

Benkert, Jean-Michel

**Working Paper**

## Bilateral trade with loss-averse agents

Working Paper, No. 188

**Provided in Cooperation with:**

Department of Economics, University of Zurich

*Suggested Citation:* Benkert, Jean-Michel (2016) : Bilateral trade with loss-averse agents, Working Paper, No. 188, University of Zurich, Department of Economics, Zurich, <https://doi.org/10.5167/uzh-109940>

This Version is available at:

<https://hdl.handle.net/10419/162412>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



**University of  
Zurich** <sup>UZH</sup>

University of Zurich  
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)  
ISSN 1664-705X (online)

---

Working Paper No. 188

## **Bilateral Trade with Loss-Averse Agents**

Jean-Michel Benkert

Revised version, September 2016

---

# Bilateral Trade with Loss-Averse Agents

Jean-Michel Benkert\*

This version: September 2016

First version: November 2014

## Abstract

We study the bilateral trade problem put forward by Myerson and Satterthwaite (1983) under the assumption that agents are loss-averse, using the model developed by Kőszegi and Rabin (2006, 2007). We show that the endowment effect increases the sellers information rent, and that the attachment effect reduces the buyer's information rent. Further, depending on the distribution of types, loss-aversion can reduce the severity of the impossibility problem. However, the result cannot be reversed. Turning to the design of optimal mechanisms, we show that in both revenue and welfare maximizing mechanisms the designer optimally provides the agents with full insurance in the money dimension and with partial insurance in the trade dimension. In fact, when the stakes are large, loss-aversion can eliminate trade altogether. We show that all results display robustness to the exact specification of the reference point and provide some results on general mechanism design problems.

*Keywords:* Bilateral Trade, Loss-Aversion, Mechanism Design

*JEL Classification:* C78, D02, D03, D82, D84

---

\*University of Zurich, Department of Economics, Bluemlisalpstrasse 10, CH-8006 Zurich, Switzerland and UBS International Center of Economics in Society at the University of Zurich. Email: jean-michel.benkert@econ.uzh.ch. I would like to thank Olivier Bochet, Juan Carlos Carbajal, Eddie Dekel, Jeff Ely, Samuel Häfner, Fabian Herweg, Heiko Karle, Botond Kőszegi, René Leal Vizcaíno, Igor Letina, Shou Liu, Daniel Martin, Konrad Mierendorff, Georg Nöldeke, Wojciech Olszewski, Anne-Katrin Roesler, Yuval Salant, Aleksei Smirnov, Ran Spiegler, Egor Starkov, Tom Wilkening, Peio Zuazo Garin, and seminar participants in Zurich and at the ZWE 2014 for helpful comments. I am especially grateful to my supervisor Nick Netzer for his guidance as well as numerous comments and suggestions. I would like to thank the University of Basel and Northwestern University for their hospitality while some of this work was conducted and the UBS International Center of Economics in Society at the University of Zurich as well as the Swiss National Science Foundation (Doc.Mobility Grant P1ZHP1\_161810) for financial support. All errors are my own.

# 1 Introduction

In many situations people evaluate an outcome relative to some reference point. For instance, whether a house owner is willing to sell her house at some price, may depend on whether or not that price is higher than the original purchase price (Genesove and Mayer, 2001). Relatedly, if a buyer expects a trade to go through, her willingness to pay for the good may increase (Ericson and Fuster, 2011). Evidence suggests that the most relevant type of reference-dependent preferences is loss-aversion (see DellaVigna, 2009, for a survey).<sup>1</sup> Kahneman and Tversky's (1979) prospect theory established the importance and relevance of loss-aversion early on, and the literature on this phenomenon has grown substantially since. In particular, a large body of literature finds evidence of loss-aversion in trade situations.<sup>2</sup> When loss-averse people trade, so-called behavioral effects may arise and interfere. The probably best-known such behavioral effect is the *endowment effect* (Thaler, 1980), which says that a person values a good she owns more, simply because she owns it. More recently, an *attachment effect*, which says that if a person expects to buy a good, her willingness to pay for it may increase, has been documented, too (Ericson and Fuster, 2011). In spite of this empirical evidence, the question of the effects of loss-aversion on trade has not been addressed by the theoretical literature. In this paper, we aim to fill this gap and study the bilateral trade problem put forward by Myerson and Satterthwaite (1983) (henceforth MS) under the assumption that agents are loss-averse. To do so, we make use of the model developed by Kőszegi and Rabin (2006, 2007) (henceforth KR), in which the reference point is formed endogenously as the expectation over the outcome.<sup>3</sup>

In the bilateral trade problem, a privately informed seller wants to sell one unit of an indivisible good to a privately informed buyer and both agents have quasi-linear utility over ownership of the good and money. We depart from the classic framework in MS and allow for both agents to have reference-dependent preferences as modeled in KR. More precisely, an agent derives the standard *material utility* from ownership of the good and money, and, in addition, experiences *gain-loss utility* with respect to both, money and ownership of the good, separately. Thus, following KR we assume that the agents bracket gains and losses narrowly. Further, we employ the choice acclimating personal equilibrium (CPE) introduced in Kőszegi and Rabin (2007) as our equilibrium concept. The reference point is thus formed endogenously as the rational expectation over the outcome and agents take an optimal action, taking into account that this action determines their reference

---

<sup>1</sup>There is a substantial literature providing evidence of loss-aversion, e.g., Fehr and Goette (2007), Post, van den Assem, Baltussen, and Thaler (2008), Crawford and Meng (2011) and Pope and Schweitzer (2011).

<sup>2</sup>See Ericson and Fuster (2014) for an excellent review on the role of loss-aversion in explaining (behavioral) effects in exchange situations, and, in particular, the endowment effect.

<sup>3</sup>Ericson and Fuster (2011), Abeler, Falk, Goette, and Huffman (2011), Crawford and Meng (2011), Gill and Prowse (2012), Karle, Kirchsteiger, and Peitz (2015), and Bartling, Brandes, and Schunk (2015) provide evidence for the assumption that the reference point is determined by expectations.

point and the eventual outcome. In particular, this framework gives rise to the endowment and attachment effect and allows us to study their effect on trade, building on established models.

The natural starting point in the analysis is the question of the feasibility of ex post efficient trade, that is, trade taking place whenever the buyer values the good more than the seller, while keeping a balanced budget. The famous impossibility result in MS shows that it is impossible to implement ex post efficient trade under incentive compatibility and individual rationality. This result is commonly interpreted in terms of the difference between the gains from trade and the information rents: trade between the buyer and the seller does not create enough gains to cover the information rents that need to be given to the agents. It turns out that a good way to approach the problem of the feasibility of ex post efficient trade with loss-averse agents is to study the effect of loss-aversion on the gains from trade and the information rents. Given that the agents are loss-averse, they dislike ex-post variations in their payoffs. Hence, on average, the presence of gain-loss utility reduces the overall utility of both agents and the gains from trade. As predicted by the attachment and endowment effect, however, we find that loss-aversion affects the behavior of the buyer and seller differently: the buyer has a weaker incentive to misreport her type in the presence of loss-aversion, while the seller has a stronger incentive to do so. As a consequence, the buyer's information rent decreases in the presence of loss-aversion, whereas the seller's information rent increases. Hence, loss-aversion has an ambiguous effect on the sum of information rents, which makes it a priori unclear whether the famous impossibility result persists in the presence of loss-aversion. Nevertheless we can show that the impossibility result persists in same generality as in MS, that is, for arbitrary and asymmetric distributions with full and overlapping support. It is noteworthy that the attachment effect can actually mitigate the impossibility problem in the sense of requiring a lower subsidy to induce materially efficient trade under incentive compatibility and individual rationality. The impossibility result cannot be reversed, however, because incentive compatibility puts limits on how loss-averse the agents can be, which limits the strength of the attachment effect. As we show at the end of the paper in a robustness section, this limiting effect of incentive compatibility on the strength of behavioral effects extends to other models of reference-dependent utility than the one considered in the main analysis.

Having confirmed the impossibility result in the presence of loss-aversion, we turn to the problem of designing optimal mechanisms. In the revenue maximizing mechanism, the designer acts as an intermediary between the seller and the buyer and tries to make a profit from the resulting trade. In the welfare maximizing mechanism, the designer wants to maximize the agents' sum of utilities subject to a budget constraint. We show that in the presence of loss-aversion any revenue or welfare maximizing mechanism features what

we call *interim deterministic transfers*, that is, the transfer of an agent is independent of the other agent's report and is thus deterministic given her own type. Turning to the optimal trade rule, we impose the assumption that types are drawn from the uniform distribution to keep the model tractable. However, it is not possible to obtain the optimal trade rule using pointwise maximization because the agents' expected utility depends on the mechanism through the reference point. In order to derive the optimal trade rule we make use of the reduced-form approach to auctions. Border (1991) characterizes which interim allocation probabilities are implementable by some ex post allocation rule in the case of single-unit auctions. Che, Kim, and Mierendorff (2013) provide a substantial generalization of this result to multi-unit auctions, and also extend the reduced-form approach from auctions to the bilateral trade setting. Thus, instead of maximizing over the ex post trade rule, we maximize directly over the interim trade probabilities and can explicitly derive the optimal trade rule. We show that the designer optimally induces less trade than in the absence of loss-aversion. Thus, the designer eliminates all ex post variation in the agents' transfers, thereby fully insuring them against any losses in the money dimension, and partially insures them against losses in the trade dimension by reducing the trade probability. Full insurance in the trade dimension boils down to trade always or never taking place, which is generally not optimal. For sufficiently high stakes and degrees of loss-aversion, however, the designer indeed provides the agents with full insurance by eliminating trade altogether.

Our final results concern the robustness of the optimal mechanisms and of the impossibility result in the presence of loss-aversion for other specifications of the formation of the reference point. KR note that their equilibrium concept CPE is similar to models of disappointment-aversion such as those introduced by Bell (1985) and Loomes and Sugden (1986). The CPE specifies the reference point endogenously as the full distribution of a lottery, whereas the reference point corresponds to the certainty equivalent of the lottery in these models of disappointment-aversion. Masatlioglu and Raymond (forthcoming) find that the intersection of preferences induced by the CPE and any of the listed disappointment-aversion models is simply expected utility. Thus, although the models seem to be very similar on first glance, the induced preferences are generically different. Nevertheless, we show that the optimal mechanisms derived in this paper for CPE are also optimal for the models by Bell (1985) and Loomes and Sugden (1986) and that the impossibility result continues to hold, too. Further, we briefly explore the possibility of an exogenously given fixed reference point. We model this using the framework from Spiegel (2012) where the agents have an exogenously given reference point and feel losses in case of negative deviations, but feel no gains in the case of positive deviations. While we cannot fully characterize the set of parameters for which the impossibility result extends to this framework, we show that it persists for a large range of parameters, for

instance, whenever the degree of loss-aversion is symmetric across the agents.

Finally, the appendix contains a section on general mechanism design problems with loss-averse agents. In particular, this allows us to prove the optimality of deterministic transfers in revenue and welfare maximizing mechanisms beyond the bilateral trade problem.

This paper is organized as follows. The next section contains a more detailed discussion of the related literature. In Section 3 we present the model, solution concept and notation used throughout the paper. In Section 4 we study the effect of loss-aversion on the gains of trade and information rents in order to address the impossibility result. Section 5 contains the derivation of the revenue and welfare maximizing mechanisms. In Section 6 we show that these optimal mechanisms display robustness to the exact specification of the reference point and Section 7 concludes. All proofs are relegated to the appendix.

## 2 Related literature

The paper contributes to three strands of literature: the literature on bilateral trade, behavioral mechanism design, and theoretical applications of reference-dependent utility in general.

Garratt and Pycia (2015) examine the bilateral trade problem relaxing the assumption that the agents have quasi-linear utility.<sup>4</sup> Allowing for risk-aversion and wealth effects, they show that ex post efficient trade is possible under some conditions. The impossibility result can be reversed in this setting, because the presence of risk-aversion and wealth effects gives rise to additional gains from trade, which then suffice to cover the agents' information rent. In earlier work, Chatterjee and Samuelson (1983) extend their analysis of the double-auction for risk-neutral agents to the case of risk-averse agents. They find that when agents "become infinitely risk-averse" an ex post efficient outcome can be achieved.

Bierbrauer and Netzer (2016) modify the standard mechanism design framework by introducing intention-based social preferences. They study the implementability of incentive-compatible social choice functions and how it depends on the designer's information on the degree of agents' reciprocity. Like the reference point in the present paper, intentions are determined endogenously. In an application to the bilateral trade problem, they show that the impossibility result can be reversed in their setting if the kindness-weights are known. Roughly, the incentive compatibility constraints can be turned slack by introducing an action which generates sufficiently strong feelings of kindness, thereby essentially eliminating any tension between ex post efficiency and the agents' incentives. Kucuksenel

---

<sup>4</sup>See also the references in Garratt and Pycia (2015) for more work on the bilateral trade problem in the classic framework with quasi-linear utility following MS.

(2012) considers the mechanism design problem under the assumption that agents are altruistic. He also considers the bilateral trade problem as an application and finds that the more altruistic agents are, the higher the probability of efficient trade taking place. Intuitively, as agents become more altruistic, their utility becomes more aligned with the expected gains from trade, reducing the tension between ex post efficiency and the agents' incentives. Wolitzky (2016) applies general results on mechanism design with agents who maximize expected maxmin utility to the bilateral trade problem. He fully characterizes when the impossibility result by MS is reversed or when it persists in this setting. A reversal is possible because in the case of maxmin agents, the gains from trade need only be larger than the sum of agents' minimal information rents.

Cavallo (2011) considers the problem of reallocating a good from an agent to other agents, using an individual rationality constraint which is required to be satisfied before an agent learns her own type, having already learned the other agents' types. This individual rationality constraint allows him to implement ex post efficient social choice functions. In this setting, he considers a mechanism in which the designer extracts all surplus from the agents, a mechanism in which all agents and the designer obtain the same expected share of surplus, and a mechanism which aims at minimizing the risk of agents being worse off ex post than if they had not participated. Finally, the three mechanisms are evaluated numerically under the assumption that agents are loss-averse, finding that the third mechanism remains efficient for higher degrees of loss-aversion than the two other mechanisms. Salant and Siegel (2016) study the efficient allocation of a divisible asset for different types of reallocation costs. For concave reallocation cost, the initial allocation can be interpreted as the reference point and deviations from the reference point lead to losses (but no gains) that are symmetric across agents. They show that expected surplus is maximized for fully concentrated initial allocations, i.e., completely opposite reference points across agents, but even then ex post efficiency may not be attained, suggesting some robustness of the impossibility result with a fixed reference point. We pick up on this robustness to exogenously given reference points in Section 6.

There are numerous theoretical papers working under the assumption of loss-averse agents, while applying it to different settings. Three papers are particularly closely related to ours. Eisenhuth (2013) considers the problem of a risk-neutral seller who wants to maximize revenue by selling a good to loss-averse buyers. Using the framework of KR, he finds that the optimal auction is an all-pay auction with reserve price when agents bracket narrowly, and that it is a first-price auction with reserve price in the case of wide bracketing. In the appendix, we generalize the all-pay-result in the case of narrow bracketing and show that it holds for any revenue maximizing mechanisms and in fact extends to welfare maximizing mechanisms, too. Rosato (2015) considers a sequential



bargaining model with a risk-neutral seller and a loss-averse buyer.<sup>5</sup> In the framework of KR and assuming wide bracketing, he shows that the buyer’s loss-aversion softens the rent-efficiency trade-off for the seller. Just as in the present paper, this is driven by the attachment effect: the buyer is willing to accept lower offers to avoid the risk of a breakdown of the negotiations. Using the dynamic model of reference-dependent utility in Kőszegi and Rabin (2009), Duraj (2015) considers the impact of news utility in mechanism design models.<sup>6</sup> In his framework, in addition to being loss-averse over consumption utility, agents are also loss-averse over changes in beliefs about their current and future consumption. He shows that any mechanism which is incentive compatible in the presence of loss-aversion on news utility, is also incentive compatible in the framework of the present paper. In the context of bilateral trade, he shows on the one hand that, when the realization of the outcome is delayed, the extra slack in the incentive compatibility constraints due to news utility is enough to reverse the impossibility result, contrasting the robustness result in the present paper. On the other hand, he shows that the optimality of deterministic transfers in revenue-maximizing mechanisms in the present paper extends to the setting with news utility and a delayed realization of the outcome. In the case without delay, which proves to be more tractable than the setting with delay as well as the setting in the present paper, he solves for the welfare maximizing mechanism. Less closely related, de Meza and Webb (2007) consider incentive design under loss-aversion, Gill and Stone (2010) model a two-player rank-order tournament when agents are loss-averse, Carbajal and Ely (2016) study optimal price discrimination when a monopolist faces a continuum of consumers with reference-dependent preferences, Rosato (2014) proposes expectations-based loss-aversion as an explanation for the “afternoon effect” observed in sequential auctions, and Karle and Peitz (2014) investigate firm strategy in imperfect competition.

## 3 Model

### 3.1 Utility, Social Choice Functions and Mechanisms

The set of agents is given by  $I = \{S, B\}$  where  $S$  and  $B$  denote seller and buyer, respectively. It is commonly known that the type of agent  $i \in I$  has distribution  $F_i$  with full support on the set  $\Theta_i = [a_i, b_i] \subset \mathbb{R}_+$ , and is private information. Let  $\Theta = \Theta_S \times \Theta_B$ . We

---

<sup>5</sup>See Shalev (2002) and Driesen, Perea, and Peters (2012) for other approaches incorporating loss-aversion to bargaining.

<sup>6</sup>Both Duraj (2015) and Duraj’s master thesis, from which said paper evolved, have been made available to us through personal communication. We thank Niccolò Lomys for making the connection. In the master thesis, Duraj also derives some results in the framework of the present paper. In particular, imposing stronger symmetry assumptions than here, he proves the robustness of the impossibility result and the optimality of deterministic transfers in revenue maximizing mechanisms.

interpret the type of an agent as her valuation of the good.<sup>7</sup> A social alternative is given by  $\mathbf{x} = (y, t_S, t_B) \in X = \{0, 1\} \times \mathbb{R}^2$ , where  $y$  indicates whether or not trade takes place and  $t_S$  and  $t_B$  denote the respective transfers of the seller and buyer.

Following KR, we allow for the agents to be loss-averse in the trade and in the money dimension. That is, the buyer derives the standard material utility from obtaining and paying for the good, and additionally, the buyer feels weighted gain-loss utility with respect to getting the good as well as weighted gain-loss utility with respect to paying for the good. Loss-aversion is captured by value functions in the sense of Kahneman and Tversky (1979) given by

$$\mu_i^k(x) = \begin{cases} x & \text{if } x \geq 0, \\ \lambda_i^k x & \text{else,} \end{cases}$$

for some  $\lambda_i^k > 1$ , which reflects the degree of loss-aversion.<sup>8</sup> Thus, the riskless total utility is given by

$$u_S(\mathbf{x}, \mathbf{r}_S, \theta_S) = (1 - y)\theta_S + t_S + \eta_S^1 \mu_S^1(r_S^1 \theta_S - y\theta_S) + \eta_S^2 \mu_S^2(t_S - r_S^2) \quad (1)$$

$$u_B(\mathbf{x}, \mathbf{r}_B, \theta_B) = y\theta_B - t_B + \eta_B^1 \mu_B^1(y\theta_B - r_B^1 \theta_B) + \eta_B^2 \mu_B^2(r_B^2 - t_B) \quad (2)$$

where  $\eta_i^k \geq 0$  are the weights put on gain-loss utility. The parameters  $\mathbf{r}_i = \{r_i^1, r_i^2\}$  are the so-called riskless reference level. Following KR we will allow the reference point to be the agent's rational expectations and therefore a probability distribution over all riskless reference levels (see more below). We will refer to  $(1 - y)\theta_S + t_S$  and  $y\theta_B - t_B$  as material utility and to the other terms as gain-loss utility in the trade and money dimension, respectively. We adopt the following assumption from Herweg, Müller, and Weinschenk (2010):<sup>9</sup>

**Assumption 1 (No Dominance of Gain-Loss Utility)**  $\Lambda_i = \eta_i^1(\lambda_i^1 - 1) \leq 1, i \in I$ .

This assumption ensures that gain-loss utility does not dominate material utility and plays an important role for incentive compatibility. We will maintain this assumption throughout the paper and discuss the implications of relaxing it after deriving the impossibility

---

<sup>7</sup>We could alternatively assume that the seller does not own the good but has to produce it. The seller's type would then represent her marginal cost of production. All the results that follow would go through in this case.

<sup>8</sup>We follow the literature by abstracting from diminishing sensitivity.

<sup>9</sup>This condition is commonly imposed, see for instance de Meza and Webb (2007), Eisenhuth and Ewers (2012), Eisenhuth (2013), Karle and Peitz (2014), and Rosato (2014). KR show that this condition ensures that agents will not choose stochastically dominated options.

result in Section 4. We follow KR by assuming that there is a separate gain-loss term for each of the two material utility dimensions, trade and money utility.<sup>10</sup>

A social choice function (SCF)  $f : \Theta \rightarrow X$  assigns a collective choice  $f(\theta_S, \theta_B) \in X$  to each possible profile of the agents' types  $(\theta_S, \theta_B) \in \Theta$ . In the present bilateral trade setting, a social choice function takes the form  $f = (y^f, t_S^f, t_B^f)$ . Let  $\mathcal{F}$  denote the set of all SCFs and  $\mathcal{Y}$  the set of all trade mechanisms, i.e., the set containing all  $y^f$ . A mechanism  $\Gamma = (M_S, M_B, g)$  is a collection of message sets  $(M_S, M_B)$  and an outcome function  $g : M_S \times M_B \rightarrow X$ . We denote the direct mechanism by  $\Gamma^d = (\Theta_S, \Theta_B, f)$ . Since agents privately observe their types, they can condition their message on their type. Consequently, a pure strategy for agent  $i$  in a mechanism  $\Gamma$  is a function  $s_i : \Theta_i \rightarrow M_i$ . Note that  $g(s_S(\theta_S), s_B(\theta_B)) \in X$ . Let  $S_i$  denote the set of all pure strategies of agent  $i$ . Further, we denote the truthful strategy  $s_i^t(\theta_i) = \theta_i$ . Throughout, the operator  $\mathbb{E}_{-i}$  denotes the expectation over the random variables  $\tilde{\theta}_{-i}$  taking the value  $\theta_i$  as given.

### 3.2 Equilibrium Concept and Revelation Principle

We use the concept of an (interim) choice-acclimating personal equilibrium (CPE) introduced in Kőszegi and Rabin (2007). The set of all riskless reference levels is given by the set of all social alternatives  $X$ . Essentially, the set  $X$  captures all the outcomes that could materialize at the end of the agents' interaction. In a mechanism  $\Gamma$ , agent  $i$ 's action induces a distribution over the set of social alternatives  $X$ , conditional on the other agent playing  $s_{-i}$ . It is this endogenously generated distribution over  $X$  that forms the agent's reference point, or rather, reference distribution in a CPE. Effectively, when an agent evaluates an outcome, she is comparing it to all other possible social alternatives that could have materialized given the distribution induced over them. Moreover, when the agent takes an action in a CPE, she takes the action anticipating that it will not only determine the outcome of the mechanism, but also the distribution over the set  $X$  and, therefore, the reference point.

Moving to the interim stage and allowing the reference point to be the agent's rational expectations, we can define the interim expected utility of the seller with type  $\theta_S$ , in the mechanism  $\Gamma$ , when playing action  $m \in M_B$ , given that the buyer plays strategy  $s_B$  as

$$U_S(m, s_B, \Gamma | \theta_S) = \int_{a_B}^{b_B} (1 - y^g(m, s_B(\theta_B)))\theta_S + t_S^g(m, s_B(\theta_B)) dF_B(\theta_B)$$

---

<sup>10</sup>The assumption that the loss-aversion parameters are commonly known may seem restrictive. However, we are essentially assuming that the functional form of the utility function is common knowledge, thereby following for instance Maskin and Riley (1984) who assume in their study of optimal auctions with risk-averse buyers that the buyers' parameter of risk-aversion is commonly known.

$$\begin{aligned}
& + \int_{a_B}^{b_B} \int_{a_B}^{b_B} \eta_S^1 \mu_S^1 (y^g(m, s_B(\theta'_B)) \theta_S - y^g(m, s_B(\theta_B)) \theta_S) dF_B(\theta'_B) dF_B(\theta_B) \quad (3) \\
& + \int_{a_B}^{b_B} \int_{a_B}^{b_B} \eta_S^2 \mu_S^2 (t^g(m, s_B(\theta_B)) - t^g(m, s_B(\theta'_B))) dF_B(\theta'_B) dF_B(\theta_B) \\
& = \theta_S \int_{a_B}^{b_B} (1 - y^g(m, s_B(\theta_B))) d\theta_B + \int_{a_B}^{b_B} t_S^g(m, s_B(\theta_B)) dF_B(\theta_B) \\
& + \theta_S \eta_S^1 \int_{a_B}^{b_B} \int_{a_B}^{b_B} \mu_S^1 (y^g(m, s_B(\theta'_B)) - y^g(m, s_B(\theta_B))) dF_B(\theta'_B) dF_B(\theta_B) \\
& + \eta_S^2 \int_{a_B}^{b_B} \int_{a_B}^{b_B} \mu_S^2 (t_S^g(m, s_B(\theta_B)) - t_S^g(m, s_B(\theta'_B))) dF_B(\theta'_B) dF_B(\theta_B).
\end{aligned}$$

The expression in (3) may require some explanation. The first line corresponds to material utility, the second to gain-loss utility in the trade dimension and the third to gain-loss utility in the money dimension. The double integral has a clear intuition. To illustrate, consider the third line containing the money gain-loss utility. Fix any  $\theta_B$  in the domain of integration of the outer integral and suppose this was the actual realization of the buyer's type. The seller would then receive a transfer of  $t_S^g(m, s_B(\theta_B))$ , which she would compare to the reference point. The reference point, or rather distribution, is induced endogenously and corresponds to the distribution of possible transfers. Thus, for every  $\theta'_B$  in the domain of the inner integral we get a possible transfer  $t_S^g(m, s_B(\theta'_B))$  given the buyer's strategy and the seller's message. The seller compares the actual transfer  $t_S^g(m, s_B(\theta_B))$  with all these other possible transfers and the value function  $\mu_S^2$  weights these comparisons differently, depending on whether they result in a loss or a gain. The inner integral then aggregates the gains and loss weighted by the induced probability distribution. Next, integrate over all the values  $\theta_B$  in the domain of the outer integral to get the familiar interim expected utility. In summary, the seller aggregates over each possible realization of transfers and for each of these possible realizations she compares the outcome with all other possible outcomes, aggregating gains and losses in each comparison.

Given our interpretation that the seller owns the good, her outside option is type-dependent and given by  $\theta_S$ . To simplify notation later, we will consider the seller's net utility from trade, which, with some abuse of notation, allows us to compactly write  $U_S(m, s_B, \Gamma | \theta_S) = -\theta_S \tilde{v}_S(m) + \tilde{t}_S(m)$ , where

$$\begin{aligned}
\tilde{v}_S(m) &= \int_{a_B}^{b_B} y^g(m, s_B(\theta_B)) dF_B(\theta_B) \\
&\quad - \eta_S^1 \int_{a_B}^{b_B} \int_{a_B}^{b_B} \mu_S^1 (y^g(m, s_B(\theta'_B)) - y^g(m, s_B(\theta_B))) dF_B(\theta'_B) dF_B(\theta_B), \\
\tilde{t}_S(m) &= \int_{a_B}^{b_B} t_S^g(m, s_B(\theta_B)) dF_B(\theta_B)
\end{aligned}$$

$$+ \eta_S^2 \int_{a_B}^{b_B} \int_{a_B}^{b_B} \mu_S^2 (t_S^g(m, s_B(\theta_B)) - t_S^g(m, s_B(\theta'_B))) dF_B(\theta'_B) dF_B(\theta_B).$$

This compact notation highlights the fact that not only material utility, but also overall utility is linear in the type. Moreover, it will turn out to be useful to further define

$$\begin{aligned} \bar{t}_S(m) &= \int_{a_B}^{b_B} t_S^g(m, s_B(\theta_B)) dF_B(\theta_B), \\ w_S(m) &= \int_{a_B}^{b_B} \int_{a_B}^{b_B} \mu_B^2 (t_S^g(m, s_B(\theta_B)) - t_S^g(s_S(\theta_S), s_B(\theta'_B))) dF_B(\theta'_B) dF_B(\theta_B), \end{aligned}$$

allowing us to write  $\tilde{t}_S(m) = \bar{t}_S(m) + \eta_S^2 w_S(m)$ . Similarly, we can write the buyer's utility as  $U_B(m, s_S, \Gamma | \theta_B) = \theta_B \tilde{v}_B(m) + \tilde{t}_B(m)$ , defining the functions  $\tilde{v}_B$  and  $\tilde{t}_B$  analogously.

We can now define our equilibrium concept, which follows Eisenhuth (2013).

**Definition 1** A strategy profile  $s^* = (s_S^*, s_B^*)$  is a CPE of the mechanism  $\Gamma = (M_S, M_B, g)$  if  $s_i^*(\theta_i) \in \arg \max_{m_i \in M_i} U_i(m_i, s_{-i}^*, \Gamma | \theta_i)$  for all  $i \in I$  and  $\theta_i \in \Theta_i$ .

**Definition 2** A mechanism  $\Gamma$  implements a SCF  $f$  if there is a CPE strategy profile  $s = (s_S, s_B)$  such that  $g(s_S(\theta_S), s_B(\theta_B)) = f(\theta_S, \theta_B)$  for all  $(\theta_S, \theta_B) \in \Theta$ .

**Definition 3** A SCF  $f$  is CPE incentive compatible (CPEIC) if the truthful profile  $s^t = (s_S^t, s_B^t)$  is a CPE strategy in the direct mechanism  $\Gamma^d$ .

As a first result we note that the revelation principle for CPE holds in our setting.

**Proposition 1 (Revelation Principle for CPE)** A social choice function  $f$  can be implemented in CPE by some mechanism  $\Gamma$  if and only if  $f$  is CPEIC.

The proof is contained in Appendix A, where we prove the revelation principle for general social choice functions, thus showing that its validity extends beyond the bilateral trade setting. Henceforth, we focus on direct mechanisms and no longer explicitly list the mechanism as an argument in the utility function.

### 3.3 Incentive Compatibility and Efficiency

In this section we characterize the set of all CPEIC social choice functions and introduce some familiar concepts, such as individual rationality and ex post budget balance. Moreover, we take a closer look at the materially efficient SCF, i.e., trade being induced whenever the buyer's valuation exceeds the seller's marginal cost of production.

**Proposition 2** The SCF  $f = (y^f, t_S^f, t_B^f)$  is CPEIC if and only if,

- (i)  $\tilde{v}_S$  is non-increasing and  $\tilde{v}_B$  is non-decreasing, and

(ii) we can write utility as

$$U_S(\theta_S, s_B^t | \theta_S) = U_S(b_S, s_B^t | b_S) + \int_{\theta_S}^{b_S} \tilde{v}_S(t) dt, \quad (4)$$

$$U_B(\theta_B, s_S^t | \theta_B) = U_B(a_B, s_S^t | a_B) + \int_{a_B}^{\theta_B} \tilde{v}_B(t) dt. \quad (5)$$

The proof is contained in Appendix A, where we prove the result for general social choice functions. Recall that the functions  $\tilde{v}_B$  and  $\tilde{v}_S$  contain terms of gain-loss utility. Thus, while the incentive-compatibility conditions in Proposition 2 take the same form as in the absence of loss-aversion, it need not follow that the set of incentive-compatible SCF coincides. We say that a SCF is individually rational if for both agents  $i \in I$

$$U_i(\theta_i, s_{-i}^t | \theta_i) \geq 0 \quad \forall \theta_i \in \Theta_i, \quad (\text{IR})$$

and that it is ex post budget balanced if

$$t_S^f(\theta_S, \theta_B) = t_B^f(\theta_S, \theta_B), \quad \forall (\theta_S, \theta_B) \in \Theta. \quad (\text{BB})$$

Setting the outside option in (IR) equal to zero is without loss of generality.<sup>11</sup> An agent could choose to walk away and not participate in the mechanism as soon as she learns her type. Doing so would rule out any possibility of trade and payment or receipt of any transfers. Therefore, the reference points of the agent would be equal to zero, as she anticipates that no trade or transfers can take place if she walks away. Consequently, there would be no feelings of gain or loss, as well as zero material utility when the agent walks away.

We say that a mechanism has interim deterministic transfers, when, given her own type, an agent's transfer does not depend on almost all types of the other agent. Similarly, a trade rule is interim deterministic, when, given her own type, the trade rule coincides for almost all types of the other agent. A mechanism with interim deterministic transfers and an interim deterministic trade rule is called interim deterministic.

A trade mechanism is materially efficient if

$$y^f(\theta_S, \theta_B) = \begin{cases} 1 & \text{if } \theta_B \geq \theta_S, \\ 0 & \text{if } \theta_B < \theta_S. \end{cases} \quad (\text{ME})$$

In the classic framework with no loss-aversion, material efficiency and budget balance taken together are equivalent to Pareto efficiency. In particular, MS's impossibility result

---

<sup>11</sup>Recall that we are considering net utility and have thus already taken care of the seller's type-dependent outside option.

shows that no Pareto efficient mechanism simultaneously satisfies individual rationality and incentive compatibility. We will thus use these concepts as a benchmark allowing us to analyze the impact of the introduction of loss-aversion in a familiar environment and to draw a clear comparison to the classic framework.

## 4 Information Rents and Gains From Trade

The impossibility theorem in MS is commonly interpreted in terms of the difference between the gains from trade and the information rents: trade between the buyer and the seller does not create enough gains to cover the information rents that need to be given to the agents. As a consequence, it is not possible to implement materially efficient trade under incentive compatibility and individual rationality without subsidizing the agents. The question arises how and to what extent the presence of loss-aversion changes this result. We have already noted that the attachment effect should facilitate trade, whereas the endowment effect should impede it. In what follows, we will formalize these notions. We begin by considering the effect of loss-aversion on the gains from trade.

**Lemma 1** *Loss-aversion decreases the gains from trade of a mechanism if and only if the mechanism is not interim deterministic.*

The proof of the lemma is straightforward. Loss-averse agents dislike ex post variations in their payoffs, which reduces their interim utility. Only in the case of an interim deterministic mechanism, ex post variations in the transfers and the trade outcome are completely eliminated (from an interim perspective) and therefore loss-aversion does not decrease the gains from trade.

The effect of loss-aversion on the information rents is more subtle and interesting. We will now illustrate this using a simple mechanism. Consider the materially efficient trade rule (ME) with transfers given by

$$\begin{aligned} t_B(\theta_S, \theta_B) &= -\theta_B \tilde{v}_B(\theta_B), \\ t_S(\theta_S, \theta_B) &= \theta_S \tilde{v}_S(\theta_S). \end{aligned}$$

This mechanism is special in two ways. First, under complete information, this mechanism fully extracts all rents from the agents. Hence, the mechanism is individual rational, but, as we will see momentarily, it is not incentive compatible. Second, the transfers are interim deterministic. Hence, the agent does not feel any gains or losses in the money dimension.

We begin by considering the effects of loss-aversion on the buyer. The expected utility of reporting type  $\theta'_B$  when  $\theta_B$  is the agent's true type (and conditional on the seller

reporting her type truthfully) is given by<sup>12</sup>

$$\begin{aligned}
U_B(\theta'_B, s_B^t | \theta_B) &= \theta_B \tilde{v}_B(\theta'_B) - \tilde{t}_B(\theta'_B) \\
&= \underbrace{(\theta_B - \theta'_B) F_S(\theta'_B)}_{\text{material utility}} + \underbrace{\Lambda_B (\theta'_B - \theta_B) (1 - F_S(\theta'_B)) F_S(\theta'_B)}_{\text{gain-loss utility}}. \tag{6}
\end{aligned}$$

In the classic framework of MS without loss-aversion (i.e., with  $\Lambda_B = 0$ ), a buyer of type  $\theta_B$  would have an incentive to imitate a lower type  $\theta'_B$ . This effect is still present as we can see from equation (6). Note that for the material utility we have  $(\theta_B - \theta'_B) F_S(\theta'_B) > 0$  for  $\theta_B > \theta'_B$ , making a downward deviation profitable for the buyer in the same way as it does in the absence of loss-aversion. However, loss-aversion adds a new, countervailing effect: there is an incentive to imitate a *higher* type. When looking at the gain-loss utility in equation (6), we indeed have  $\Lambda_B (\theta'_B - \theta_B) (1 - F_S(\theta'_B)) F_S(\theta'_B) > 0$  for  $\theta_B < \theta'_B$ . The intuition is as follows. Loss-averse agents dislike payoff uncertainty. Since overall utility and, in particular, gain-loss utility is linear in the type, a higher buyer type dislikes the uncertainty more than a lower type. Recall that the mechanism we are considering in this example is fully rent-extracting. This allows us to decompose the transfer in two parts, one extracting the material utility, and the other extracting the gain-loss utility. When a buyer of type  $\theta_B$  truthfully reports her type this yields a gain-loss utility of  $-\Lambda_B \theta_B (1 - F_S(\theta_B)) F_S(\theta_B)$  and the corresponding component in the transfer is given by  $\Lambda_B \theta_B (1 - F_S(\theta_B)) F_S(\theta_B)$  so that the gain-loss (dis-)utility is fully extracted. A deviation to a higher buyer type yields a transfer with a gain-loss component intended to extract the gain-loss utility of type, who values gains and losses more strongly. Thus, imitating a higher type is profitable, as it yields a transfer which extracts the gain-loss (dis-)utility of a higher buyer type and, therefore, leaves the buyer with some rent. The assumption that gain-loss utility does not dominate material utility ( $\Lambda_B \leq 1$ ) ensures that overall the buyer still has an incentive to imitate a lower type. However, in the presence of loss-aversion this incentive is diminished. As a consequence, the buyer's information rent is smaller in the presence of loss-aversion. This reduction in the incentives to imitate a lower type and, in conjunction with that, the decrease in the information rent is precisely the attachment effect. Formally, and more generally, we can observe the reduction in the information rent of the buyer due to the attachment effect in an incentive compatible mechanism using the integral representation of the utility (see Proposition 2).

Turning to the seller, we can write the expected utility of reporting type  $\theta'_S$  when  $\theta_S$  is her true type as

$$U_S(\theta'_S, s_B^t | \theta_S) = -\theta_S \tilde{v}_S(\theta'_S) + \tilde{t}_S(\theta'_S)$$

---

<sup>12</sup>We omit the derivations as they mirror the steps in the proof of Proposition 3 in Appendix B.1.



$$= \underbrace{(\theta'_S - \theta_S)(1 - F_B(\theta'_S))}_{\text{material utility}} + \underbrace{\Lambda_S(\theta'_S - \theta_S)F_B(\theta'_S)(1 - F_B(\theta'_S))}_{\text{gain-loss utility}}. \quad (7)$$

In contrast to the case of the buyer, the analogous exercise as above reveals that the presence of loss-aversion *amplifies* the seller's incentive to imitate a high type. This increase in the incentives to imitate a higher type and in the information rent captures precisely the endowment effect. We summarize these findings in the following lemma.

**Lemma 2** *Loss-aversion in the trade dimension decreases the buyer's and increases the seller's information rent, respectively.*

The overall effect of loss-aversion on the sum of information rents is ambiguous and as a consequence it is a priori unclear whether the impossibility result persists. As we will see, although the severity of the impossibility problem can be mitigated by loss-aversion, it cannot be reversed. The result follows in two steps. First, observe that Lemmas 1 and 2 imply that it is sufficient to show the impossibility in the case when neither the seller nor the buyer are loss-averse in the money dimension, and, moreover, the seller is not loss-averse in the trade dimension either. To see this, note that loss-aversion in the money dimension does not affect the agents' information rents, but may decrease the gains from trade. Thus, loss-aversion in the money dimension makes the problem unambiguously harder. The above discussion of the endowment effect showed that the seller's information rent increases in the presence of loss-aversion in the trade dimension. In addition, loss-aversion in the trade dimension decreases the gains from trade, since the materially efficient trade rule is not interim deterministic. Thus, any loss-aversion on the seller's side makes the problem unambiguously harder. Hence, it suffices to consider the case when the seller is not loss-averse and the buyer is loss-averse in the trade dimension only. Put differently, only the attachment effect could potentially reverse the impossibility result. Making use of this insight, the second step is to proceed analogously to the proof in MS. That is, impose budget balance as well as incentive compatibility to obtain an expression for the sum of utilities of the "worst" buyer and seller types in the materially efficient mechanism and show that it is strictly negative. Indeed, we obtain

$$\begin{aligned} U_B(a_B) + U_S(b_S) &= \\ &- \int_{\max\{a_B, a_S\}}^{\min\{b_S, b_B\}} (1 - F_B(x))F_S(x)(1 - \Lambda_B(1 - F_S(x))) + \Lambda_B(1 - F_S(x))F_S(x)xf_B(x) dx \\ &< 0, \end{aligned} \quad (8)$$

which violates individual rationality for any  $\Lambda_B \leq 1$ . This proves our first main result (see Appendix B.1 for the details).

**Proposition 3** *For any degree of loss-aversion in the money or good dimension there exists no SCF simultaneously satisfying CPEIC, IR, ME and BB.*

The minimal subsidy needed to induce materially efficient trade under CPEIC and IR (see equation (8)) can be interpreted as a measure of the severity of the impossibility problem and will generally depend on the degree of loss-aversion and the distribution of the buyer's types. Indeed, taking the derivative of the minimal subsidy in equation (8) with respect to  $\Lambda_B$ , we can see that the attachment effect mitigates the impossibility problem by dominating the diminishing effect of loss-aversion on the gains from trade whenever

$$\int_{\max\{a_B, a_S\}}^{\min\{b_S, b_B\}} (1 - F_B(x))F_S(x)(1 - F_S(x)) - (1 - F_S(x))F_S(x)xf_B(x) dx \geq 0.$$

To get a feel for this condition, consider the families of distributions  $F_S(x) = x^s$  and  $F_B(x) = x^b$  on  $[0, 1]$  for  $b, s > 0$ . Whenever  $b > 2s^2 - 1$  the buyer's loss-aversion makes the problem easier. In words, the likelier low seller types and high buyer types are, the less severe is the impossibility problem. This is in line with the intuition underlying the attachment effect. When low seller types are likely, a buyer puts a relatively high probability on trade taking place and thus has a strong attachment to the good (a high reference point). Hence, when low seller types and high buyer types are likely, on average the buyer will have a high attachment effect, thereby mitigating the impossibility problem. Note that in the absence of loss-aversion, it is also true that the minimal subsidy is lower the likelier low seller types and high buyer types are. In the presence of the attachment effect, however, this is reinforced.

Another noteworthy point is that for the extreme types, i.e., types who lie outside the intersection of the intervals, loss-aversion does not matter. This finding is very intuitive. To see this, observe that for these types trade is interim deterministic and hence there is no gain-loss utility as there is no room for ex post variations in payoffs. Put differently, expectations-based loss-aversion only has bite when there is unresolved uncertainty, which is only the case for types lying strictly in the intersection of the type spaces.

The fact that the impossibility result is not reversed is linked to the assumption that  $\Lambda_B \leq 1$ , i.e., that gain-loss utility does not dominate. For instance, when types are drawn from  $[0, 1]$  with distributions  $F_S(x) = x$  and  $F_B(x) = x^{10}$  the subsidy in equation (8) turns into a surplus for  $\Lambda_B \geq 13/3$ . However, in this example  $\Lambda_B \leq 1$  is a necessary condition for the materially efficient mechanism to be incentive compatible for the buyer. Hence, incentive compatibility puts limits on the feasible degree of loss-aversion, and, as a consequence, on the strength of the attachment effect, meaning that the impossibility result cannot be reversed. However, as we will discuss next,  $\Lambda_B \leq 1$  is in general only a sufficient condition for incentive compatibility and not always necessary.

KR showed that the assumption  $\Lambda_i \leq 1$  ensures that agents do not choose stochastically dominated options. As we noted when introducing the assumption, it is commonly imposed in the literature, typically for technical reasons. In the present context, it is easy to show that the assumption is a sufficient condition for the materially efficient trade rule to be incentive compatible in the presence of loss-aversion. Moreover, whenever  $F_S(a_B) = 0$  the assumption is not only sufficient, but also necessary. That is, whenever the smallest buyer type has a zero probability of trading, the materially efficient trading rule is CPEIC if and only if  $\Lambda_B \leq 1$ . In particular, this is true when the types of both agents are drawn from the same support. It turns out, however, that when  $F_S(a_B) > 0$  the assumption is no longer necessary.<sup>13</sup> Indeed, when  $F_S(a_B) < 1/2$  the necessary condition reads  $\Lambda_B \leq 1/(1 - 2F_S(a_B))$  and when  $F_S(a_B) \geq 1/2$  no restrictions need to be put on  $\Lambda_B$ . In the light of the above result the question thus arises whether the impossibility result persists when  $F_S(a_B) > 0$  and the assumption is relaxed, as this would allow us to strengthen the attachment effect and possibly set the required subsidy in equation (8) equal to zero.

To this end, one can show that the impossibility result continues to hold for  $\Lambda_B \leq 1/(1 - F_S(a_B))$ . This condition ensures that the lowest buyer type  $a_B$  is in fact the “worst” buyer type.<sup>14</sup> For  $\Lambda_B > 1/(1 - F_S(a_B))$ , the worst buyer type is some intermediate type and the above approach to proving the impossibility result fails: if the lowest buyer type is no longer the worst type, satisfying individual rationality for the lowest buyer type does no longer guarantee satisfying individual rationality for all types. The observation that an intermediate type is the worst type is reminiscent of the related model of partnership dissolution (Cramton, Gibbons, and Klemperer, 1987; Fieseler, Kittsteiner, and Moldovanu, 2003). In this model, the good is initially not exclusively owned by one agent only, but by several agents. As a result, the worst type of an agent may be an intermediate type. However, in spite of this similarity, the approach taken in that model cannot be extended to the present context due to the endogeneity of the reference point. In sum, although counter examples have proved elusive, a reversal of the impossibility for when  $\Lambda_B > 1/(1 - F_S(a_B))$  cannot be ruled out. Note, however, that for sufficiently high degrees of loss-aversion the gains-from trade disappear completely. Thus, even if the buyer’s information rent can be reduced using the attachment effect, impossibility will obtain for sufficiently high degrees of loss-aversion because it will eliminate off the gains from trade.

In the above we have only discussed the degree of loss-aversion of the buyer. Analogous arguments regarding the necessity and sufficiency of  $\Lambda_S \leq 1$  for incentive-compatibility of

---

<sup>13</sup>In Herweg et al. (2010), who first introduced this assumption, the assumption plays a similar role as here. It provides a sufficient but not necessary condition to satisfy incentive compatibility of certain contracts.

<sup>14</sup>Rosato (2014) makes the assumption that gain-loss utility does not dominate precisely to ensure that the lowest type of an agent is the worst type.

the seller apply. However, as loss-aversion on the side of the seller makes the impossibility problem only harder, relaxing the assumption that gain-loss utility does not dominate does not affect our result.

## 5 Optimal Mechanisms

### 5.1 The Revenue Maximization Problem

The preceding section has confirmed the impossibility result in a framework with loss-averse agents under the standard assumption that gain-loss utility does not dominate. In particular, a designer who wants to ensure materially efficient trade while satisfying incentive compatibility and individual rationality cannot make a positive profit. A natural question is thus whether a materially *inefficient* trade mechanism satisfying incentive compatibility and individual rationality can lead to a positive profit for the designer. To answer this question we consider the design of revenue maximizing mechanisms in the presence of loss-averse agents. We will first consider the case of general distributions and prove that the designer insures the agents against ex post variations in their payoffs. More specifically, we show that in the presence of loss-aversion optimal transfers are interim deterministic. We then restrict attention to the case where both the seller and buyer types are distributed uniformly on  $[a, b]$  with  $b = a + 1$ . The preceding, more general analysis of the impossibility result suggests that the symmetry of the type spaces is not a too restrictive assumption, as loss-aversion does not matter for the extreme types for whom trade is interim deterministic. We focus on the uniform distribution for tractability and because it allows us to derive the trade rule explicitly.

The revenue-maximizing designer's problem reads

$$\begin{aligned} & \max_{(y^f, t_S^f, t_B^f) \in \mathcal{F}} \int_{a_B}^{b_B} \int_{a_S}^{b_S} \left( t_B^f(\theta_S, \theta_B) - t_S^f(\theta_S, \theta_B) \right) dF_S(\theta_S) dF_B(\theta_B), \\ & \text{subject to CPEIC and IR.} \end{aligned} \tag{RM}$$

We begin by rewriting this problem into a more accessible form which will allow us to gain some intuition first. The complete derivations and proofs of this section are contained in Appendix B.2. The first step is to impose the envelope representation of the utility due to the CPEIC and the individual rationality constraint. The objective function then reads

$$\int_{a_B}^{b_B} \left( \eta_B^2 w_B(\theta_B) + \theta_B \tilde{v}_B(\theta_B) - \int_{a_B}^{\theta_B} \tilde{v}_B(t) dt \right) dF_B(\theta_B)$$

$$+ \int_{a_S}^{b_S} \left( \eta_S^2 w_S(\theta_S) - \theta_S \tilde{v}_S(\theta_S) - \int_{\theta_S}^{b_B} \tilde{v}_S(t) dt \right) dF_S(\theta_S).$$

In the absence of loss-aversion, the envelope representation of utility would allow us to maximize over the trade rule only instead of both the trade rule and transfers. With loss-aversion in the money dimension, however, this is not the case. Indeed, recall that we defined

$$w_S(\theta_S) = \int_{a_B}^{b_B} \int_{a_B}^{b_B} \mu_S^2 \left( t_S^f(\theta_S, \theta_B) - t_S^f(\theta_S, \theta'_B) \right) dF_B(\theta'_B) dF_B(\theta_B),$$

and thus the objective function still depends on transfers. This expression and its analog for the buyer collect all gain-loss utility with respect to money. Nevertheless, the problem can be reduced to only choosing the optimal trade rule, because in any optimal mechanism the transfers of the seller will not depend on the buyer's type, and vice versa. To see this, note that

$$\begin{aligned} w_S(\theta_S) &= \int_{a_B}^{b_B} \int_{a_B}^{b_B} \mu_S^2 \left( t_S^f(\theta_S, \theta_B) - t_S^f(\theta_S, \theta'_B) \right) dF_B(\theta'_B) dF_B(\theta_B) \\ &= \int_{a_B}^{b_B} \int_{a_B}^{b_B} \left( t_S^f(\theta_S, \theta_B) - t_S^f(\theta_S, \theta'_B) \right) \mathbb{1}[t_S^f(\theta_S, \theta_B) > t_S^f(\theta_S, \theta'_B)] dF_B(\theta'_B) dF_B(\theta_B) \\ &\quad + \int_{a_B}^{b_B} \int_{a_B}^{b_B} \lambda_S^2 \left( t_S^f(\theta_S, \theta_B) - t_S^f(\theta_S, \theta'_B) \right) \mathbb{1}[t_S^f(\theta_S, \theta_B) < t_S^f(\theta_S, \theta'_B)] dF_B(\theta'_B) dF_B(\theta_B) \\ &= \int_{a_B}^{b_B} \int_{a_B}^{b_B} \left( t_S^f(\theta_S, \theta_B) - t_S^f(\theta_S, \theta'_B) \right) \mathbb{1}[t_S^f(\theta_S, \theta_B) > t_S^f(\theta_S, \theta'_B)] dF_B(\theta'_B) dF_B(\theta_B) \\ &\quad - \lambda_S^2 \int_{a_B}^{b_B} \int_{a_B}^{b_B} \left( t_S^f(\theta_S, \theta'_B) - t_S^f(\theta_S, \theta_B) \right) \mathbb{1}[t_S^f(\theta_S, \theta'_B) > t_S^f(\theta_S, \theta_B)] dF_B(\theta'_B) dF_B(\theta_B) \\ &= (1 - \lambda_S^2) \int_{a_B}^{b_B} \int_{a_B}^{b_B} \left( t_S^f(\theta_S, \theta'_B) - t_S^f(\theta_S, \theta_B) \right) \mathbb{1}[t_S^f(\theta_S, \theta'_B) > t_S^f(\theta_S, \theta_B)] dF_B(\theta'_B) dF_B(\theta_B), \end{aligned}$$

where  $\mathbb{1}$  denotes the indicator function. The key step in the above derivation lies in the last equality. Comparing the two integrands on the third and second to last lines, we notice that they look the same but that  $\theta_B$  and  $\theta'_B$  are interchanged. To see the equality, change the order of integration in the integral on the second to last line and perform a change of variables for the resulting integral. This shows that the two integrals are actually the same and allows us to sum them. Thus, since  $\lambda_S^2 > 1$  we find  $w_S(\theta_S) \leq 0$ . As the expression enters the designer's maximization problem positively, she optimally sets  $w_S(\theta_S) = 0$ . Note that a transfer achieves  $w_S(\theta_S) = 0$  if and only if the transfer is independent of almost all buyer types. Thus, interim deterministic transfers are the only transfers that achieve  $w_S(\theta_S) = 0$ . The argument for the transfers of the buyer is analogous.

**Proposition 4** *If the mechanism  $(y^f, t_S^f, t_B^f)$  is a solution to the revenue maximization problem (RM), then the transfer functions  $t_S^f$  and  $t_B^f$  are interim deterministic.*

Intuitively, loss-averse agents dislike ex post variations in their payoffs. By making the transfers independent of the other agent's type, the designer completely insures the agents from any ex post variation in the transfers. Thus, starting from any mechanism with non-interim deterministic transfers, the designer can extract more surplus from the agents by choosing appropriate interim deterministic transfers, effectively selling the agents insurance. Note that interim deterministic transfers are also a solution in the optimal mechanism in MS. However, in the presence of loss-aversion interim deterministic transfers are the *only* solution. Proposition 4 in fact extends beyond the bilateral trade setting to general social choice functions, as we show in Appendix A, thus generalizing the corresponding result on auctions in Eisenhuth (2013).

For the remainder of this section we will assume that the seller and buyer types are distributed uniformly on  $[a, b]$  with  $b = a + 1$  and explicitly derive the optimal trade rule. The assumption allows us to rewrite the maximization problem to

$$\begin{aligned} \max_{y^f \in \mathcal{Y}} & \int_a^b (2\theta_B - 1 - a)y_B(\theta_B) (1 + \Lambda_B [y_B(\theta_B) - 1]) d\theta_B \\ & - \int_a^b (2\theta_S - a)y_S(\theta_S) (1 - \Lambda_S [y_S(\theta_S) - 1]) d\theta_S, \end{aligned} \quad (\text{RM}')$$

subject to  $y_B(\theta_B)$  being non-decreasing and  $y_S(\theta_S)$  being non-increasing,

where  $y_B(\theta_B) = \int_a^b y^f(\theta_S, \theta_B) d\theta_S$  and  $y_S(\theta_S) = \int_a^b y^f(\theta_S, \theta_B) d\theta_B$  denote the interim trade probabilities of the buyer and seller, respectively. Let us inspect the objective function in (RM') more closely. The first integral corresponds to the expected payment the designer receives from the buyer and the second integral to the expected payment the designer makes to the seller. Note that the seller integral is always positive. The buyer integral is positive whenever  $(2\theta_B - 1 - a) \geq 0$ . Clearly, any optimal mechanism will therefore only induce trade for buyer types  $\theta_B \geq (1 + a)/2$ . Given this, both integrals are increasing in the trade probabilities  $y_B$  and  $y_S$ , respectively. Thus, the designer faces the intuitive trade-off that inducing trade comes at cost in the form of the payment due to the seller and with a benefit in the form of the payment from the buyer. Further, the form of the objective function suggests that even in the presence of loss-aversion the designer wants to induce trade between high buyer and low seller types in particular. Put differently, the designer wants to buy the good from a low-value seller and sell it to a high-value buyer, as this yields a large profit margin. However, as a consequence of expectations-based loss-aversion it matters for an agent's utility whether trade takes place with only a few or many types of the other agent, as this affects her expectations, which in turn determine her expected gain-loss utility. Thus, there are in some sense externalities between the outcomes of different types. Indeed, because the agent's expected utility depends on the mechanism through the reference point, pointwise maximization of the

objective function is not possible. In order to nevertheless explicitly derive the optimal trade rule, we make use of the reduced-form approach developed first by Border (1991) and recently generalized by Che et al. (2013). In the case of single-unit auctions, Border (1991) characterized which interim allocation probabilities are implementable by some ex post allocation rule. Che et al. (2013) generalize this to the case of multi-unit auctions when agents may face capacity constraints. In particular, the results in Che et al. (2013) extend to the bilateral trade setting, allowing us to revert to this reduced-form approach. The conditions derived in Che et al. (2013) allow us to maximize directly over the interim trade probabilities  $y_B$  and  $y_S$  instead of the ex post trade rule  $y^f$ . Using the conditions that ensure that these trade probabilities can actually be implemented by some ex post trade rule, we can eliminate the seller's trade probability from the problem and maximize over  $y_B$  only. This allows us to transform the problem into one which can be solved using standard techniques from calculus of variations.

**Proposition 5** *The revenue maximizing trade rule is given by*

$$y^{RM}(\theta_S, \theta_B) = \begin{cases} 1 & \text{if } \theta_S \leq \delta^{RM}(\theta_B), \\ 0 & \text{otherwise.} \end{cases}$$

If  $\Lambda_S \leq (1 - \Lambda_B(a + 1))/a$  and  $\Lambda_B \leq 1/(1 + a)$ , there exists  $\bar{\theta}_B \in [a, a + 1]$  such that  $\delta^{RM}(\theta_B) = a$  for  $\theta_B < \bar{\theta}_B$ , and

$$\delta^{RM}(\theta_B) = \frac{(2\theta_B - 1 - a)(1 - \Lambda_B(2a + 1) + a\Lambda_S) + a - \Lambda_S a^2}{2(1 - \Lambda_B(2\theta_B - a - 1) + \Lambda_S(2\theta_B - 1 - 2a))},$$

for  $\theta_B \geq \bar{\theta}_B$ . If  $\Lambda_S > (1 - \Lambda_B(a + 1))/a$  or  $\Lambda_B > 1/(1 + a)$  we have  $\delta^{RM}(\theta_B) = a$  for all  $\theta_B \in [a, a + 1]$ .

This result requires some discussion as it has several noteworthy features. First, in the absence of loss-aversion in the trade dimension, i.e., for  $\Lambda_S = \Lambda_B = 0$ , we obtain the mechanism from MS in the framework without loss-aversion given by  $\delta(\theta_B) = \theta_B - 1/2$ . Second, the amount of trade taking place is monotonically decreasing in the degree of loss-aversion and for sufficiently high degrees of loss-aversion no trade takes place at all. Third, the trade-reducing effect of buyer loss-aversion is stronger than the one of seller loss-aversion.<sup>15</sup> This may come as a surprise in view of the endowment and attachment effect. In particular, when confirming the impossibility result under loss-aversion, the endowment effect made the problem unambiguously harder, while the attachment effect had the potential to mitigate it, depending on the distribution of types. However, when

---

<sup>15</sup>For any value of  $a$ , as loss-aversion increases, the buyer loss-aversion will always lead to no trade taking place more quickly than seller loss-aversion.

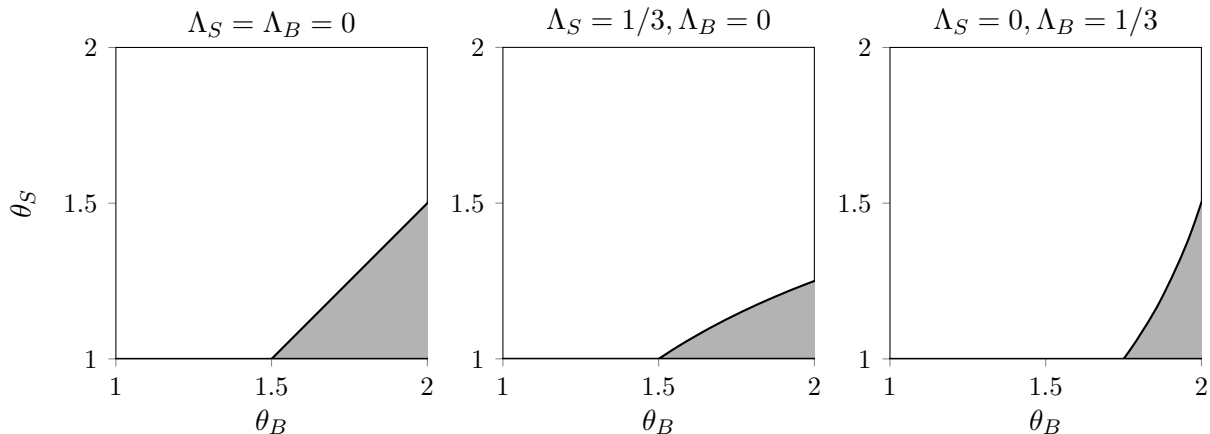


Figure 1: Illustration of the optimal trade rules for  $a = 1$ . The shaded area indicates for which pairs of types trade is taking place.

types are distributed uniformly, the attachment effect does not mitigate the impossibility problem. Moreover, loss-aversion affects the types of buyers and sellers the designer is most interested in differently. Indeed, the attachment and endowment effect are generally stronger for higher types, as these types value the gain-loss utility more strongly than low types. Moreover, as we already noted above, inducing trade increases the payment received from the buyer but it also increases the payment made to the seller. It is for this reason that the designer wants trade to take place in particular with high buyer types and low seller types. Hence, the effect of loss-aversion is more pronounced for the buyer types than the seller types which are attractive from the revenue maximizing designer's point of view. Put differently, the adverse effect of loss-aversion is increasing in the type of the agents. Since the designer cares most about high buyer types and low seller types, buyer loss-aversion has a stronger impact on the trade frequency than seller loss-aversion.

Fourth, and perhaps most interestingly, the optimal mechanism depends on the type space. In the context of loss-aversion, this suggests that the size of the stakes matters. In particular, for high stakes, i.e., high values of  $a$ , less trade takes place for any degree of loss-aversion. This is in sharp contrast to the case without loss-aversion, where the optimal mechanism is independent of the size of the stakes. Intuitively, the potential material gains from trade remain the same even when the stakes are high, because only the difference between valuations matters. However, as the stakes increase, the potential losses increase. Since the designer needs to compensate the agents for these losses with appropriate transfers to maintain individual rationality, the losses eventually eat up all the potential material gains. Hence, at some point the best the designer can do is to induce no trade at all.

Finally, as already noted, by optimally making transfers interim deterministic, the



designer provides the agents with insurance in the money dimension. Similarly, one can interpret the reduction in the trade dimension as partial insurance. Full insurance in this dimension would correspond to trade always or trade never taking place, which in general is not optimal. However, reducing the probability for trade lowers expectations and, as a consequence, there is less room for losses which benefits the agents.

## 5.2 The Welfare Maximization Problem

In this section, we put ourselves in the shoes of a benevolent designer who wants to maximize ex ante welfare by maximizing the sum of ex ante expected utilities. In addition to CPEIC and IR, we impose a budget balance condition. Namely, we do not want the designer to inject money in the economy on average. This is in line with the preceding section, where we looked at ex ante revenue maximization. We say that a mechanism is ex ante budget balanced if

$$\int_{a_S}^{b_S} \int_{a_B}^{b_B} \left( t_S^f(\theta_S, \theta_B) - t_B^f(\theta_S, \theta_B) \right) dF_S(\theta_S) dF_B(\theta_B) = 0. \quad (\text{AB})$$

The designer's problem reads

$$\begin{aligned} \max_{(y^f, t_B^f, t_S^f) \in \mathcal{F}} \int_{a_S}^{b_S} U_S(\theta_S, s_B^t | \theta_S) dF_S(\theta_S) + \int_{a_B}^{b_B} U_B(\theta_B, s_S^t | \theta_B) dF_B(\theta_B), \\ \text{subject to CPEIC, IR and AB.} \end{aligned} \quad (\text{WM})$$

To solve this problem we will proceed as we did in the preceding section. We also obtain the result that in any welfare maximizing mechanisms transfers will be interim deterministic.

**Proposition 6** *If the mechanism  $(y^f, t_S^f, t_B^f)$  is a solution to the welfare maximization problem (WM), then the transfer functions  $t_S^f$  and  $t_B^f$  are interim deterministic.*

The proof is analogous to the revenue maximization problem. In fact, just as in the case of revenue maximizing mechanisms, this result extends beyond the bilateral trade setting and applies to general social choice functions (see Appendix A). To make further progress we again impose that types are uniformly distributed on  $[a, b] = [a, a + 1]$ . However, the presence of the budget constraint makes the problem less tractable, as we need to pin down the Lagrange multiplier. As a consequence, we need to impose symmetric degrees of loss-aversion in the trade dimension, i.e.,  $\Lambda_B = \Lambda_S = \Lambda$ . Imposing incentive compatibility and

budget balance we can obtain the Lagrangian to the problem which reads

$$\begin{aligned} \mathcal{L}(y^f, \gamma) = & \int_a^b (\theta_B + \gamma(2\theta_B - 1 - a))y_B(\theta_B) (1 + \Lambda_B [y_B(\theta_B) - 1]) d\theta_B \\ & - \int_a^b (\theta_S + \gamma(2\theta_S - a))y_S(\theta_S) (1 - \Lambda_S [y_S(\theta_S) - 1]) d\theta_S. \end{aligned}$$

To derive the optimal trade rule we proceed as before for the revenue maximizing mechanism. That is, we make use of the reduced-form implementability conditions in Che et al. (2013) to derive the optimal interim trade probabilities. From there we recover an ex post allocation rule which implements these probabilities and therefore is an optimal trade rule.

**Proposition 7** *The welfare maximizing trade rule is given by*

$$y^{WM}(\theta_S, \theta_B) = \begin{cases} 1 & \text{if } \theta_S \leq \delta^{WM}(\theta_B), \\ 0 & \text{otherwise.} \end{cases}$$

If  $\Lambda < 1/(1+a)$ , there exists  $\bar{\theta}_B \in [a, a+1]$  such that  $\delta^{WM}(\theta_B) = a$  for  $\theta_B < \bar{\theta}_B$ , and

$$\begin{aligned} \delta^{WM}(\theta_B) = & \frac{(2a\Lambda + \Lambda - 1)((2a^2 + 2a + 1)\Lambda^2 - M - (2a + 1)\Lambda)}{2(a\Lambda - 1)(M + a\Lambda^2 - (a + 1)\Lambda + 1)} \\ & + \theta_B \frac{(a\Lambda + \Lambda - 1)(M - a(\Lambda + 1)\Lambda - \Lambda^2 + 1)}{(a\Lambda - 1)(M + a\Lambda^2 - (a + 1)\Lambda + 1)}, \end{aligned}$$

for  $\theta_B \geq \bar{\theta}_B$ , where

$$M = \sqrt{(3a^2 + 3a + 1)\Lambda^4 - (2a + 1)\Lambda^3 + a(a + 1)\Lambda^2 - (2a + 1)\Lambda + 1}.$$

If  $\Lambda \geq 1/(1+a)$  we have  $\delta^{WM}(\theta_B) = a$  for all  $\theta_B \in [a, a+1]$ .

The optimal mechanism once more has some noteworthy features. First, in the absence of loss-aversion we get  $\delta^{WM}(\theta_B) = \theta_B - 1/4$  which is the mechanism from MS in the framework without loss-aversion.

Second, we can compare the condition for trade taking place with the corresponding conditions in the revenue maximizing mechanism. There trade took place if  $\Lambda_S \leq (1 - \Lambda_B(a + 1))/a$  and  $\Lambda_B \leq 1/(1 + a)$ . Clearly, the condition on the buyer loss-aversion is the more restrictive one. When either of these conditions are violated, any mechanism which induces trade yields a negative expected revenue. Hence, for  $\Lambda \geq 1/(1 + a)$  any mechanism which induces trade violates the budget balance constraint. Consequently, no trade is the only feasible welfare maximizing mechanism. Moreover, just as in the case of revenue maximization, the size of the stakes matter.

Finally, as in the revenue maximization problem, optimal transfers are interim deterministic. Thus, the designer provides the agents with insurance in the money dimension. As already noted, this result is not specific to the bilateral trade setting, but applies to any welfare maximizing mechanism.

## 6 Alternative reference-point formation

The model by KR used in this paper has arguably become the workhorse model in the context of reference-dependent utility. A particularly appealing feature of the model is the endogenously determined reference point using the agent's rational expectations. As noted earlier (see footnote 3), a number of studies provide evidence for the assumption that a person's reference point is determined by her expectations. However, there are different ways one can model this. KR note that the equilibrium concepts in the models on disappointment-aversion by Bell (1985) and Loomes and Sugden (1986) are closely related to the CPE. The CPE specifies the reference point as the full distribution of a lottery, whereas the reference point corresponds to the certainty equivalent of the lottery in these models of disappointment-aversion. However, Masatlioglu and Raymond (forthcoming) find that the intersection of preferences induced by the CPE and any of these disappointment-aversion models is only expected utility. Thus, although the models seem to be very similar, the induced preferences do generally not coincide. Nevertheless, the impossibility result in Section 4 remains valid and the optimal mechanisms derived in Section 5 coincide if we specify the reference point as the certainty equivalent of the lottery as in Bell (1985) and Loomes and Sugden (1986). Hence, the optimal mechanisms we derived earlier exhibits robustness to the specific formation of the reference-point.<sup>16</sup> To keep the analysis concise, we focus on the seller only. The arguments are essentially the same for the buyer. Under the alternative specification of the reference point the utility of the seller reads

$$\begin{aligned}
U_S(\theta_S, s_B^t | \theta_S) &= \int_{a_B}^{b_B} \left( -y^f(\theta_S, \theta_B)\theta_S + t_S^f(\theta_S, \theta_B) \right) dF_B(\theta_B) \\
&+ \int_{a_B}^{b_B} \eta_S^1 \mu_S^1 \left( \mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)]\theta_S - y^f(\theta_S, \theta_B)\theta_S \right) dF_B(\theta_B) \\
&+ \int_{a_B}^{b_B} \eta_S^2 \mu_S^2 \left( t_S^f(\theta_S, \theta_B) - \mathbb{E}_B[t_S^f(\theta_S, \tilde{\theta}_B)] \right) dF_B(\theta_B).
\end{aligned}$$

Comparing this alternative expression to the expected utility we worked with (see equation (3)), we notice that the material utility on the first line remains unchanged, while the

---

<sup>16</sup>Copic and Ponsatí (2008) have studied the bilateral trade problem in the context of robust mechanism design in the vein of Bergemann and Morris (2005). The robustness we have in mind here is closer to the behaviorally robust mechanisms in Bierbrauer and Netzer (2016).

gain-loss utility in the second line takes a new form. Indeed, instead of comparing the induced outcome to every single potential outcome in the reference lottery, the agent now compares the outcome only to the certainty equivalent of the reference lottery, which enters the value function directly. Two observations about the alternative gain-loss utility yield the robustness result. Consider the money dimension first and recall that  $\mu_S^2$  is a concave function. Thus, by Jensen's inequality we get

$$\begin{aligned} & \int_{a_B}^{b_B} \eta_S^2 \mu_S^2 \left( t_S^f(\theta_S, \theta_B) - \mathbb{E}_B[t_S^f(\theta_S, \tilde{\theta}_B)] \right) dF_B(\theta_B) \\ & \leq \eta_S^2 \mu_S^2 \left( \int_{a_B}^{b_B} \left( t_S^f(\theta_S, \theta_B) - \mathbb{E}_B[t_S^f(\theta_S, \tilde{\theta}_B)] \right) dF_B(\theta_B) \right) = 0, \end{aligned}$$

as  $\int_{a_B}^{b_B} t_S^f(\theta_S, \theta_B) dF_B(\theta_B) = \mathbb{E}_B[t_S^f(\theta_S, \tilde{\theta}_B)]$  by definition. Therefore, the result in Lemma 3 in Appendix A that  $w_S(\theta_S) \leq 0$  carries through to this specification. Hence, irrespective of which of the two specifications of the reference point we use, interim deterministic transfers are optimal.

Consider the trade dimension next and notice that  $\mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)] \in [0, 1]$  while  $y^f(\theta_S, \theta_B) \in \{0, 1\}$ . Thus, the binary nature of trade implies that an agent feels only either gains or losses in the trade dimension, irrespective of the reference lottery and outcome. We can thus rewrite

$$\begin{aligned} & \int_{a_B}^{b_B} \eta_S^1 \mu_S^1 \left( \mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)] \theta_S - y^f(\theta_S, \theta_B) \theta_S \right) dF_B(\theta_B) \\ & = \theta_S \eta_S^1 \int_{a_B}^{b_B} \left( \lambda_S^1 y^f(\theta_S, \theta_B) \left( \mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)] - 1 \right) + (1 - y^f(\theta_S, \theta_B)) \mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)] \right) dF_B(\theta_B) \\ & = \theta_S \eta_S^1 \int_{a_B}^{b_B} \int_{a_B}^{b_B} \left( \lambda_S^1 y^f(\theta_S, \theta_B) (y^f(\theta_S, \theta'_B) - 1) + (1 - y^f(\theta_S, \theta_B)) y^f(\theta_S, \theta'_B) \right) dF_B(\theta'_B) dF_B(\theta_B) \\ & = \theta_S \eta_S^1 \int_{a_B}^{b_B} \int_{a_B}^{b_B} \mu_S^1 (y^f(\theta_S, \theta'_B) - y^f(\theta_S, \theta_B)) dF_B(\theta'_B) dF_B(\theta_B), \end{aligned}$$

where the final line is the very expression of gain-loss utility in the trade dimension under the specification used throughout the paper. Thus, regarding gain-loss utility in the trade dimension the two different specifications of the reference point are equivalent.<sup>17</sup> Consequently, all of our results continue to hold under the alternative specification of the reference point, as the two specifications are equivalent conditional on interim deterministic transfers.

While the two formulations disagree on the precise way the reference-point is formed, they agree that it is the agents' expectations which determine the reference-point endogenously. Alternatively, one could consider a model in which the reference-point is exogenously given and not determined by the agent's expectations. We briefly explore

---

<sup>17</sup>Notice that this finding does not hinge on the piece-wise linearity of  $\mu_i^1$ , but is solely due to the binary nature of trade.

this direction using the model of loss-aversion used in Spiegler (2012) and reconsider the impossibility result in this framework. In the model by Spiegler (2012) agents have an exogenously given reference point  $r_i$  and feel losses in case of negative deviations, but they feel no gains in case of positive deviations. Thus, a buyer feels a loss of  $\lambda_B r_B \theta_B$  when no trade happens, while the seller feels a loss of  $\lambda_S(1 - r_S)\theta_S$  when trade does happen. Similarly to the model by KR, loss-aversion in the money dimension will only make the impossibility problem harder, as it decreases gains from trade without affecting information rents. We can write agents' expected utility as

$$U_B(\theta_B, r_B) = \theta_B y_B(\theta_B) - \bar{t}_B(\theta_B) - (1 - y_B(\theta_B))\lambda_B r_B \theta_B$$

and

$$U_S(\theta_S, r_S) = -\theta_S y_S(\theta_S) + \bar{t}_S(\theta_S) - y_S(\theta_S)\lambda_S(1 - r_S)\theta_S.$$

Collecting terms we observe that, as in the analysis in Section 4, seller-loss aversion makes the problem unambiguously harder while the effect is ambiguous in case of the buyer. Hence, the endowment and attachment effect are once more at work. One can then follow essentially the same steps as we did for the proof of Proposition 3 to obtain that an incentive compatible, materially efficient, and budget balanced mechanism implies

$$\begin{aligned} U_B(a_B) + U_S(b_S) = & \int \int \left( (1 + \lambda_B r_B) \left( \theta_B - \frac{1 - F_B(\theta_B)}{f_B(\theta_B)} \right) - (1 + \lambda_S(1 - r_S)) \left( \theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)} \right) \right) y(\theta_S, \theta_B) dF_B(\theta_B) dF_S(\theta_S) \\ & - \lambda_B r_B \int \left( \theta_B - \frac{1 - F_B(\theta_B)}{f_B(\theta_B)} \right) dF_B(\theta_B). \end{aligned}$$

Thus, making use of the result in MS, one can see that a sufficient condition for the impossibility result to persist is given by  $\lambda_B r_B \leq \lambda_S(1 - r_S)$ . Whether the impossibility result extends in full generality, is not clear however.

## 7 Conclusion

There are countless papers on mechanism design and vast evidence of the prevalence of loss-aversion in people's behavior. Yet, as highlighted in a recent survey by Kőszegi (2014), work combining these two highly relevant fields is scarce. The present paper contributes to this literature by investigating optimal mechanisms in a bilateral trade setting with loss-averse agents.

We address three problems in the bilateral trade context. First, the traditionally important question of inducing materially efficient trade; second, the economically relevant issue of revenue maximization; third, the socially important design of welfare maximizing

institutions. We find that the presence of loss-aversion generally makes all three problems harder, as a higher subsidy is required to induce materially efficient trade, and maximal revenue and welfare are reduced. The endowment and attachment effects, which are well-documented empirically, are apparent in our results and provide an intuitive explanation. The common theme in all three problems is that of insurance. In both, welfare and revenue maximizing mechanisms, interim deterministic transfers are optimal, providing agents with full insurance in the money dimension. Additionally, less trade takes place in the presence of loss-aversion, which can be interpreted as partial insurance in the trade dimension. Further, loss-aversion affects the optimal mechanisms in a surprising and yet intuitive fashion. First, while both buyer and seller loss-aversion reduce the optimal amount of trade, buyer loss-aversion has a more pronounced impact, because loss-aversion affects high types more strongly than low types, and the designer is particularly interested in high buyer types and low seller types. Second, the size of the stakes matter for the optimal mechanism: when the stakes are high, the designer optimally induces less trade, because the agents need to be compensated for risking large losses.

Interestingly and somewhat surprisingly, all of these findings display robustness to the exact specification of the endogenous reference point. This is of practical relevance, as the designer of some economic institution may have evidence that individuals are loss-averse, but be unsure about the precise formation process of the reference point. The robustness result suggests that lacking this information may not be too much of a problem, as long as loss-averse individuals are provided with insurance.

## A A General Mechanism Design Approach

In this section we briefly consider general mechanisms, that is, we do not limit ourselves to the bilateral trade problem. This allows us to generalize the result in Eisenhuth (2013) that the optimal auction is an all-pay auction to the result that any revenue maximizing mechanism features interim deterministic transfer and further extend this to welfare maximizing mechanisms.

An environment  $E = [I, X, (\Theta_i, v_i)_{i \in I}, F_i]$  is characterized by the following components. There is a finite set of  $N$  agents denoted by  $I = \{1, \dots, N\}$ . The set of social alternatives is given by  $X = Y \times T$  with typical element  $\mathbf{x} = (y, t_1, t_2, \dots, t_N)$ . The (general) set  $Y$  is the set of projects and the set  $T \subseteq \mathbb{R}^N$  is the set of transfers. We consider an independent private values setting. Hence, the type of agent  $i$  is private information and is independently drawn from a distribution  $F_i$  with bounded support  $\Theta_i = [a_i, b_i] \subset \mathbb{R}_+$ . Throughout, we use the conventional notation  $\Theta = \prod_{i=1}^N \Theta_i$ , with typical element  $\theta$ , and  $\Theta_{-i} = \prod_{j \neq i} \Theta_j$ , with typical element  $\theta_{-i}$ . The agents and the principal have identical prior beliefs. Following KR, agents' riskless total utility is additively separable in material utility and in gain-loss utility, and is defined as

$$u_i(\mathbf{x}, \mathbf{r}_i, \theta_i) = \theta_i v_i(y) + t_i + \eta_i^1 \mu_i^1 (\theta_i v_i(y) - \theta_i v_i(r_i^1)) + \eta_i^2 \mu_i^2 (t_i - r_i^2), \quad (9)$$

with some  $\eta_i^1, \eta_i^2 \geq 0$ , and where

$$\mu_i^j(s) = \begin{cases} s & s \geq 0, \\ \lambda_i^j s & s < 0, \end{cases}$$

for  $j = 1, 2$  is a value function in the sense of Kahneman and Tversky (1979), with  $\lambda_i^j > 1$ , thereby capturing loss-aversion. The parameters  $\mathbf{r}_i = (r_i^1, r_i^2)$  are the riskless reference levels. Following KR we allow for the reference point to be stochastic, i.e., to be a reference lottery over all riskless reference levels. More specifically, the reference point is equal to the agent's rational expectations.

A social choice function (SCF)  $f : \Theta \rightarrow X$  assigns a collective choice  $f(\theta_1, \dots, \theta_N) \in X$  to each possible profile of the agents' types  $(\theta_1, \dots, \theta_N) \in \Theta$ . We denote the set of all SCFs  $\mathcal{F}$ . A mechanism  $\Gamma = (M_1, \dots, M_N, g)$  is a collection of  $N$  message sets  $(M_1, \dots, M_N)$  and an outcome function  $g : M_1 \times \dots \times M_N \rightarrow X$ . We denote the direct mechanism by  $\Gamma^d = (\Theta_1, \dots, \Theta_N, f)$ . Since agents privately observe their types, they can condition their message on their type. Consequently, a pure strategy for agent  $i$  in a mechanism  $\Gamma$  is a function  $s_i : \Theta_i \rightarrow M_i$ . Note that  $g(s_1(\theta_1), \dots, s_N(\theta_N)) = x \in X$ . Let  $S_i$  denote the set of all pure strategies of agent  $i$ . Further, we denote the truthful strategy  $s_i^t(\theta_i) = \theta_i$ .

Our equilibrium concept is based on the choice-acclimating personal equilibrium (CPE) (Kőszegi and Rabin, 2007). We thus allow for the reference point to be a distribution over the set  $X$ . In a mechanism  $\Gamma$ , this distribution is induced endogenously for each agent: conditional on the other agents playing  $s_{-i}$ , agent  $i$  induces a distribution over the set of social alternatives,  $X$ , by playing the strategy  $s_i$ . Hence, the loss-averse agent will compare any given social alternative to all possible social alternatives, allowing for gain or loss feelings in every comparison. Moving to the interim stage and allowing for a reference lottery, we can define the interim expected utility of agent  $i$  with type  $\theta_i$ , in the mechanism  $\Gamma$ , when reporting  $m_i$ , given that the other agents play  $s_{-i}$  as

$$\begin{aligned} U_i(m_i, s_{-i}, \Gamma | \theta_i) &= \theta_i \int_{\Theta_{-i}} v_i(y^g(m_i, \theta_{-i})) dF_{-i}(\theta_{-i}) + \int_{\Theta_{-i}} t_i^g(m_i, \theta_{-i}) dF_{-i}(\theta_{-i}) \\ &+ \theta_i \eta_i^1 \int_{\Theta_{-i}} \int_{\Theta_{-i}} \mu_i^1 (v_i(y^g(m_i, \theta_{-i})) - v_i(y^g(m_i, \theta'_{-i}))) dF_{-i}(\theta'_{-i}) dF_{-i}(\theta_{-i}) \\ &+ \eta_i^2 \int_{\Theta_{-i}} \int_{\Theta_{-i}} \mu_i^2 (t_i^g(m_i, \theta_{-i}) - t_i^g(m_i, \theta'_{-i})) dF_{-i}(\theta'_{-i}) dF_{-i}(\theta_{-i}). \end{aligned}$$

We can now define our equilibrium concept, which follows Eisenhuth (2013).

**Definition 4** A strategy profile  $s^* = (s_1^*, \dots, s_N^*)$  is an interim CPE of the mechanism  $\Gamma = (M_1, \dots, M_N, g)$  if for all  $i \in I$  and  $\theta_i \in \Theta_i$ ,

$$s_i^*(\theta_i) \in \arg \max_{m_i \in M_i} U_i(m_i, s_{-i}^*, \Gamma | \theta_i).$$

**Definition 5** A mechanism  $\Gamma$  implements a social choice function  $f$  in CPE if there is a CPE strategy profile,  $s = (s_1, \dots, s_N)$  of  $\Gamma$ , such that

$$g(s_1(\theta_1), \dots, s_N(\theta_N)) = f(\theta_1, \dots, \theta_N)$$

for all  $(\theta_1, \dots, \theta_N) \in \Theta$ .

**Definition 6** A social choice function  $f$  is CPEIC if the truthful profile  $s^t = (s_1^t, \dots, s_N^t)$  is a CPE strategy in the direct mechanism  $\Gamma^d$ .

With these definitions in hand we can now prove the revelation principle for CPE.

**Proposition 8 (Revelation Principle for CPE)** A social choice function  $f$  can be implemented in CPE by some mechanism  $\Gamma$  if and only if  $f$  is CPEIC.

**Proof.** Suppose  $f$  was CPEIC. Then, by definition the strategy profile  $s^t$  a CPE in the direct mechanism  $\Gamma^d$  and thus, again by definition, the direct mechanism implements  $f$  in CPE. Conversely, suppose there is a mechanism  $\Gamma = (M_1, \dots, M_N, g)$  that implements



$f$  in CPE. If  $s^* = (s_1^*, \dots, s_N^*)$  is a CPE, then for all  $i, m'_i \in M_i$  and  $\theta_i$

$$U_i(s_i^*(\theta_i), s_{-i}^*, \Gamma|\theta_i) \geq U_i(m'_i, s_{-i}^*, \Gamma|\theta_i)$$

by definition of the CPE. In particular, this is also true for  $m'_i = s_i^*(\hat{\theta}_i)$  for all  $i \in I, \hat{\theta}_i \in \Theta_i$ . Therefore, given that  $s^* = (s_1^*, \dots, s_N^*)$  is a CPE we have for all  $i \in I, \theta_i, \hat{\theta}_i \in \Theta_i$ ,

$$U_i(s_i^*(\theta_i), s_{-i}^*, \Gamma|\theta_i) \geq U_i(s_i^*(\hat{\theta}_i), s_{-i}^*, \Gamma|\theta_i)$$

Since  $\Gamma$  implements  $f$  in CPE we have

$$g(s_1^*(\theta_1), \dots, s_N^*(\theta_N)) = f(\theta_1, \dots, \theta_N),$$

implying

$$U_i(s_i^t(\theta_i), s_{-i}^t, \Gamma^d|\theta_i) \geq U_i(s_i^t(\hat{\theta}_i), s_{-i}^t, \Gamma^d|\theta_i)$$

for all  $i \in I, \theta_i, \hat{\theta}_i \in \Theta_i$ . Thus, the truthful strategy profile  $s^t$  is a CPE in the direct mechanism and therefore the social choice function  $f$  is CPEIC. ■

Henceforth, we restrict attention to direct mechanisms and no longer explicitly lost the mechanism as an argument in the utility function. Proceeding as in the main text we can write  $U_i(m_i, s_{-i}^t|\theta_i) = \theta_i \tilde{v}_i(m_i) + \bar{t}_i(m_i)$ . It will turn out to be useful to further define

$$\begin{aligned} \bar{t}_i(m_i) &= \int_{\Theta_{-i}} t_i^f(m_i, \theta_{-i}) dF_{-i}(\theta_{-i}), \\ w_i(m_i) &= \int_{\Theta_{-i}} \int_{\Theta_{-i}} \mu_i^2 \left( t_i^f(m_i, \theta_{-i}) - t_i^f(m_i, \theta'_{-i}) \right) dF_{-i}(\theta'_{-i}) dF_{-i}(\theta_{-i}), \end{aligned}$$

which allows us to write  $\tilde{t}_i(m_i) = \bar{t}_i(m_i) + \eta_i^2 w_i(m_i)$ . With this in hand we get the following condition for a social choice function  $f$  to be CPEIC:

$$U_i(\theta_i, s_{-i}^t|\theta_i) \geq U_i(\hat{\theta}_i, s_{-i}^t|\theta_i) \quad \forall i \in I, \forall \hat{\theta}_i \in \Theta_i. \quad (\text{CPEIC})$$

We are now in a position to characterize the set of all CPEIC social choice functions.

**Proposition 9** *The social choice function  $f = (y^f, t_1^f, \dots, t_N^f)$  is CPEIC if and only if, for all  $i \in I$ ,*

(i)  $\tilde{v}_i$  is non-decreasing, and

(ii)  $U_i(\theta_i, s_{-i}^t|\theta_i) = U_i(a_i, s_{-i}^t|a_i) + \int_{a_i}^{\theta_i} \tilde{v}_i(s) ds$  for all  $\theta_i \in \Theta_i$ .

**Proof.** Suppose the social choice function  $f$  is CPEIC. Take some  $\hat{\theta}_i > \theta_i$ , then by CPEIC

$$U_i(\theta_i, s_{-i}^t | \theta_i) \geq \theta_i \tilde{v}_i(\hat{\theta}_i) + \tilde{t}_i(\hat{\theta}_i) = U_i(\hat{\theta}_i, s_{-i}^t | \hat{\theta}_i) + (\theta_i - \hat{\theta}_i) \tilde{v}_i(\hat{\theta}_i)$$

and analogously

$$U_i(\hat{\theta}_i, s_{-i}^t | \hat{\theta}_i) \geq \hat{\theta}_i \tilde{v}_i(\theta_i) + \tilde{t}_i(\theta_i) = U_i(\theta_i, s_{-i}^t | \theta_i) + (\hat{\theta}_i - \theta_i) \tilde{v}_i(\theta_i).$$

Thus,

$$\tilde{v}_i(\hat{\theta}_i) \geq \frac{U_i(\hat{\theta}_i, s_{-i}^t | \hat{\theta}_i) - U_i(\theta_i, s_{-i}^t | \theta_i)}{\hat{\theta}_i - \theta_i} \geq \tilde{v}_i(\theta_i),$$

implying that  $\tilde{v}_i$  is non-decreasing because we assumed  $\hat{\theta}_i > \theta_i$ . Now, letting  $\hat{\theta}_i \rightarrow \theta_i$  we get that for all  $\theta_i$  we have

$$\frac{\partial U_i(\theta_i, s_{-i}^t | \theta_i)}{\partial \theta_i} = \tilde{v}_i(\theta_i)$$

and so

$$U_i(\theta_i, s_{-i}^t | \theta_i) = U_i(a_i, s_{-i}^t | a_i) + \int_{a_i}^{\theta_i} \tilde{v}_i(s) ds$$

for all  $\theta_i \in \Theta_i$ . Conversely, suppose that conditions (i) and (ii) hold. Without loss of generality, take any  $\theta_i > \hat{\theta}_i$ . Then,

$$\begin{aligned} U_i(\theta_i, s_{-i}^t | \theta_i) - U_i(\hat{\theta}_i, s_{-i}^t | \hat{\theta}_i) &= \int_{\hat{\theta}_i}^{\theta_i} \tilde{v}_i(s) ds \\ &\geq \int_{\hat{\theta}_i}^{\theta_i} \tilde{v}_i(\hat{\theta}_i) ds \\ &= (\theta_i - \hat{\theta}_i) \tilde{v}_i(\hat{\theta}_i). \end{aligned}$$

Hence,

$$U_i(\theta_i, s_{-i}^t | \theta_i) \geq U_i(\hat{\theta}_i, s_{-i}^t | \hat{\theta}_i) + (\theta_i - \hat{\theta}_i) \tilde{v}_i(\hat{\theta}_i) = \theta_i \tilde{v}_i(\hat{\theta}_i) + \tilde{t}_i(\hat{\theta}_i)$$

and similarly

$$U_i(\hat{\theta}_i, s_{-i}^t | \hat{\theta}_i) \geq U_i(\theta_i, s_{-i}^t | \theta_i) + (\hat{\theta}_i - \theta_i) \tilde{v}_i(\theta_i) = \hat{\theta}_i \tilde{v}_i(\theta_i) + \tilde{t}_i(\theta_i).$$

Consequently,  $f$  is CPEIC. ■

We are now in a position to prove that interim deterministic transfers are part of any

revenue or welfare maximizing mechanism.

**Proposition 10** *Deterministic transfers are part of a solution to the problem*

$$\min_{(y^f, t_1^f, \dots, t_N^f) \in \mathcal{F}} \sum_{i=1}^N \int_{a_i}^{b_i} \bar{t}_i(\theta_i) dF_i(\theta_i),$$

subject to CPEIC and IR.

We first prove a lemma which we will use repeatedly.

**Lemma 3** *We have  $w_i(\theta_i) \leq 0$  for all  $i$  and  $\theta_i \in \Theta_i$ .*

**Proof.** Recall that we defined

$$w_i(\theta_i) = \int_{\Theta_{-i}} \int_{\Theta_{-i}} \mu_i^2 \left( t_i^f(\theta_i, \theta_{-i}) - t_i^f(\theta_i, \theta'_{-i}) \right) dF_{-i}(\theta'_{-i}) dF_{-i}(\theta_{-i}).$$

We can rewrite these expressions as follows

$$\begin{aligned} w_i(\theta_i) &= \int_{\Theta_{-i}} \int_{\Theta_{-i}} \mu_i^2 \left( t_i^f(\theta_i, \theta_{-i}) - t_i^f(\theta_i, \theta'_{-i}) \right) dF_{-i}(\theta'_{-i}) dF_{-i}(\theta_{-i}) \\ &= \int_{\Theta_{-i}} \int_{\Theta_{-i}} \left( t_i^f(\theta_i, \theta_{-i}) - t_i^f(\theta_i, \theta'_{-i}) \right) \mathbb{1}[t_i^f(\theta_i, \theta_{-i}) - t_i^f(\theta_i, \theta'_{-i}) > 0] dF_{-i}(\theta'_{-i}) dF_{-i}(\theta_{-i}) \\ &\quad + \int_{\Theta_{-i}} \int_{\Theta_{-i}} \lambda_i^2 \left( t_i^f(\theta_i, \theta_{-i}) - t_i^f(\theta_i, \theta'_{-i}) \right) \mathbb{1}[t_i^f(\theta_i, \theta_{-i}) - t_i^f(\theta_i, \theta'_{-i}) < 0] dF_{-i}(\theta'_{-i}) dF_{-i}(\theta_{-i}) \\ &= \int_{\Theta_{-i}} \int_{\Theta_{-i}} \left( t_i^f(\theta_i, \theta_{-i}) - t_i^f(\theta_i, \theta'_{-i}) \right) \mathbb{1}[t_i^f(\theta_i, \theta_{-i}) - t_i^f(\theta_i, \theta'_{-i}) > 0] dF_{-i}(\theta'_{-i}) dF_{-i}(\theta_{-i}) \\ &\quad - \lambda_i^2 \int_{\Theta_{-i}} \int_{\Theta_{-i}} \left( t_i^f(\theta_i, \theta'_{-i}) - t_i^f(\theta_i, \theta_{-i}) \right) \mathbb{1}[t_i^f(\theta_i, \theta'_{-i}) - t_i^f(\theta_i, \theta_{-i}) > 0] dF_{-i}(\theta'_{-i}) dF_{-i}(\theta_{-i}) \\ &= (1 - \lambda_i^2) \int_{\Theta_{-i}} \int_{\Theta_{-i}} \left( t_i^f(\theta_i, \theta'_{-i}) - t_i^f(\theta_i, \theta_{-i}) \right) \mathbb{1}[t_i^f(\theta_i, \theta'_{-i}) - t_i^f(\theta_i, \theta_{-i}) > 0] dF_{-i}(\theta'_{-i}) dF_{-i}(\theta_{-i}), \end{aligned}$$

where  $\mathbb{1}$  denotes the indicator function. Thus, since  $\lambda_i^2 > 1$  we find  $w_i(\theta_i) \leq 0$ . ■

Note that any transfers achieve  $w_i(\theta_i) = 0$  if and only if the transfer does not depend on almost all types of the other agents, i.e., for interim deterministic transfers.

**Proof of Proposition 10.** We begin by simplifying the problem. In order for the CPEIC constraint to be satisfied, conditions (i) and (ii) from Proposition 9 must be satisfied. Using the utility functions from condition (ii), we can rewrite the minimization problem to

$$\min_{(y^f, t_1^f, \dots, t_N^f) \in \mathcal{F}} \sum_{i=1}^N \int_{a_i}^{b_i} \left( \bar{t}_i(a_i) + \eta_i^2 w_i(a_i) - \theta_i \tilde{v}_i(\theta_i) - \eta_i^2 w_i(\theta_i) + \int_{a_i}^{\theta_i} \tilde{v}_i(s) ds \right) dF_i(\theta_i),$$

subject to  $\tilde{v}_i$  being non-decreasing for all  $i \in I$  and IR.

By Lemma 3 we have  $w_i(\theta_i) \leq 0$  for all  $i$  and  $\theta_i \in \Theta_i$ . Note that these terms enter the problem negatively. Since we want to minimize the objective function, we optimally choose transfers such that  $w_i(\theta_i) = 0$  for all  $\theta_i \in \Theta_i$  to minimize the integrands pointwise and therefore minimize the integrals. Doing so does not contradict the IR constraint, on the contrary, it relaxes it. Thus, choosing interim deterministic transfers is optimal and part of a solution to the problem. ■

**Proposition 11** *Deterministic transfers are part of a solution to the problem*

$$\min_{(y^f, t_1^f, \dots, t_N^f) \in \mathcal{F}} \sum_{i=1}^N \int_{a_i}^{b_i} U_i(\theta_i, s_{-i}^t | \theta_i) dF_i(\theta_i),$$

*subject to CPEIC, IR and AB.*

**Proof.** In order for the CPEIC constraint to be satisfied, conditions (i) and (ii) from Proposition 9 must be satisfied. Using the utility functions from condition (ii), we can rewrite the objective function in the problem to

$$\sum_{i=1}^N \left( U_i(a_i, s_{-i}^t | a_i) + \int_{a_i}^{b_i} \int_{a_i}^{\theta_i} \tilde{v}_i(s) ds dF_i(\theta_i) \right). \quad (10)$$

We still have condition (i) from Proposition 9, as well as the IR and AB to keep as constraints. Recall that we can write utility as

$$U_i(\theta_i, s_{-i}^t | \theta_i) = \theta_i \tilde{v}_i(\theta_i) + \bar{t}_i(\theta_i) + \eta_i^2 w_i(\theta_i),$$

and, further, using the same notation, we can write the AB constraint as

$$\sum_{i=1}^N \int_{a_i}^{b_i} \bar{t}_i(\theta_i) dF_i(\theta_i) = 0.$$

Thus, given the CPEIC constraint (condition (ii) in particular) we can write the AB constraint as

$$\sum_{i=1}^N \int_{a_i}^{b_i} \left( \eta_i^2 w_i(\theta_i) + \theta_i \tilde{v}_i(\theta_i) - U_i(a_i, s_{-i}^t | a_i) - \int_{a_i}^{\theta_i} \tilde{v}_i(t) dt \right) dF_i(\theta_i) = 0. \quad (11)$$

Using the rewritten objective function in (10) and using the form of the AB constraint in (11), we can set up a Lagrangian:

$$\mathcal{L}(y^f, t_1^f, \dots, t_N^f, \gamma) = \sum_{i=1}^N (1 - \gamma) U_i(a_i, s_{-i}^t | a_i) + \sum_{i=1}^N (1 - \gamma) \int_{a_i}^{b_i} \int_{a_i}^{\theta_i} \tilde{v}_i(s) ds dF_i(\theta_i)$$

$$+ \gamma \sum_{i=1}^N \int_{a_i}^{b_i} \eta_i^2 w_i(\theta_i) dF_i(\theta_i) + \gamma \sum_{i=1}^N \int_{a_i}^{b_i} \theta_i \tilde{v}_i(\theta_i) dF_i(\theta_i),$$

where  $\gamma$  is the Lagrange multiplier. By Lemma 3 we have  $w_i(\theta_i) \leq 0$  for  $i \in I$ , which enter the Lagrangian positively. In order to maximize the Lagrangian, we can choose interim deterministic transfers which result in  $w_i(\theta_i) = 0$  for  $i \in I$ . This is in line with the remaining constraints given by condition (i) from Proposition 9 and the IR constraint. ■

## B Proofs

### B.1 Impossibility Result

We begin by noting that

$$\begin{aligned} & \tilde{v}_B(\theta_B) \\ &= \int_{a_S}^{b_S} y^f(\theta_S, \theta_B) dF_S(\theta_S) + \eta_B^1 \int_{a_S}^{b_S} \int_{a_S}^{b_S} \mu_B^1 (y^f(\theta_S, \theta_B) - y^f(\theta'_S, \theta_B)) dF_S(\theta'_S) dF_S(\theta_S), \\ &= y_B(\theta_B) + \eta_B^1 \int_{a_S}^{b_S} \int_{a_S}^{b_S} y^f(\theta_S, \theta_B) (1 - y^f(\theta'_S, \theta_B)) - \lambda_B^1 (1 - y^f(\theta_S, \theta_B)) y^f(\theta'_S, \theta_B) dF_S(\theta'_S) dF_S(\theta_S), \\ &= y_B(\theta_B) (1 + \Lambda_B (y_B(\theta_B) - 1)) \end{aligned}$$

and analogously  $\tilde{v}_S(\theta_S) = y_S(\theta_S) (1 - \Lambda_S (y_S(\theta_S) - 1))$ , where

$$y_B(\theta_B) = \int_{a_S}^{b_S} y^f(\theta_S, \theta_B) dF_S(\theta_S), \quad y_S(\theta_S) = \int_{a_B}^{b_B} y^f(\theta_S, \theta_B) dF_B(\theta_B).$$

Imposing CPEIC we can write the sum of the agents' ex ante expected utilities as

$$\begin{aligned} & \int_{a_B}^{b_B} U_B(\theta_B) f_B(\theta_B) d\theta_B + \int_{a_S}^{b_S} U_S(\theta_S) f_S(\theta_S) d\theta_S \\ &= U_B(a_B) + \int_{a_B}^{b_B} \int_{a_B}^{\theta_B} y_B(t) (1 + \Lambda_B (y_B(t) - 1)) dt f_B(\theta_B) d\theta_B \\ &+ U_S(b_S) + \int_{a_S}^{b_S} \int_{\theta_S}^{b_S} y_S(t) (1 - \Lambda_S (y_S(t) - 1)) dt f_S(\theta_S) d\theta_S \\ &= U_B(a_B) + \int_{a_B}^{b_B} y_B(\theta_B) (1 + \Lambda_B (y_B(\theta_B) - 1)) (1 - F_B(\theta_B)) d\theta_B \\ &+ U_S(b_S) + \int_{a_S}^{b_S} y_S(\theta_S) (1 - \Lambda_S (y_S(\theta_S) - 1)) F_S(\theta_S) d\theta_S. \end{aligned}$$

Note that the monotonicity constraints are satisfied due to Assumption 1, i.e.,  $\Lambda_B, \Lambda_S \leq 1$ . Further, from Lemmas 1 and 2 and the corresponding discussion in the main text we know that we can set the loss-aversion in the money dimension to zero. This allows us to express

the sum of the agents' ex ante expected utilities as

$$\begin{aligned}
& \int_{a_B}^{b_B} U_B(\theta_B) f_B(\theta_B) d\theta_B + \int_{a_S}^{b_S} U_S(\theta_S) f_S(\theta_S) d\theta_S \\
&= \int_{a_B}^{b_B} \int_{a_S}^{b_S} (\theta_B - \theta_S) y(\theta_S, \theta_B) f_S(\theta_S) f_B(\theta_B) d\theta_S d\theta_B \\
&+ \int_{a_S}^{b_S} \theta_S y_S(\theta_S) \Lambda_S(y_S(\theta_S) - 1) f_S(\theta_S) d\theta_S + \int_{a_B}^{b_B} \theta_B y_B(\theta_B) \Lambda_B(y_B(\theta_B) - 1) f_B(\theta_B) d\theta_B
\end{aligned}$$

where we used CPEIC and integration by parts towards the end. Putting these two equations together we get

$$\begin{aligned}
& U_B(a_B) + U_S(b_S) \\
&= \int_{a_B}^{b_B} \int_{a_S}^{b_S} (\theta_B - \theta_S) y(\theta_S, \theta_B) f_S(\theta_S) f_B(\theta_B) d\theta_S d\theta_B \\
&+ \int_{a_S}^{b_S} \theta_S y_S(\theta_S) \Lambda_S(y_S(\theta_S) - 1) f_S(\theta_S) d\theta_S + \int_{a_B}^{b_B} \theta_B y_B(\theta_B) \Lambda_B(y_B(\theta_B) - 1) f_B(\theta_B) d\theta_B \\
&- \int_{a_B}^{b_B} y_B(\theta_B) (1 + \Lambda_B(y_B(\theta_B) - 1)) (1 - F_B(\theta_B)) d\theta_B - \int_{a_S}^{b_S} y_S(\theta_S) (1 - \Lambda_S(y_S(\theta_S) - 1)) F_S(\theta_S) d\theta_S.
\end{aligned}$$

Individual rationality requires  $U_B(a_B) + U_S(b_S) \geq 0$ . We will now show that this condition is never satisfied for any combination of buyer and seller loss-aversion. From our discussion in the main text, we know that it is sufficient to consider the case  $\Lambda_S = 0$ , i.e., no loss-aversion on the trade-dimension for the seller. This allows us to simplify and rewrite to

$$\begin{aligned}
& U_B(a_B) + U_S(b_S) \\
&= \int_{a_B}^{b_B} \int_{a_S}^{b_S} \left( \left[ \theta_B - \frac{1 - F_B(\theta_B)}{f_B(\theta_B)} \right] - \left[ \theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)} \right] \right) y(\theta_S, \theta_B) f_B(\theta_B) f_S(\theta_S) d\theta_S d\theta_B \\
&+ \Lambda_B \int_{a_B}^{b_B} y_B(\theta_B) (y_B(\theta_B) - 1) \left[ \theta_B - \frac{1 - F_B(\theta_B)}{f_B(\theta_B)} \right] f_B(\theta_B) d\theta_B.
\end{aligned}$$

MS show in their proof of Theorem 1 (p. 269) that

$$\begin{aligned}
& \int_{a_B}^{b_B} \int_{a_S}^{b_S} \left( \left[ \theta_B - \frac{1 - F_B(\theta_B)}{f_B(\theta_B)} \right] - \left[ \theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)} \right] \right) y(\theta_S, \theta_B) f_B(\theta_B) f_S(\theta_S) d\theta_S d\theta_B \\
&= - \int_{a_B}^{b_S} (1 - F_B(x)) F_S(x) dx.
\end{aligned}$$

Further, we have  $y_B(\theta_B) = F_S(\theta_B)$  since we are considering the ex post efficient mechanism. Putting this together yields

$$\begin{aligned}
U_B(a_B) + U_S(b_S) &= - \int_{a_B}^{b_S} (1 - F_B(x))F_S(x) dx \\
&\quad + \Lambda_B \int_{a_B}^{b_B} F_S(x)(F_S(x) - 1) \left[ x - \frac{1 - F_B(x)}{f_B(x)} \right] f_B(x) dx.
\end{aligned}$$

Careful inspection of the limits of the integrals shows that

$$\begin{aligned}
U_B(a_B) + U_S(b_S) &= - \int_{\max\{a_B, a_S\}}^{\min\{b_S, b_B\}} (1 - F_B(x))F_S(x) dx \\
&\quad + \Lambda_B \int_{\max\{a_B, a_S\}}^{\min\{b_S, b_B\}} F_S(x)(F_S(x) - 1) \left[ x - \frac{1 - F_B(x)}{f_B(x)} \right] f_B(x) dx \\
&= - \int_{\max\{a_B, a_S\}}^{\min\{b_S, b_B\}} (1 - F_B(x))F_S(x) + \Lambda_B(1 - F_S(x))F_S(x) \left[ x - \frac{1 - F_B(x)}{f_B(x)} \right] f_B(x) dx \\
&= - \int_{\max\{a_B, a_S\}}^{\min\{b_S, b_B\}} (1 - F_B(x))F_S(x)(1 - \Lambda_B(1 - F_S(x))) + \Lambda_B(1 - F_S(x))F_S(x)x f_B(x) dx \\
&< 0,
\end{aligned}$$

violating individual rationality. To conclude the proof, recall from our discussion of the information rents, that loss-aversion in the money dimension makes the problem unambiguously harder, as it reduces the gains from trade without affecting the information rents. Thus, impossibility in the absence of loss-aversion in the money dimension implies impossibility in the presence of loss-aversion in the money dimension.

## B.2 Revenue Maximizing Mechanism

*Step 1.* We begin by imposing CPEIC. In order for the CPEIC constraint to be satisfied, conditions (i) and (ii) from Proposition 2 must be satisfied. Using the utility functions given in equations (4) and (5) from condition (ii), we can rewrite the objective function in the problem (RM) to

$$\begin{aligned}
&\int_{a_B}^{b_B} \left( \eta_B^2 w_B(\theta_B) + \theta_B \tilde{v}_B(\theta_B) - U_B(a_B, s_S^t|a_B) - \int_{a_B}^{\theta_B} \tilde{v}_B(t) dt \right) dF_B(\theta_B) \\
&+ \int_{a_S}^{b_S} \left( \eta_S^2 w_S(\theta_S) - \theta_S \tilde{v}_S(\theta_S) - U_S(b_S, s_B^t|b_S) - \int_{\theta_S}^{b_S} \tilde{v}_S(t) dt \right) dF_S(\theta_S).
\end{aligned}$$

From the IR constraint we have  $U_B(a_B, \theta_S|a_B) \geq 0$  and  $U_S(b_S, \theta_B|b_S) \geq 0$ , which enter the objective function negatively. Since we are maximizing the objective function, we choose transfers such that  $U_B(a_B, \theta_S|a_B) = 0$  and  $U_S(b_S, \theta_B|b_S) = 0$ . If the expected utility of these “worst” types was not equal to zero in the optimal mechanism, we could modify the transfers by adding lump-sum transfers and reduce their expected utility to

zero without affecting CPEIC. Moreover,  $w_B$  and  $w_S$ , which are negative by Lemma 3, enter positively. Thus, we impose an additional restriction on transfers, namely that they are interim deterministic, which leads to  $w_B(\theta_B) = w_S(\theta_S) = 0$  for all  $\theta_B, \theta_S \in [a, b]$ . Note that these two restrictions on transfers do not contradict each other. Given this, the problem reduces to

$$\begin{aligned} \max_{(y^f)} & \int_{a_B}^{b_B} \left( \theta_B \tilde{v}_B(\theta_B) - \int_{a_B}^{\theta_B} \tilde{v}_B(t) dt \right) dF_B(\theta_B) \\ & + \int_{a_S}^{b_S} \left( -\theta_S \tilde{v}_S(\theta_S) - \int_{\theta_S}^{b_S} \tilde{v}_S(t) dt \right) dF_S(\theta_S) \end{aligned}$$

subject to  $\tilde{v}_S$  being non-increasing,  $\tilde{v}_B$  being non-decreasing,

which proves Proposition 4.

*Step 2.* We next impose that types are uniformly distributed on  $[a, a + 1]$  and rewrite the objective function in this reduced problem. Using integration by parts we get

$$\begin{aligned} & \int_a^b \left( \theta_B \tilde{v}_B(\theta_B) - \int_a^{\theta_B} \tilde{v}_B(t) dt \right) d\theta_B + \int_a^b \left( -\theta_S \tilde{v}_S(\theta_S) - \int_{\theta_S}^b \tilde{v}_S(t) dt \right) d\theta_S \\ & = \int_a^b (2\theta_B - 1 - a) \tilde{v}_B(\theta_B) d\theta_B - \int_a^b (2\theta_S - a) \tilde{v}_S(\theta_S) d\theta_S. \end{aligned}$$

Further, we can write

$$\begin{aligned} \tilde{v}_B(\theta_B) &= \int_a^b y^f(\theta_S, \theta_B) d\theta_S + \eta_B^1 \int_a^b \int_a^b \mu^1 (y^f(\theta_S, \theta_B) - y^f(\theta'_S, \theta_B)) d\theta'_S d\theta_S \\ &= y_B(\theta_B) + \eta_B^1 [y_B(\theta_B)(1 - y_B(\theta_B)) - \lambda_B^1 (1 - y_B(\theta_B)) y_B(\theta_B)] \\ &= y_B(\theta_B) + y_B(\theta_B) \Lambda_B (y_B(\theta_B) - 1) \\ &= y_B(\theta_B) (1 + \Lambda_B (y_B(\theta_B) - 1)), \end{aligned}$$

where  $\int_a^b y^f(\theta_S, \theta_B) d\theta_S = y_B(\theta_B)$ . Analogously, we can write  $\tilde{v}_S(\theta_S) = y_S(\theta_S)(1 - \Lambda_S(y_S(\theta_S) - 1))$ . Note that therefore the constraints that  $\tilde{v}_S$  is non-increasing and  $\tilde{v}_B$  non-decreasing are equivalent to  $y_S$  being non-increasing and  $y_B$  being non-decreasing given the assumption that gain-loss utility does not dominate. Thus, we have reduced the maximization problem to

$$\begin{aligned} \max_{y^f \in \mathcal{Y}} & \int_a^b (2\theta_B - 1 - a) y_B(\theta_B) (1 + \Lambda_B (y_B(\theta_B) - 1)) d\theta_B \\ & - \int_a^b (2\theta_S - a) y_S(\theta_S) (1 - \Lambda_S (y_S(\theta_S) - 1)) d\theta_S, \end{aligned} \tag{RM'}$$

subject to  $y_B$  being non-decreasing and  $y_S$  being non-increasing.



*Step 3.* We will make use of the reduced-form approach as in Che et al. (2013) to maximize directly over the interim trade probabilities  $y_B$  and  $y_S$  instead of the ex post allocation rule  $y^f$ . First, we perform a change of variables to rewrite the objective function to

$$\max_{y^f \in \mathcal{Y}} \int_0^1 (2x - 1 + a)q_B(x)(1 + \Lambda_B(q_B(x) - 1)) dx - \int_0^1 (2x + a)q_S(x)(1 - \Lambda_S(q_S(x) - 1)) dx,$$

where  $q_i(x) = y_i(x + a)$  for all  $x \in [0, 1]$ . Making use of Corollary 6 in Che et al. (2013), we maximize directly over  $q_B$  and  $q_S$  subject to an allocation and an aggregate constraint. The problem then reads

$$\max_{q_B, q_S} \int_0^1 (2x - 1 + a)q_B(x)(1 + \Lambda_B(q_B(x) - 1)) dx - \int_0^1 (2x + a)q_S(x)(1 - \Lambda_S(q_S(x) - 1)) dx,$$

subject to  $q_B$  being non-decreasing,  $q_S$  being non-increasing, the allocation constraint

$$\int_{\theta_S}^1 (1 - q_S(t)) dt + \int_{\theta_B}^1 q_B(t) dt \leq 1 - \theta_B \theta_S$$

for all  $(\theta_B, \theta_S) \in [0, 1]^2$  and the aggregate constraint

$$\int_0^1 (1 - q_S(t)) dt + \int_0^1 q_B(t) dt = 1.$$

The allocation constraint is the condition known from Border (1991) and aggregate constraint ensures that the good is either allocated to the buyer or the seller. Following the proof of Lemma 4 in Mierendorff (2016) we can rewrite the allocation constraint to

$$\int_{\theta_S}^1 (1 - q_S(t)) dt \leq \min_{\theta_B \in [0, 1]} \left[ 1 - \theta_S \theta_B - \int_{\theta_S}^1 q_B(t) dt \right]$$

for all  $\theta_B \in [0, 1]$  and since we are minimizing a convex function on the right-hand side, we obtain

$$\int_{\theta_S}^1 (1 - q_S(t)) dt \leq 1 - q_B^{-1}(\theta_S) \theta_S - \int_{y_B^{-1}(\theta_S)}^1 q_B(t) dt$$

for all  $\theta_S \in [0, 1]$ . This constraint is satisfied with equality when  $q_S^*(t) = 1 - q_B^{-1}(t)$ , where  $q_B^{-1}$  denotes the generalized inverse. In what follows, we will show that for a given, non-decreasing function  $q_B$ , the function  $q_S^*(t) = 1 - q_B^{-1}(t)$  minimizes

$$\int_0^1 (2x + a)q_S(x)(1 - \Lambda_S(q_S(x) - 1)) dx$$

subject to the allocation and aggregate constraint and to  $q_S$  being non-increasing. This

implies that is enough to maximize over the set of all non-decreasing trade probabilities  $q_B$  such that  $q_S(t) = 1 - q_B^{-1}(t)$ . Consider some other candidate to the solution,  $\tilde{q}_S$  which satisfies the allocation constraints and is different from  $q_S^*$  on a set of positive measure. Then there must exist an interval  $[\underline{u}, \bar{u}]$  such that

$$\int_{\theta_S}^1 (1 - \tilde{q}_S(t)) dt < \int_{\theta_S}^1 (1 - q_S^*(t)) dt$$

for all  $\theta_S \in [\underline{u}, \bar{u}]$ . We will now construct a function  $\hat{q}_S$  which does better than the candidate  $\tilde{q}_S$ , thereby proving that  $q_S^*$  is indeed optimal. To do this, we show that there exist  $\bar{p}, \underline{p} \in [0, 1]$  and  $p \in (\underline{p}, \bar{p})$  such that (1)  $\hat{q}_S(t) = \tilde{q}_S(t)$  for all  $t \notin [\underline{p}, \bar{p}]$ , (2)  $\hat{q}_S(t) \geq \tilde{q}_S(t)$  for all  $t \in [\underline{p}, \bar{p}]$ , (3)  $\hat{q}_S(t) \leq \tilde{q}_S(t)$  for all  $t \in [p, \bar{p}]$ , (4)

$$\int_{\underline{p}}^{\bar{p}} q_S^*(t) - \tilde{q}_S(t) dt = 0,$$

and (5)

$$\int_{\theta_S}^1 (1 - \hat{q}_S(t)) dt \leq \int_{\theta_S}^1 (1 - q_S^*(t)) dt$$

for all  $\theta_S \in [0, 1]$ . Fix some  $\bar{p} \in [\underline{u}, \bar{u}]$  and define  $\hat{q}_S(t) = \tilde{q}_S(t)$  for all  $t > \bar{p}$ . Note that by the aggregate constraint there must exist  $0 \leq p < \bar{p}$  such that

$$\int_{\underline{p}}^1 (1 - \hat{q}_S(t)) dt = \int_{\underline{p}}^1 (1 - q_S^*(t)) dt$$

when  $\hat{q}_S(t) = \tilde{q}_S(\bar{p})$  for all  $t \in [p, \bar{p}]$ . This construction satisfies the monotonicity and the allocation constraint. If there now exists a  $0 \leq \underline{p} < p$  such that  $\hat{q}_S(t) = \tilde{q}_S(\underline{p})$  for all  $t \in [\underline{p}, p]$  and  $\hat{q}_S(t) = \tilde{q}_S(t)$  for all  $t < \underline{p}$  with

$$\int_{\underline{p}}^{\bar{p}} \hat{q}_S(t) - \tilde{q}_S(t) dt = 0$$

we are done. If not, then we must have even with  $\underline{p} = 0$  that

$$\int_{\underline{p}}^{\bar{p}} \hat{q}_S(t) - \tilde{q}_S(t) dt + \int_0^{\underline{p}} \hat{q}_S(t) - \tilde{q}_S(t) dt > 0.$$

If

$$\int_{\underline{p}}^{\bar{p}} \hat{q}_S(t) - \tilde{q}_S(t) dt - \int_0^{\underline{p}} \tilde{q}_S(t) dt < 0,$$

then there must exist  $c > 0$  such that  $\hat{q}_S(t) = c$  for  $t \in [0, p)$  yields

$$\int_p^{\bar{p}} \hat{q}_S(t) - \tilde{q}_S(t) dt + \int_0^p \hat{q}_S(t) - \tilde{q}_S(t) dt = 0.$$

If not, then increase  $p$  until

$$\int_p^{\bar{p}} \hat{q}_S(t) - \tilde{q}_S(t) dt - \int_0^p \tilde{q}_S(t) dt = 0.$$

Such a  $p$  exists and the such constructed  $\hat{q}_S$  satisfies the above (1) to (5). Thus, we have constructed  $\hat{q}_S$  from  $\tilde{q}_S$  by shifting trade probability from high types to low types, while satisfying the allocation constraint. This was possible, because  $\tilde{q}_S$  is different from  $q_S^*$  on a set of positive measure and the aggregate constraint needs to be satisfied.

We will now show, that

$$\int_0^1 (2x + a) \hat{q}_S(x) (1 - \Lambda_S [\hat{q}_S(x) - 1]) dx \leq \int_0^1 (2x + a) \tilde{q}_S(x) (1 - \Lambda_S [\tilde{q}_S(x) - 1]) dx,$$

implying that  $\tilde{q}_S$  cannot be a minimizer. We have

$$\begin{aligned} & \int_0^1 (2x + a) (\hat{q}_S(x) (1 - \Lambda_S [\hat{q}_S(x) - 1]) - \tilde{q}_S(x) (1 - \Lambda_S [\tilde{q}_S(x) - 1])) dx \\ &= \int_p^{\bar{p}} (2x + a) (\hat{q}_S(x) (1 - \Lambda_S [\hat{q}_S(x) - 1]) - \tilde{q}_S(x) (1 - \Lambda_S [\tilde{q}_S(x) - 1])) dx \end{aligned}$$

by our construction of  $\hat{q}_S$ . Furthermore, whenever  $\hat{q}_S(x) > \tilde{q}_S(x)$ , we also have  $\hat{q}_S(x) (1 - \Lambda_S [\hat{q}_S(x) - 1]) > \tilde{q}_S(x) (1 - \Lambda_S [\tilde{q}_S(x) - 1])$ . Thus, we obtain

$$\begin{aligned} & \int_p^{\bar{p}} (2x + a) (\hat{q}_S(x) (1 - \Lambda_S [\hat{q}_S(x) - 1]) - \tilde{q}_S(x) (1 - \Lambda_S [\tilde{q}_S(x) - 1])) dx \\ & \leq (2p + a) \int_p^{\bar{p}} (\hat{q}_S(x) (1 - \Lambda_S [\hat{q}_S(x) - 1]) - \tilde{q}_S(x) (1 - \Lambda_S [\tilde{q}_S(x) - 1])) dx \end{aligned}$$

because the difference in the brackets is positive until  $p$  and then negative. Rewrite this difference to obtain

$$\begin{aligned} & \int_p^{\bar{p}} (\hat{q}_S(x) (1 - \Lambda_S [\hat{q}_S(x) - 1]) - \tilde{q}_S(x) (1 - \Lambda_S [\tilde{q}_S(x) - 1])) dx \\ &= (1 + \Lambda_S) \int_p^{\bar{p}} (\hat{q}_S(x) - \tilde{q}_S(x)) dx + \Lambda_S \int_p^{\bar{p}} (\tilde{q}_S(x) - \hat{q}_S(x)) (\hat{q}_S(x) + \tilde{q}_S(x)) dx. \end{aligned}$$

The first integral is equal to zero by construction. In the second integral, note that the first bracket is negative until  $p$  and then positive and the second bracket is a decreasing

function. Thus,

$$\begin{aligned}
& \Lambda_S \int_p^{\bar{p}} (\tilde{q}_S(x) - \hat{q}_S(x)) (\hat{q}_S(x) + \tilde{q}_S(x)) dx \\
& \leq (\hat{q}_S(p) + \tilde{q}_S(p)) \Lambda_S \int_p^{\bar{p}} (\tilde{q}_S(x) - \hat{q}_S(x)) dx \\
& = 0.
\end{aligned}$$

Overall, we have showed that

$$\int_0^1 (2x + a) (\hat{q}_S(x) (1 - \Lambda_S [\hat{q}_S(x) - 1]) - \tilde{q}_S(x) (1 - \Lambda_S [\tilde{q}_S(x) - 1])) dx \leq 0$$

proving that  $\tilde{q}_S$  was not a minimizer and that  $q_S^*$  indeed is the solution to the problem.

*Step 4.* Having eliminated the seller's interim trade probability from the problem using the allocation and aggregate constraints, the maximization problem reads

$$\begin{aligned}
& \max_{q_B} \int_0^1 (2x - 1 + a) q_B(x) (1 + \Lambda_B [q_B(x) - 1]) dx \\
& - \int_0^1 (2x + a) (1 - q_B^{-1}(x)) (1 + \Lambda_S q_B^{-1}(x)) dx
\end{aligned}$$

subject to  $q_B$  being non-decreasing. We now use the substitution  $x = q_B(t)$  to eliminate  $q_B^{-1}$  from the problem and obtain

$$\int_0^1 (2x - 1 + a) q_B(x) (1 + \Lambda_B [q_B(x) - 1]) dx - \int_{q_B^{-1}(0)}^{q_B^{-1}(1)} (2y_B(x) + a) (1 - x) (1 + \Lambda_S x) q_B'(x) dx$$

Note that  $q_B$  is differentiable almost everywhere and therefore the substitution is well-defined. We will guess and verify that  $y_B^{-1}(0) = 0$  and  $y_B^{-1}(1) = 1$ . The objective then becomes

$$\int_0^1 (2x - 1 + a) q_B(x) (1 + \Lambda_B [q_B(x) - 1]) - (2q_B(x) + a) (1 - x) (1 + \Lambda_S x) y_B'(x) dx.$$

We perform one final substitution to ensure the positivity of  $q_B$  and let  $q_B(t) = u^2(t)$ . We then obtain

$$\begin{aligned}
& \int_0^1 (2x - 1 + a) u^2(x) (1 + \Lambda_B [u^2(x) - 1]) - (2u^2(x) + a) (1 - x) (1 + \Lambda_S x) 2u(x) u'(x) dx \\
& = \int_0^1 J(x, u, u') dx.
\end{aligned}$$

We know from methods of calculus of variations that a necessary condition for a solution

to the problem is characterized by

$$\frac{d}{dx} J_{u'}(x, u, u') = J_u(x, u, u').$$

We obtain the candidates for a maximum given by

$$u(x) = 0 \text{ and } u(x) = \pm \frac{\sqrt{(2x-1)(1-\Lambda_B) + 2a^2\Lambda_S - a((2x-1)\Lambda_S + 2 - \Lambda_B)}}{\sqrt{2(1-(2x-1-a)\Lambda_B + (2x-1-2a)\Lambda_S)}}$$

where the second candidate is only well-defined for all

$$x \geq \bar{x} = \frac{2a^2\Lambda_S + a\Lambda_B + a\Lambda_S - 2a + \Lambda_B - 1}{2(a\Lambda_S + \Lambda_B - 1)}.$$

Note that  $x \leq a + 1$  when  $\Lambda_S \leq (1 - \Lambda_B(a + 1))/a$  and  $\Lambda_B \leq 1/(1 + a)$ . Reversing the substitutions we obtain that the optimal interim trade probability for the buyer is

$$y_B^*(\theta_B) = \frac{2\theta_B(1 - 2\Lambda_B - 2\Lambda_S a) + 2a^2\Lambda_S + a(\Lambda_B + \Lambda_S - 2) + \Lambda_B - 1}{2(1 - \Lambda_B(2\theta_B - 1 - a) + \Lambda_S(2\theta_B - 1 - 2a))}$$

if  $\Lambda_S \leq (1 - \Lambda_B(a + 1))/a$  and  $\Lambda_B \leq 1/(1 + a)$  and  $y_B^*(\theta_B) = 0$  otherwise. This interim trade probability (and the corresponding for the seller) can be obtained by the ex post trade rule described in Proposition 5.

*Step 5.* One can easily verify that the IR constraints are satisfied.

### B.3 Welfare Maximizing Mechanism

*Step 1.* We first rewrite the problem as a function of the trade rule only. We can rewrite the objective function to (imposing  $\Lambda_B = \Lambda_S = \Lambda$ )

$$\begin{aligned} & \int_a^b U_B(\theta_B, s_B^t | \theta_B) d\theta_B + \int_a^b U_S(\theta_S, s_B^t | \theta_S) d\theta_S \\ &= \int_a^b (\theta_B y_B(\theta_B)(1 + \Lambda(y_B(\theta_B) - 1)) - \bar{t}_B(\theta_B) + \eta_B^2 w_B(\theta_B)) d\theta_B \\ & \quad - \int_a^b (\theta_S y_S(\theta_S)(1 - \Lambda(y_S(\theta_S) - 1)) - \bar{t}_S(\theta_S) - \eta_S^2 w_S(\theta_S)) d\theta_S. \end{aligned}$$

Note that by the budget constraint (AB) we have

$$\int_a^b t_B(\theta_B) d\theta_B = \int_a^b t_S(\theta_S) d\theta_S.$$

Further,  $w_B(\theta_B)$  and  $w_S(\theta_S)$  enter the objective positively. By Lemma 3 both are negative and, hence, optimally set to zero by choosing interim deterministic transfers. This yields

$$\begin{aligned}
& \int_a^b U_B(\theta_B, s_S^t | \theta_B) d\theta_B + \int_a^b U_S(\theta_S, s_B^t | \theta_S) d\theta_S \\
&= \int_a^b \theta_B y_B(\theta_B) (1 + \Lambda(y_B(\theta_B) - 1)) d\theta_B - \int_a^b \theta_S y_S(\theta_S) (1 - \Lambda(y_S(\theta_S) - 1)) d\theta_S.
\end{aligned}$$

Mirroring the arguments in the proof of the revenue maximizing mechanism, the budget constraint AB and the CPEIC can be jointly written as

$$\begin{aligned}
& \int_a^b (2\theta_B - 1 - a) y_B(\theta_B) (1 + \Lambda[y_B(\theta_B) - 1]) d\theta_B \\
&= \int_a^b (2\theta_S - a) y_S(\theta_S) (1 - \Lambda[y_S(\theta_S) - 1]) d\theta_S,
\end{aligned}$$

as well as the monotonicity constraints. Thus, the maximization problem is a function of the trade rule only.

*Step 2.* We can set up the Lagrangian as

$$\begin{aligned}
\mathcal{L}(y^f, \gamma) &= \int_a^b (\theta_B + \gamma(2\theta_B - 1 - a)) y_B(\theta_B) (1 + \Lambda[y_B(\theta_B) - 1]) d\theta_B \\
&\quad - \int_a^b (\theta_S + \gamma(2\theta_S - a)) y_S(\theta_S) (1 - \Lambda[y_S(\theta_S) - 1]) d\theta_S.
\end{aligned}$$

Note that we must have  $\gamma \geq 0$ , because relaxing the budget constraint (i.e., allowing the designer to run a deficit) can only increase the objective. Hence,  $(\theta_B + \gamma(2\theta_B - 1 - a))$  and  $(\theta_S + \gamma(2\theta_S - a))$  are strictly increasing in  $\theta_B$  and  $\theta_S$ , respectively. Therefore, the arguments in the proof of the revenue maximizing mechanism carry through and we can again maximize over the interim trade probabilities directly and eliminate  $y_S$  from the problem.

*Step 3.* Mirroring the steps in the proof of the revenue maximizing mechanism we obtain an expression for the interim trade probability of the buyer. Using the budget constraint and the assumption that  $\Lambda = \Lambda_B = \Lambda_S$  we can eliminate the Lagrange multiplier from this expression. Next, reversing the change in variables we obtain the buyer's interim trade probability given by from which we can recover the ex post allocation rule which gives rise to the interim trade probabilities and is given in Proposition 7. The optimality of no trade for large enough stakes follows directly from the revenue maximizing mechanism. We know from there that for  $\Lambda \leq 1/(1 + a)$  any mechanism which induces trade yields a negative expected revenue. Hence, any mechanism which induces trade violates the budget balance constraint. Consequently, for  $\Lambda \leq 1/(1 + a)$  no trade is the only feasible welfare maximizing mechanism.

*Step 4.* One can easily verify that the IR constraints are satisfied.

## References

- ABELER, J., A. FALK, L. GOETTE, AND D. HUFFMAN (2011): “Reference Points and Effort Provision,” *American Economic Review*, 101, 470–492.
- BARTLING, B., L. BRANDES, AND D. SCHUNK (2015): “Expectations as Reference Points: Field Evidence from Professional Soccer,” *Management Science*, 61, 2646–2661.
- BELL, D. E. (1985): “Disappointment in Decision Making under Uncertainty,” *Operations Research*, 33, 1–27.
- BERGEMANN, D. AND S. MORRIS (2005): “Robust Mechanism Design,” *Econometrica*, 73, 1771–1813.
- BIERBRAUER, F. AND N. NETZER (2016): “Mechanism Design and Intentions,” *Journal of Economic Theory*, 163, 557–603.
- BORDER, K. C. (1991): “Implementation of Reduced Form Auctions: A Geometric Approach,” *Econometrica*, 59, 1175–1187.
- CARBAJAL, J. C. AND J. C. ELY (2016): “A Model of Price Discrimination under Loss Aversion and State-Contingent Reference Points,” *Theoretical Economics*, 11, 455–485.
- CAVALLO, R. (2011): “Efficient Mechanisms with Risky Participation,” in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, ed. by T. Walsh, AAAI Press/International Joint Conferences on Artificial Intelligence.
- CHATTERJEE, K. AND W. SAMUELSON (1983): “Bargaining under Incomplete Information,” *Operations Research*, 31, 835–851.
- CHE, Y.-K., J. KIM, AND K. MIERENDORFF (2013): “Generalized Reduced-Form Auction: A Network-Flow Approach,” *Econometrica*, 81, 2487–2520.
- COPIC, J. AND C. PONSATÍ (2008): “Ex-Post Constrained-Efficient Bilateral Trade with Risk-Averse Traders,” Mimeo.
- CRAMTON, P., R. GIBBONS, AND P. KLEMPERER (1987): “Dissolving a partnership Efficiently,” *Econometrica*, 55, 615–632.
- CRAWFORD, V. P. AND J. MENG (2011): “New York City Cab Drivers’ Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income,” *American Economic Review*, 101, 1912–1932.

- DE MEZA, D. AND D. C. WEBB (2007): “Incentive Design under Loss Aversion,” *Journal of the European Economic Association*, 5, 66–92.
- DELLAVIGNA, S. (2009): “Psychology and Economics: Evidence from the Field,” *Journal of Economic Literature*, 47, 315–372.
- DRIESEN, B., A. PEREA, AND H. PETERS (2012): “Alternating offers Bargaining with loss aversion,” *Mathematical Social Sciences*, 64, 103–118.
- DURAJ, J. (2015): “Mechanism Design with News Utility,” Personal Communication.
- EISENHUTH, R. (2013): “Reference Dependent Mechanism Design,” Mimeo.
- EISENHUTH, R. AND M. EWERS (2012): “Auctions with Loss Averse Bidders,” Working paper, Northwestern University.
- ERICSON, K. M. M. AND A. FUSTER (2011): “Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments,” *Quarterly Journal of Economics*, 126, 1879–1907.
- (2014): “The Endowment Effect,” *Annual Review of Economics*, 6, 555–579.
- FEHR, E. AND L. GOETTE (2007): “Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment,” *American Economic Review*, 97, 298–317.
- FIESELER, K., T. KITTSTEINER, AND B. MOLDOVANU (2003): “Partnerships, lemons, and efficient trade,” *Journal of Economic Theory*, 113, 223–234.
- GARRATT, R. AND M. PYCIA (2015): “Efficient Bilateral Trade,” Mimeo, FRBNY and UCLA.
- GENESOVE, D. AND C. MAYER (2001): “Loss Aversion and Seller Behavior: Evidence from the Housing Market,” *The Quarterly Journal of Economics*, 116, 1233–1260.
- GILL, D. AND V. PROWSE (2012): “A Structural Analysis of Disappointment Aversion in a Real Effort Competition,” *American Economic Review*, 102, 469–503.
- GILL, D. AND R. STONE (2010): “Fairness and desert in tournaments,” *Games and Economic Behavior*, 69, 346–364.
- HERWEG, F., D. MÜLLER, AND P. WEINSCHENK (2010): “Binary Payment Schemes: Moral Hazard and Loss Aversion,” *American Economic Review*, 100, 2451–2477.
- KAHNEMAN, D. AND A. TVERSKY (1979): “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 47, 263–291.



- KARLE, H., G. KIRCHSTEIGER, AND M. PEITZ (2015): “Loss Aversion and Consumption Choice: Theory and Experimental Evidence,” *American Economic Journal: Microeconomics*, 7, 101–120.
- KARLE, H. AND M. PEITZ (2014): “Competition under consumer loss aversion,” *The RAND Journal of Economics*, 45, 1–31.
- KŐSZEGI, B. (2014): “Behavioral Contract Theory,” *Journal of Economic Literature*, 52, 1075–1118.
- KŐSZEGI, B. AND M. RABIN (2006): “A Model of Reference-Dependent Preferences,” *The Quarterly Journal of Economics*, 121, 1133–1165.
- (2007): “Reference-Dependent Risk Attitudes,” *The American Economic Review*, 97, 1047–1073.
- (2009): “Reference-Dependent Consumption Plans,” *American Economic Review*, 99, 909–936.
- KUCUKSENEL, S. (2012): “Behavioral Mechanism Design,” *Journal of Public Economic Theory*, 14, 767–789.
- LOOMES, G. AND R. SUGDEN (1986): “Disappointment and Dynamic Consistency in Choice under Uncertainty,” *The Review of Economic Studies*, 53, 271–282.
- MASATLIOGLU, Y. AND C. RAYMOND (forthcoming): “A Behavioral Analysis of Stochastic Reference Dependence,” *The American Economic Review*.
- MASKIN, E. AND J. RILEY (1984): “Optimal Auctions with Risk Averse Buyers,” *Econometrica*, 52, 1473 – 1518.
- MIERENDORFF, K. (2016): “Optimal dynamic mechanism design with deadlines,” *Journal of Economic Theory*, 161, 190–222.
- MYERSON, R. B. AND M. A. SATTERTHWAITE (1983): “Efficient Mechanisms for Bilateral Trading,” *Journal of Economic Theory*, 29, 265 – 281.
- POPE, D. G. AND M. E. SCHWEITZER (2011): “Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes,” *American Economic Review*, 101, 129–157.
- POST, T., M. J. VAN DEN ASSEM, G. BALTUSSEN, AND R. H. THALER (2008): “Dear or No Deal? Decision Making under Risk in a Large-Payoff Game Show,” *American Economic Review*, 98, 38–71.

- ROSATO, A. (2014): “Loss Aversion in Sequential Auctions: Endogenous Interdependence, Informational Externalities and the “Afternoon Effect”,” Working paper, University of Technology Sydney.
- (2015): “Sequential Negotiations with Loss-Averse Buyers,” Working paper, University of Technology Sydney.
- SALANT, Y. AND R. SIEGEL (2016): “Reallocation Costs and Efficiency,” *American Economic Journal: Microeconomics*, 8, 203–227.
- SHALEV, J. (2002): “Loss Aversion and Bargaining,” *Theory and Decisions*, 52, 201–232.
- SPIEGLER, R. (2012): “Monopoly Pricing when Consumers are Antagonized by Unexpected Price Increases: A “Cover Versio“ of the Heidhues-Koszegi-Rabin Model,” *Economic Theory*, 51, 695–711.
- THALER, R. H. (1980): “Toward a positive theory of consumer choice,” *Journal of Economic Behavior and Organization*, 1, 39–60.
- WOLITZKY, A. (2016): “Mechanism Design with Maxmin Agents: Theory and an Application to Bilateral Trade,” *Theoretical Economics*, 11, 971–1004.