B}(b)$$

The first term stands for the marginal loss of trade with buyers of small valuations. Since $u(0;0) = 0$ (no profit for a seller with valuation 0 from trade with a buyer of valuation 0), this marginal loss is null. Generally, we should expect both $\frac{d}{d\varepsilon} u(b;0)$ and $\frac{d}{d\varepsilon} p(b;0)$ to be non-positive: The former because the overall potential gains from trade (and hence the part of it that goes to the seller) between buyer b and seller ε decrease with ε; and the latter because longer delays and/or a higher chance of negotiation breakdown is to be expected when the difference $b - \varepsilon$ between the buyer and seller valuations decreases. So overall, we have that $\frac{d}{d\varepsilon} f_s(0,0) > 0$ if and only if

$$\int_0^1 p(b;0) d\mathcal{B}(b) > -\int_0^1 \frac{d}{d\varepsilon} \left(u(b;0)p(b;0) \right) d\mathcal{B}(b)$$

The left-hand side is the overall probability of trade with no bias, and the right-hand side is the marginal decrease in the seller's perceived profits when she turns to be slightly tough. This inequality obtains in all of the examples we analyze in this paper.

A similar marginal analysis can be carried out now for the buyer, as well as for the case of asymmetric information on both sides.

5 Other bargaining Protocols

In this section we bring further examples of bargaining procedures and the biases that emerge in them under payoff-monotonic dynamics. Toughness arises under all of these procedures.

---

[19] With the weak-$*$ topology on distributions $\mathcal{T}$ of buyer biases τ.

5.1 Screening and Signaling

What happens if the seller is the one to make the first offer in the first bargaining game of section 3? Typically (Cramton 1992), the seller will immediately make an offer which will appeal to a bunch of high-valuation buyer types $\left[\tilde{B}, \overline{B}\right]$, who will prefer to trade in the proposed terms rather than bargain further at the cost of delay. The remaining, low-valuation buyer types $\left(S, \tilde{B}\right)$ will prefer to reject the seller's initial offer, and delay their counter-offer (beyond the minimal delay t^0 which incurs discounting by δ) so as to signal credibly their valuation by the length of the delay, as in (3.2) above[20].[21] The higher the price offered initially by the seller, the higher is the seller's profit given that her offer is accepted, but the smaller will be the interval of types $\left[\tilde{B}, \overline{B}\right]$ who will indeed accept it, and the longer will be the delay exercised by each of the rejecting types in $\left(S, \tilde{B}\right)$ until he makes the counter-offer to trade at the Rubinstein (1982) price $p(B, S)$ in (3.2). Understanding this, the seller will tune her initial offer $P_{\tilde{B}}$ so as to balance optimally between these considerations, in order to maximize her expected payoff

$$\int_{\tilde{B}}^{\overline{B}} \left(P_{\tilde{B}} - S\right) \mathrm{d}B + \int_{S}^{\tilde{B}} \delta e^{-r\beta(B)} \left(p(B,S) - S\right) \mathrm{d}B \tag{5.1}$$

where $P_{\tilde{B}}$ is the price for which the buyer with valuation $\tilde{B}$ is just indifferent between trading at $P_{\tilde{B}}$ or trading at $p(\tilde{B}, S)$ after the minimal delay t^0 :

$$\tilde{B} - P_{\tilde{B}} = \delta \left(\tilde{B} - p(\tilde{B}, S)\right) \tag{5.2}$$

Proposition 3 *(The emergence of toughness in bargaining with screening followed by signaling). Given any full-support initial distribution $(\mathcal{E}_0, \mathcal{T}_0)$ of seller and buyer biases, as time goes by the distribution of biases $(\mathcal{E}_t, \mathcal{T}_t)$ will converge in distribution towards a unit mass on a pair of positive biases (ε^*, τ^*) under any regular, payoff-monotonic dynamics. When $\delta \to 1$ the emerging biases tend to $(\frac{4}{17}, \frac{3}{17})$.*

The proof is in appendix 2.

5.2 Alternating offers with two-sided uncertainty

Here we consider the full Cramton (1992) equilibrium with alternating offers and two-sided uncertainty. The informational structure is as in the sealed-bid double-auction

---

[20] With $\tilde{B}$ replacing $\overline{B}$ in (3.2).

[21] As before, the buyer types with valuations B smaller than the known valaution S of the seller will leave the arena even before the initial offer, and no trade will occur.

model, with the caveat in footnote 12. The identity of the first offeror and the timing of the offer is endogenous: Eager-to-trade players – buyers with valuations closer to $\overline{B}$ and sellers with valuations closer to $\underline{S}$ – will make an offer before other, more patient players. The exact timing is given in Cramton (1992, proposition 4). The offer of the first offeror reveals her valuation. If it is the seller, the game now proceeds as in section 5.1 above (screening and signaling). If it is the buyer, the roles of the seller and buyer in section 5.1 are interchanged in a symmetric way.

Proposition 4 *(Emergence of biases with alternating offers and two-sided uncertainty). Given any full-support initial distribution* $(\mathcal{E}_0, \mathcal{T}_0)$ *of seller and buyer biases with highest values* $\bar{\varepsilon}, \bar{\tau} < \frac{5}{14}$*, as time goes by and as the minimal delay* t^0 *between offers vanishes (so that* $\delta \rightarrow 1$*), the distribution of biases* $(\mathcal{E}_t, \mathcal{T}_t)$ *will converge in distribution towards a unit mass on a pair of negative biases* $(\frac{1}{10}, \frac{1}{10})$ *under any regular, payoff-monotonic dynamics.*

The proof is in appendix 2.

5.3 Take-it-or-leave-it Offer

In this much simpler model, the seller with a known valuation S has a unique chance to offer an ask price p to the buyer, which the buyer can either accept or reject. The buyer's valuations are uniformly distributed in $\left[\underline{B}, \overline{B}\right]$. Buyer types with valuations in $\left[p, \overline{B}\right]$ will therefore accept the offer, while those in $[\underline{B}, p)$ will reject it. The seller will balance optimally between the probability of trade and her profit given trade. The optimal offer is

$$p^* \equiv \arg\max_p \left(\overline{B} - p\right)(p - S) = \frac{S + \overline{B}}{2}$$

As in the previous models, we maintain the normalization by which the seller's true valuation is $s = 0$, and the buyer's true valuations are uniformly distributed on the interval $[0, 1]$.[22]

In contrast with the previous models (and the that follows), in this simple model the behavior of the buyer (acceptance or rejection) does not depend on the valuation S he perceives the seller to have, but only on the actual offer p. Hence the seller cannot influence the buyer's behavior by misperceiving her own valuation, and can only lose by miscalculating her true stakes. Therefore, if this is the bargaining procedure that is played when people meet to bargain, no bias of misperception will evolve for the seller, and all non-zero biases will be weeded out under a payoff-monotonic dynamics.

However, a toughness bias τ of the buyer *will* influence the price p^* offered by the seller, since $\overline{B} = 1 - \tau$ and therefore

$$p^* = \frac{S + (1 - \tau)}{2}.$$

---

[22] See footnote 10 above

>From the point of view of the "gene" of the buyer, the optimal bias to plant in the buyer's mind is the one which maximizes the true ex ante expected payoff of the buyer

$$\tau^* = \arg\max_{\tau} \int_{\frac{S+(1-\tau)}{2}}^{1-\tau} ((B+\tau) - p^*)\,\mathrm{d}B = \frac{1-S}{3}$$

Since the dominant bias of the seller is zero, $S = s + 0 = 0$, the only bias for the buyer which is not iteratively eliminated is $\tau^* = \frac{1}{3}$. We have thus shown:

Proposition 5 *(Emergence of biases with a take-it-or-leave-it offer). Given any full-support initial distribution* $(\mathcal{E}_0, \mathcal{T}_0)$ *of seller and buyer biases, as time goes by the distribution of biases* $(\mathcal{E}_t, \mathcal{T}_t)$ *will converge in distribution towards a unit mass on the pair of biases* $(0, \frac{1}{3})$ *under any regular, payoff-monotonic dynamics.*

5.4 Repeated Screening

In the model of Sobel and Takahashi (1983), a seller with a known valuation S makes repeated trading-price offers to the buyer, whose valuation is uniformly distributed[23] in the range $\left[\underline{B}, \overline{B}\right]$, with delay t^0 between the offers. At a subgame-perfect equilibrium[24], the i-th offer P_i is accepted by a bunch of high-valuation buyer types $(\overline{B}_i, \overline{B}_{i-1}]$ [25] who have rejected all previous offers, while the remaining, lower-valuation buyer types $(S, \overline{B}_i]$ prefer to wait for better, future offers. After each round in which trade did not occur, the seller faces an analogous optimization problem, only with a smaller interval of buyer types. The structure of the equilibrium is thus recursive as well, with

$$\overline{B}_i - S = d\left(\overline{B}_{i-1} - S\right)$$

and

$$P_i = S + c\left(\overline{B}_i - S\right) = S + cd^{i-1}\left(\overline{B} - S\right)$$

for some $c, d \in (0,1)$ which are determined by the players discounting δ for each round of forgone trade.

As before, we continue to assume that the seller's true valuation is $s = 0$, and the buyer's true valuations are uniformly distributed in the interval $[0,1]$.[26]

---

[23] Sobel and Takahashi (1983) consider also other distributions.

[24] In a subgame-perfect equilibrium the seller cannot commit to a sequence of offers in advance, but rather has to make an optimal offer after any given number of rejections.

[25] with the convention $\overline{B}_0 = \overline{B}$

[26] See footnote 10 above.

Proposition 6 *(Emergence of biases in bargaining with repeated one-sided offers). Given any full-support initial distribution* $(\mathcal{E}_0, \mathcal{T}_0)$ *of seller and buyer biases, as time goes by the distribution of biases* $(\mathcal{E}_t, \mathcal{T}_t)$ *will converge in distribution towards a unit mass on a pair of positive biases* (ε^*, τ^*) *under any regular, payoff-monotonic dynamics. When* $\delta \to 1$ *the emerging biases tend to* $(\frac{1}{2}, 0)$*.*

The proof is in appendix 2.

When the delay t^0 between offers tends to zero (so that $\delta \to 1$), more and more of the gains from trade are extracted by the buyer: When the delay t^0 is very small, the buyer knows that better offers are forthcoming very soon, many buyer types are willing to wait for them, and hence the seller's offers must be very attractive to the buyer. Hence the incentives for the buyer's "gene" to bias and misrepresent the buyer's valuations in the seller's eyes are small, since the buyer is getting almost the entire pie anyhow. Thus, the only way that the seller can sustain an actual profit is by having an "endowment effect" $\varepsilon > 0$, so that her offers will never get below ε. With a small delay between offers, the seller trades with all the buyer types with valuation $B > \varepsilon$, at an average price slightly above ε. In the limit as $\delta \to 1$, the bias $\varepsilon^* = \frac{1}{2}$ has the best balance between the actual profit ε and the probability of realizing the trade.

However, we emphasize again that for every positive delay t^0 between the offers, the emerging buyer bias τ^* is positive as well.

6 Refinements of the model

In this section we relax our assumption that the biases of the bargainers are mutually observable. In a first refinement, we assume that the biases are observed only in some of the interactions, while in the rest of them a Bayesian game is played, based only on knowledge of the current distributions of biases in the population. In a second refinement, we assume that a player may be inclined to pretend that he is tougher or softer than he actually is, bearing a deception cost.

We exemplify these refinements in the simplest of the bargaining settings - that with a take-it-or-leave-it offer by the seller. In that setting, we know that no bias will evolve for the seller (see section 5.3 above), and thus we confine our analysis to the buyer side. This enables us to obtain analytic results regarding the asymptotic behavior of the evolutionary dynamics. The analysis shows how the qualitative features of our analysis are maintained with the refinements.

6.1 Partial Observability

Suppose that in the take-it-or-leave-it model (section 5.3.), the seller observes the buyer's bias only with probability w. In the remaining cases, she only knows the current distribution of the buyers' biases, $\mathcal{T}$ (with a density function $f(\tau)$ an a support $[\underline{\tau}, \overline{\tau}]$). Therefore she should maximize:

$$\max_p \int_{\underline{\tau}}^{1-p} p(1-\tau-p)d\mathcal{T}(\tau)$$

We derive it with respect to p to get the first order condition:

$$\frac{d}{dp}\left(\int_{\underline{\tau}}^{1-p} p(1-\tau-p)d\mathcal{T}(\tau)\right) = \int_{\underline{\tau}}^{1-p} (1-\tau-2p)d\mathcal{T}(\tau) = 0$$

or

$$(1-2p)\int_{\underline{\tau}}^{1-p} d\mathcal{T}(\tau) - \int_{\underline{\tau}}^{1-p} \tau d\mathcal{T}(\tau) = 0$$

yielding the solution

$$p^* = \frac{1-\tilde{\tau}|_{1-p^*}}{2}$$

where $\tilde{\tau}|_{1-p^*}$ is the average of those biases smaller than $1-p^*$.

The true payoff of a buyer with a perceived valuation B and bias τ is $B-p^*+\tau$. Therefore, this buyer's true ex-ante payoff (in case of no observability) is

$$U_B(\tau) = \int_{p^*}^{1-\tau} (B-p^*+\tau)dB = \frac{1}{2}\left(p^*-1+\tau\right)\left(p^*-1-\tau\right)$$

Clearly, the bias that maximizes this expression is $\tau = 0$ (the un-biased type "hides" behind the biased types: Given that p^* is independent of τ, which the seller cannot observe, the optimizing type $\tau = 0$ has the best performance).

With full observability of the buyer's bias (or, more precisely, the distribution of perceived valuations), the seller will offer the price

$$p^* = \frac{1-\tau}{2}$$

to a buyer with a bias τ, whose true expected payoff will therefore be

$$U_B(\tau) = \int_{\frac{1}{2}(1-\tau)}^{1-\tau} (B+\tau-\frac{1}{2}(1-\tau))dB = \frac{1}{8}\left(3\tau+1\right)\left(1-\tau\right)$$

Now, if the probability of observability is $w \in [0,1]$, the expected payoff for the buyer of type τ is

$$U_B(\tau;\pi) = w\frac{1}{8}\left(3\tau+1\right)\left(1-\tau\right) + (1-w)\frac{1}{2}\left(p^*-1+\tau\right)\left(p^*-1-\tau\right)$$

The type that maximizes this expected payoff is

$$\tau^* = \frac{w}{4-w}$$

which is indeed $\frac{1}{3}$ for $w=1$, as in section 5.3, and 0 for $w=0$. The higher the percentage of interactions in which the bias is observed, the larger will be the bias selected for in the evolutionary process.

6.2 Deception

Suppose now that a buyer of bias τ can be inclined to make an effort to appear as if he has the bias m, by paying the deception cost

$$a(\tau - m)^2$$

for some parameter $a > 0$. Thus, m is the "mask" that the buyer puts on his face. The type of the buyer then becomes two-dimensional – (τ, m). The seller only observes the mask m, but she is not naive: although she does not know τ, she knows the current distribution $\mathcal{T}(\tau; m)$ of biases τ among the buyers who use the mask m. Her offer p^*_m when she trades with a buyer of mask m is as we computed before:

$$p^*_m = \frac{1 - \tilde{\tau}|_{1-p^*_m}}{2}$$

where $\tilde{\tau}|_{1-p^*_m}$ is the average of the biases smaller than $1 - p^*_m$ according to the distribution $\mathcal{T}(\tau; m)$.

The true ex ante average payoff of a buyer of type (τ, m) is therefore

$$U_B(\tau) = \int_{p^*_m}^{1-\tau} (B - p^*_m + \tau)dB - a(\tau - m)^2 = \frac{1}{2}(p^*_m - 1 + \tau)(p^*_m - 1 - \tau) - a(\tau - m)^2$$

The bias τ^*_m that maximizes this expression is

$$\tau^*_m \equiv \arg\max_{\tau} U_B(\tau) = \frac{m}{\frac{1}{2a} + 1} < m$$

The type τ^*_m has the optimal balance between two competing considerations. On one hand, avoiding deception ($\tau = m$) avoids the deception cost, but involves a relatively large deviation from rationality and optimization. On the other hand, optimization ($\tau = 0$) involves a relatively high deception cost. So when $a = 0$ (no deception costs), it's best to be un-biased ($\tau^*_m = 0$) as long as there is nevertheless a positive fraction of biased types with the same mask m, behind which the un-biased type can "hide". When $a \to \infty$ and the deception costs become prohibitive, we have $\tau^*_m \to m$. When $0 < a << \infty$, we have $0 < \tau^*_m < m$.

Thus, under payoff-monotonic dynamics the distribution of biases τ with mask m will converge in distribution to the unit mass at τ^*_m. Consequently, p^*_m will converge with time to $\frac{1-\tau^*_m}{2}$. The true average ex ante payoff of a buyer with bias τ^*_m and with mask m will therefore converge to

$$\begin{aligned}
&\frac{1}{2}\left(\frac{1-\tau^*_m}{2} - 1 + \tau^*_m\right)\left(\frac{1-\tau^*_m}{2} - 1 - \tau^*_m\right) - a(\tau^*_m - m)^2 \\
&= \frac{1}{2}\left(\frac{1 - \frac{m}{\frac{1}{2a}+1}}{2} - 1 + \frac{m}{\frac{1}{2a}+1}\right)\left(\frac{1 - \frac{m}{\frac{1}{2a}+1}}{2} - 1 - \frac{m}{\frac{1}{2a}+1}\right) - a(\frac{m}{\frac{1}{2a}+1} - m)^2
\end{aligned}$$

The mask m which maximizes this expression is

$$m^* = \frac{1+2a}{2\,(3a+2)}$$

and the corresponding bias is

$$\tau^*_{m^*} = \frac{m^*}{\frac{1}{2a}+1} = \frac{\frac{1+2a}{2(3a+2)}}{\frac{1}{2a}+1} = \frac{a}{3a+2} = m^* - \frac{1}{2(3a+2)}$$

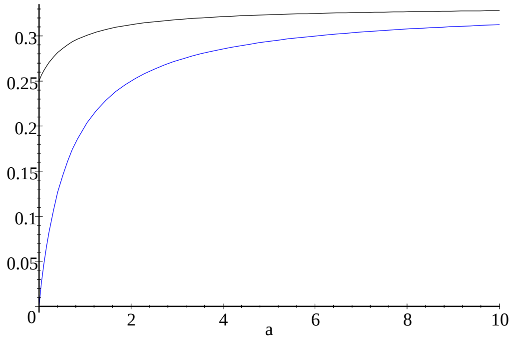

Figure 1: Emerging masks and biases as a function of deception costs

Under payoff-monotonic dynamics, the distribution of pairs (τ, m) will therefore converge in distribution to $(m^*, \tau^*_{m^*})$. In the limit, all the buyers will pretend to be tougher than they really are ($m^* > \tau^*_{m^*}$), even though asymptotically the true bias[27] is known by the seller. Bearing the deception costs $a\,(m^* - \tau^*_{m^*})^2$ is needed in order to make the sellers believe that the bias is indeed $\tau^*_{m^*}$: If we had $m^* = \tau^*_{m^*}$, a marginal decrease in τ from $\tau^*_{m^*}$ (while keeping the mask m^*) would entail zero marginal deception costs, but would marginally enhance the actual efficiency of the buyer's decision making – the buyer would reject less often the seller's offer when it's actually good for him. Thus $m^* = \tau^*_{m^*}$ cannot be a stable configuration of the evolutionary dynamics.

7 Conclusion

We have shown how a toughness bias may emerge under various bargaining settings. This is a particular example of a more general phenomenon: Each setting of strategic interaction would typically give rise to some non-trivial biases (Heifetz and Spiegel 2000). This suggests a new perspective on issues of mechanism design. If biases are endogenous to the mechanism, what is the bargaining mechanism that maximizes the bargainers actual gains from trade? What is the mechanism that maximizes the *perceived* gains

---

[27] Or, more precisely, the actual distribution of perceieved valuations.

from trade – the one that will make the traders happiest? Policy questions arise as well: If a third party tries to mitigate bargaining inefficiencies by a subsidy for trade, what will be the eventual implications on the biases that may evolve? Taking these biases into account, what is the policy needed to maximize actual or perceived efficiency, or the policy maker interests?

Similar questions arise, of course, not only in the context of bargaining, but in many other areas – from auction design to the design of legislation. We plan to address these issues in future work.

8 Appendix 1: Regular Payoff-Monotonic Dynamics

In this appendix we provide the formal definitions of the evolutionary dynamics employed in the paper.

At any point t in time, the population is characterized by a joint distribution $G_t = (\mathcal{E}_t, \mathcal{T}_t)$ of seller and buyer biases $\theta = (\varepsilon, \tau)$ from the compact support $\Theta = \mathsf{E} \times \mathsf{T} = [\underline{\varepsilon}, \bar{\varepsilon}] \times [\underline{\tau}, \overline{\tau}]$. We emphasize that what follows applies in fact to any compact space Θ of biases. A continuous growth-rate function $g : \Theta \times \Delta(\Theta) \to R$ determines how the distribution of biases $G_t \in \Delta(\Theta)$ in the population evolves over time by the following differential equation in the space of probability measures $\Delta(\Theta)$:

$$\dot{G}_t(A) = \int_A g(\theta, G_t) \mathrm{d}G_t(\theta), \quad A \subseteq \Theta \text{ measurable.} \tag{A1.1}$$

For the time path G_t to remain in $\Delta(\Theta)$ (and not to protrude to the general space of signed measures over Θ), the growth-rate function g should preserve the population size:

$$\int_\Theta g(\theta, G) \mathrm{d}G(\theta) = 0. \tag{A1.2}$$

We say that g is *regular* if it can be extended to the domain of signed measures G with variational norm smaller than 2, and on this extended domain it is uniformly bounded and uniformly Lipschitz continuous:

$$\sup_{\theta \in \Theta} |g(\theta, G)| < \infty,$$

$$\sup_{\theta \in \Theta} \left| g(\theta, G) - g(\theta, \tilde{G}) \right| < K \left\| G - \tilde{G} \right\|, \qquad G, \tilde{G} \in X,$$

for some constant K, where $\|G\| = \sup_{|h| \leq 1} \left| \int_T h \mathrm{d}G \right|$ is the variational norm on signed measures.

Oechssler and Riedel (2001, Lemma 3) proved that the regularity of g guarantees that the mapping $G \to \int_T g(\cdot, G) \mathrm{d}G$ is bounded and Lipschitz continuous in the variational norm, which implies that together with (A1.2), the differential equation (A1.1) in the space of distributions $\Delta(\Theta)$ has a unique solution for any initial distribution G_0.

Since individuals turn to be sellers or buyers with equal probabilities,

$$f(\theta,\theta') = \frac{f_s(\varepsilon,\tau') + f_b(\tau,\varepsilon')}{2}$$

is the average ex ante payoff to an individual of type θ when bargaining with one with type θ'. We interpret $f(\theta,\theta')$ as a payoff function in a game between the "genes" of the players, who "choose" the biases so as to maximize actual wealth. We denote by

$$f(\theta,G) = \frac{f_s(\varepsilon,\mathcal{T}) + f_b(\tau,\mathcal{E})}{2}$$

the average ex ante payoff to bias $\theta = (\varepsilon,\tau)$ when the distribution of biases in the population is $G = (\mathcal{E},\mathcal{T})$. Thus, we say that g is *payoff monotonic* if

$$g(\theta,G) > g(\theta',G) \quad \Longleftrightarrow \quad f(\theta,G) > f(\theta',G).$$

We say that θ' is dominated by θ if $f(\theta,\theta'') > f(\theta',\theta'')$ for every $\theta'' \in \Theta$. Denote by D_1 the set of types θ' that are dominated by some $\theta \in \Theta$, and $U_1 = \Theta \setminus D_1$. Inductively, if D_n is the set of biases dominated after at most n iterations and $U_n = \Theta \setminus D_n$, we say that $\theta' \in U_n$ is dominated in iteration $n+1$ by $\theta \in U_n$ if $f(\theta,\theta'') > f(\theta',\theta'')$ for every $\theta'' \in U_n$. We say that θ' is *serially dominated* if it is dominated after some number of iterations. The "genes game" is dominance solvable if there is a unique $\theta^* = (\varepsilon^*,\tau^*)$ which is not serially dominated.

The following theorem, which generalizes a result of Samuelson and Zhang (1992), asserts that serially dominated biases get extinct in the limit:[28]

Theorem (Heifetz and Spiegel 2000)**.** *When the support Θ of the initial distribution G_0 is compact, biases θ which are serially dominated are asymptotically weeded out – they have a neighborhood $V \ni \theta$ for which $\lim_{t\to\infty} G_t(V) = 0$. In particular, when the game is dominance solvable with equilibrium $\theta^* = (\varepsilon^*,\tau^*)$, G_t converges in distribution to a unit mass at θ^*.*

In particular, $\theta' = (\varepsilon',\tau)$ (or $\hat{\theta} = (\varepsilon,\hat{\tau})$) is dominated by $\theta = (\varepsilon,\tau)$ in iteration $n+1$ if for every $\theta'' = (\varepsilon'',\tau'') \in U_n$ we have $f_s(\varepsilon,\tau'') > f_s(\varepsilon',\tau'')$, (or $f_b(\tau,\varepsilon'') > f_b(\hat{\tau},\varepsilon'')$, respectively). Thus, dominance solvability in the asymmetric game between the "gene of the seller", with payoff $f_s(\varepsilon,\tau)$, and the "gene of the buyer", with payoff $f_b(\tau,\varepsilon)$, implies dominance solvability of the symmetric game $f(\theta,\theta')$.This sufficient condition for domination is the one applied throughout the paper.

---

[28] In Heifetz and Spiegel (2000), the theorem is phrased for the case in which the support Θ of the initial distribution is a closed interval, but the proof only relies on the compactness of Θ.

9 Appendix 2: Proofs

This appendix provides the computations of the emerging biases for the remaining bargaining protocols.

9.1 A Sealed-Bid Double Auction

Proof of Proposition 2. In Chatterjee and Samuelson's (1983) equilibrium, a seller with valuation S bids

$$\frac{2}{3}S+\frac{1}{4}\overline{B}+\frac{1}{12}\underline{S}$$

and a buyer with valuation B bids

$$\frac{2}{3}B+\frac{1}{12}\overline{B}+\frac{1}{4}\underline{S}$$

This means that trade takes place only when

$$B-S\geq\frac{1}{4}\left(\overline{B}-\underline{S}\right).$$

With biases ε and τ for the buyer and seller, respectively, we have $\underline{S}=\varepsilon$ and $\overline{B}=1-\tau$. Ex Ante, the perceived profits for each of the players are therefore half of the total (perceived) gains from trade, which are

$$G\left(\varepsilon,\tau\right)=\int_{\varepsilon}^{\varepsilon+\frac{3}{4}(1-\tau-\varepsilon)}\int_{S+\frac{1}{4}(1-\tau-\varepsilon)}^{1-\tau}(B-S)dBdS=\frac{9}{64}\left(1-\tau-\varepsilon\right)^{3}$$

and the probability of trade is

$$P\left(\varepsilon,\tau\right)=\int_{\varepsilon}^{\varepsilon+\frac{3}{4}(1-\tau-\varepsilon)}\int_{S+\frac{1}{4}(1-\tau-\varepsilon)}^{1-\tau}dBdS=\frac{9}{32}\left(1-\tau-\varepsilon\right)^{2}$$

The actual ex ante profits for the seller are hence

$$f_{s}\left(\varepsilon,\tau\right)=\frac{G\left(\varepsilon,\tau\right)}{2}+\varepsilon P\left(\varepsilon,\tau\right)=\frac{9}{128}\left(1-\tau-\varepsilon\right)^{3}+\frac{9}{32}\left(1-\tau-\varepsilon\right)^{2}\varepsilon.$$

Similarly, the actual ex ante profits for the buyer are

$$f_{b}\left(\tau,\varepsilon\right)=\frac{G\left(\varepsilon,\tau\right)}{2}+\tau P\left(\varepsilon,\tau\right)=\frac{9}{128}\left(1-\tau-\varepsilon\right)^{3}+\frac{9}{32}\left(1-\tau-\varepsilon\right)^{2}\tau.$$

These payoff functions of the “genes game” are twice continuously differentiable and strictly concave in the gene’s bias in the area defined by: $\varepsilon < \frac{5}{9} - \frac{5}{9}\tau$ and $\tau < \frac{5}{9} - \frac{5}{9}\varepsilon$, so in particular for $\varepsilon, \tau < \frac{5}{14}$. The reaction functions

$$\begin{aligned} \varepsilon(\tau) &\equiv \arg\max_{\varepsilon} f_s(\varepsilon, \tau) = \frac{1}{9} - \frac{1}{9}\tau \\ \tau(\varepsilon) &\equiv \arg\max_{\tau} f_b(\tau, \varepsilon) = \frac{1}{9} - \frac{1}{9}\varepsilon \end{aligned}$$

have slopes smaller than 1 in absolute value. Thus, under any regular payoff-monotonic dynamics, the bias distributions $(\mathcal{E}_t, \mathcal{T}_t)$ will therefore converge in distribution to the unit mass at the unique Nash equilibrium of the “genes game”, which are

$$\varepsilon^* = \tau^* = \frac{1}{10}$$

■

9.2 Screening and Signaling

Proof of Proposition 3. Expressions (3.2) and (5.2) yield

$$P_{\hat{B}} = \hat{B} - \delta\left(\hat{B} - p(\hat{B}, S)\right) = \hat{B} - \delta\left(\hat{B} - \frac{\delta\hat{B} + S}{1+\delta}\right) = \frac{\hat{B} + \delta S}{1+\delta}$$

Therefore, by (5.1) and (3.1) the seller chooses $\hat{B}$ so as to maximize

$$\begin{aligned} &\int_{\hat{B}}^{\overline{B}} \left[P_{\hat{B}} - S\right] \mathrm{d}B + \int_{S}^{\hat{B}} \delta e^{-r\beta(B)} \left(p(B,S) - S\right) \mathrm{d}B \\ &= \left(\overline{B} - \hat{B}\right)\left(\frac{\hat{B} + \delta S}{1+\delta} - S\right) + \int_{S}^{\hat{B}} \delta \left(\frac{B - S}{\hat{B} - S}\right)^{\delta} \left(\frac{\delta B + S}{1+\delta} - S\right) \mathrm{d}B \\ &= \frac{\delta^2 \left(\hat{B} - S\right)^2}{(\delta + 2)(1+\delta)} + \frac{\left(\overline{B} - \hat{B}\right)\left(\hat{B} - S\right)}{1+\delta} \end{aligned}$$

and the maximizing $\hat{B}$ is

$$\hat{B} = \frac{S(2 + \delta - 2\delta^2) + \overline{B}(2+\delta)}{2(1+\delta)(2-\delta)}$$

With biases ε, τ for the seller and the buyer we have $S = \varepsilon$ and $\overline{B} = 1 - \tau$. Thus, the true ex ante expected payoffs of the seller and the buyer, i.e. their payoffs in the “game between the genes" are

$$f_s(\varepsilon_s,\varepsilon_b)=\frac{\delta^2\left(\bar{B}-\varepsilon\right)^2}{(\delta+2)\,(1+\delta)}+\frac{\left(1-\tau-\bar{B}\right)\left(\bar{B}-\varepsilon\right)}{1+\delta}$$
$$=\frac{1}{4}\,(1-\tau-\varepsilon)\,\frac{-2\tau-\tau\delta-4\varepsilon\delta^3+\delta+9\varepsilon\delta+2+2\varepsilon}{(1+\delta)^2\,(2-\delta)}$$

$$f_b(\tau,\varepsilon)=\int_{\bar{B}}^{1-\varepsilon_b}(B+\tau-\frac{\bar{B}+\delta\varepsilon}{1+\delta})dB+\int_{\varepsilon}^{\bar{B}}\delta e^{-r\beta(B)}(B+\tau-p(B,\varepsilon_s))dB$$
$$=\frac{1}{8}\,(1-\tau-\varepsilon)\,\frac{-4\delta^3\tau+11\tau\delta+6\tau-2\varepsilon+2-9\varepsilon\delta+9\delta+4\varepsilon\delta^3-4\delta^3}{(1+\delta)^2\,(2-\delta)}$$

These payoff functions are continuous, twice differentiable, and strictly concave with respect to the gene's chosen bias. The reaction functions

$$\varepsilon(\tau)\equiv\arg\max_{\varepsilon}f_s(\varepsilon,\tau)=\frac{2\delta\left(2-\delta^2\right)}{2+9\delta-4\delta^3}\,(1-\tau)$$
$$\tau(\varepsilon)\equiv\arg\max_{\tau}f_b(\tau,\varepsilon)=\frac{\delta+2}{6+11\delta-4\delta^3}\,(1-\varepsilon)$$

have slope smaller than 1 in absolute value (since $\delta\in(0,1)$). Thus, under any regular payoff-monotonic dynamics, the bias distributions $(\mathcal{E}_t,\mathcal{T}_t)$ will converge in distribution to the unit mass at the unique Nash equilibrium of the "genes game", which is

$$\varepsilon^*=4\delta\frac{2-\delta^2}{6+19\delta-8\delta^3}$$
$$\tau^*=\frac{2+\delta}{6+19\delta-8\delta^3}$$

It is now straightforward to verify that $\varepsilon^*,\tau^*>0$ for $0<\delta<1$, and that $(\varepsilon^*,\tau^*)\rightarrow_{\delta\rightarrow1}(\frac{4}{17},\frac{3}{17})$. $\blacksquare$

9.3 Alternating offers with two-sided uncertainty

Proof of Proposition 4 (sketch). Tedious but straightforward computations[29] show that without biases, the average probability of trade[30] is $P=\frac{3}{16}$, and the average ex ante profit of each trader is $U=\frac{3}{64}$. The bargaining procedure is symmetric in the sense

---

[29] Available from the authors.

[30] When the discounting caused by the delay before the trade is interpreted as the probability of that trade.

of section 4.1. With the uniform distribution, $\alpha = 1$ in the notation of that section. Substitution in (4.8) yields the emerging biases

$$\varepsilon^* = \tau^* = \frac{P - (2\alpha + 1)\,U}{P + (2\alpha + 1)\,(P - 2U)} = \frac{1}{10}$$

Inequality (4.5)

$$\varepsilon, \tau < \frac{2P - (2\alpha + 1)\,U}{2P + (P - 2U)\,(2\alpha + 1)} = \frac{5}{14}$$

determines the upper bounds for the bias intervals E and T. $\blacksquare$

9.4 Repeated Screening

Proof of Proposition 6. With biases ε, τ for the seller and the buyer, respectively, we have that $S = \varepsilon$ and B is uniformly distributed in the interval $[-\tau, 1-\tau]$. At the Sobel-Takahashi (1983) equilibrium, the true expected ex ante payoff for the seller and for the buyer are therefore, respectively

$$\begin{aligned} f_s(\varepsilon,\tau) &= \int_{\varepsilon}^{1-\tau} \Pi_s(S,B;\varepsilon,\tau)dB = \sum_{i=1} \delta^{i-1} P_i \left(\overline{B}_{i-1} - \overline{B}_i\right) \\ &= \sum_{i=1} \delta^{i-1} \left(\varepsilon + cd^{i-1}\,(1-\tau-\varepsilon)\right)(d^{i-1} - d^i)\,(1-\tau-\varepsilon) \\ &= (1-\tau-\varepsilon)\,(1-d)\left(\frac{c(1-\delta d)\,(1-\tau-\varepsilon) + \varepsilon(1-\delta d^2)}{(1-\delta d^2)\,(1-\delta d)}\right) \end{aligned}$$

$$\begin{aligned} f_b(\tau,\varepsilon) &= \sum_{i=1} \delta^{i-1} \int_{\overline{B}_i}^{\overline{B}_{i-1}} (B + \tau - P_i)dB \\ &= \sum_{i=1} \delta^{i-1} \int_{\varepsilon + d^i(1-\tau-\varepsilon)}^{\varepsilon + d^{i-1}(1-\tau-\varepsilon)} \left(B + \tau - \left(\varepsilon + cd^{i-1}\,(1-\tau-\varepsilon)\right)\right)dB \end{aligned}$$

$$= (1-\tau-\varepsilon)\,(1-d)\left(\frac{(2c(1-\delta d) - d(1-\delta))(1-\tau) + \varepsilon\,(1+d-2c)\,(1-\delta d) - (1-\delta d^2)(\tau+1)}{2\,(1-\delta d^2)\,(1-\delta d)}\right)$$

These payoff functions in the "game between the genes" are continuous, twice differentiable, and strictly concave with respect to the gene's chosen bias. The reaction functions are hence

$$\varepsilon(\tau) \equiv \arg\max_{\varepsilon} f_s(\varepsilon,\tau) = \frac{1}{2}\frac{(1-\tau)(2c(1-\delta d) + \delta d^2 - 1)}{c(1-\delta d) - (1-\delta d^2)}$$

$$\tau(\varepsilon) \equiv \arg\max_{\tau} f_b(\tau,\varepsilon) = \frac{(1-\varepsilon)\,(2c(1-\delta d) - d(1-\delta))}{2c(1-\delta d) - d(1-\delta) + (1-\delta d^2)}$$

It is straightforward to check these reaction functions have slopes smaller than 1 in absolute value. Thus, under any regular payoff-monotonic dynamics, the bias distributions $(\mathcal{E}_t, \mathcal{T}_t)$ will converge in distribution to the unit mass at the unique Nash equilibrium of the "genes game", which are

$$\varepsilon^* = \frac{(1-\delta d)\left((1-\delta d^2) - 2c(1-\delta d)\right)}{(1-\delta d^2)(2-d(1+\delta)) + 2c\delta d(1-d)(1-\delta d)}$$
$$\tau^* = \frac{(1-\delta d^2)\left(2c(1-\delta d) + d(\delta-1)\right)}{(1-\delta d^2)(2-d(1+\delta)) + 2c\delta d(1-d)(1-\delta d)}$$

At the Sobel-Takahashi (1983) equilibrium, it turns out that

$$c = \frac{\sqrt{1-\delta} - (1-\delta)}{\delta}$$
$$d = \left(\frac{\sqrt{1-\delta} - (1-\delta)}{\delta\sqrt{1-\delta}}\right)$$

Substitution then yields

$$\varepsilon^* = \frac{2\left(\sqrt{(1-\delta)}(\delta-4) - 3\delta + 4\right)}{10 - 6\delta + \sqrt{(1-\delta)}(\delta-10)}$$
$$\tau^* = \frac{4 - 5\delta + \delta^3 + \sqrt{(1-\delta)}(\delta^2 - 4 + 3\delta)}{\left(10 - 6\delta + \sqrt{(1-\delta)}(\delta-10)\right)\left(\sqrt{(1-\delta)}(1+\delta) - 1 + \delta\right)}$$

It is now straightforward to verify that $\varepsilon^*, \tau^* > 0$ for $0 < \delta < 1$, and that $(\varepsilon^*, \tau^*) \rightarrow_{\delta\rightarrow 1} (\frac{1}{2}, 0)$.

10 References

Admati A. R. and M. Perry (1987), "Strategic Delay in Bargaining", *Review of Economic Studies* **54**, pp. 345-364.

Arlen, J., M. Spitzer and E. Talley (2001), "Endowment Effects Within Corporate Agency Relationships", mimeo, University of Southern California.

Babcock L. and G. Loewenstein (1997), "Explaining Bargaining Impasse: The Role of Self-Serving Biases," *Journal of Economic Perspectives,"* **11(1)**, pp. 109-126.

Bester H. and W. Güth (1998), "Is Altruism Evolutionary Stable?", *Journal of Economic Behavior and Organization* **34(2),** pp. 211-221.

Bolle F. (2000), “Is Altruism Evolutionarily Stable? And Envy and Malevolence? - Remarks on Bester and Güth,” *Journal of Economic Behavior and Organization* **42(1),** pp. 131-133.

Chatterjee K. and W. Samuelson (1983), “Bargaining under Incomplete Information”, *Operations Research* **31**, pp. 835-851.

Cramton P. (1992), “Strategic Delay in Bargaining with Two Sided Uncertainty”, *Review of Economic Studies* **59**, pp. 205-225.

Fershtman C. and Y. Weiss (1997), “Why do We Care about what Others Think about Us?,” in: Ben Ner, A. and L. Putterman (eds.), *Economics, Values and Organization*, Cambridge University Press, Cambridge MA.

Fershtman C. and Y. Weiss (1998), “Social Rewards, Externalities and Stable Preferences,” *Journal of Public Economics* **70**, pp. 53-74.

Frank R.H. (1988), *Passions Within Reason – The Strategic Role of the Emotions,* W.W. Norton & Company, New York.

Güth W. and M. Yaari (1992), “Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach,” in Witt, U. (ed.), *Explaining Forces and Changes: Approaches to Evolutionary Economics,* University of Michigan Press.

Heifetz, A. and Y. Spiegel (2000), “On the Evolutionary Emergence of Optimism”, mimeo,
http://econ.tau.ac.il/papers/foerder/24-2000.pdf

Horowitz, J.K. and K.E. McConnel (2000), “A Review of WTA/WTP Studies,” mimeo, University of Maryland.

Huck S., G. Kirchsteiger, and J. Oechssler (1997), “Learning To Like What You Have – Explaining the Endowment Effect,” mimeo, Humboldt University, Berlin.

Huck S. and J. Oechssler (1998), “The Indirect Evolutionary Approach to Explaining Fair Allocations,” *Games and Economic Behavior* **28**, pp. 13-24.

Kahneman, D., Knetch, J.L and R.H. Thaler (1990), “Experimental Tests of the Endowment Effect and the Coase Theorem”, *Journal of Political Economy* **98**, pp. 1325-1348

Koçkesen L., E.A. Ok, and R. Sethi (1998), “Evolution of Interdependent Preferences in Aggregative Games,” C.V. Starr RR# 98-19, New York University.

Leininger W., P.B. Linhart and R. Radner (1989), “Equilibria of the Sealed-Bid Mechanism for Bargaining with Incomplete Information”, *Journal of Economic Theory* **48**, pp.63-106.

Loewenstein, G. and D. Adler (1995), "A Bias in the Prediction of Tastes", *The Economic Journal* **105**, pp. 929-937

Moulin H. (1984), "Dominance Solvability and Cournot Stability," *Mathematical Social Sciences,* **7(1)**, pp. 83-102.

Marshall, J.D., Knetch, J. L. and J.A. Sinden (1987), " 'Agents' Evaluations of the Disparity in Measures of Economic Loss", *Journal of Economic Behavior and Organization* **7**, pp. 115-127.

Myerson R. and M. Satterthwaite (1983), "Efficient Mechanisms for Bilateral Trading", *Journal of Economic Theory* **29**, pp.265-281.

Oechssler J. and F. Riedel (2001), "Evolutionary Dynamics on Infinite Strategy Spaces," *Economic Theory* **17**, pp. 141-162.

Possajennikov A. (2000), "On The Evolutionary Stability of Altruistic and Spiteful Preferences," *Journal of Economic Behavior and Organization* **42(1)** pp. 125-129

Radner R. and A. Schotter (1989), "The Sealed-Bid Mechanism: An Experimental Study", *Journal of Economic Theory* **48**, pp.179-220.

Rubinstein A. (1982), "Perfect Equilibrium in a Bargaining Model", *Econometrica* **50**, pp. 97-110.

Samuelson L. and J. Zhang (1992), "Evolutionary Stability in Asymmetric Games," *Journal of Economic Theory* **57**, pp. 363-391.

Sobel J. and I. Takahashi (1983), "A Multistage Model of Bargaining", *Review of Economic Studies* **50**, pp. 411-426.