

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 (2022) : 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/10419/251431>

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**^{UZH}

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

First version: November 2014
This version: January 2022

Bilateral Trade with Loss-Averse Agents^{*}

Jean-Michel Benkert[†]

This version: January 2022
First version: November 2014

Abstract

The endowment and attachment effect are empirically well-documented in bilateral trade situations. Yet, the theoretical literature has so far failed to formally identify these effects. We fill this gap by introducing expectations-based loss aversion, which can explain both effects, into the classical setting by Myerson and Satterthwaite (1983). This allows us to formally identify the endowment and attachment effect and study their impact on information rents, allowing us to show that, in contrast to other behavioral approaches to the bilateral trade problem, the impossibility of inducing materially efficient trade persists in the presence of loss aversion. We then turn to the design of optimal mechanisms and consider the problem of maximizing the designer's revenue as well as gains from trade. We find that the designer optimally provides the agents with full insurance in the money dimension and with partial insurance in the trade dimension, thereby reducing ex-post variation in agents' payoffs.

Keywords: Bilateral trade, loss aversion, mechanism design, endowment and attachment effect

JEL Classification: C78, D01, D02, D82, D84, D90

^{*}This paper is a revised version of the first chapter of my PhD thesis submitted at the University of Zurich. I would like to thank Zoltán Balogh, Olivier Bochet, Juan Carlos Carbajal, Eddie Dekel, Jeff Ely, Samuel Häfner, Fabian Herweg, Heiko Karle, Botond Köszegi, René Leal Vizcaíno, Igor Letina, Shangen Li, Shou Liu, Daniel Martin, Konrad Mierendorff, Marc Möller, Wojciech Olszewski, Anne-Katrin Roesler, Yuval Salant, Aleksei Smirnov, Gerhard Sorger, Ran Spiegler, Egor Starkov, Tom Wilkenning, Peio Zuazo Garin, and seminar participants in Bern, Zurich, at the Workshop on Mechanism Design and Behavioural Economic in Glasgow, at the ZWE 2014 and ESEM 2016 for helpful comments. I am especially grateful to my supervisors Nick Netzer and Georg Nöldeke for their 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.

[†]University of Bern, Department of Economics, Schanzeneckstrasse 1, 3001 Bern, Switzerland. Email: jean-michel.benkert@vwi.unibe.ch.

1 Introduction

The bilateral trade setting describes a simple, yet economically important situation. There is a seller, who owns a good and might be willing to sell it, and a buyer, who might be interested in buying it. Different parts of the economic literature have approached this setting differently. In the field of mechanism design we assume that the agents' valuation of the good is private information and study what outcomes a designer can achieve through different institutions. Famously, Myerson and Satterthwaite (1983) have shown that some potential gains from trade will be left on the table as a rule. This impossibility result on efficiency constitutes a cornerstone within economics overall.

In the empirical literature, especially by means of experiments, we have studied how people behave in such trade situations and how the institution affects behavior. Here, two notable effects have been documented: the endowment effect, going back to Thaler (1980), and, more recently, the attachment effect (Ericson and Fuster, 2011). The endowment effect tells us that ownership of the good drives up the seller's valuation of the good. The attachment effect tells us that a buyer can get attached to a good she does not own (yet) and that this attachment drives up her valuation for it. Both of these empirical findings can be explained by the model of expectations-based loss aversion by Köszegi and Rabin (2006, 2007), which builds on the seminal work by Kahneman and Tversky (1979).¹ In their model, essentially, people compare an outcome to some reference point, which is given by their initial, rational expectations of the outcome. In the case of the buyer, for instance, it is the expectation that she will buy the good which leads to her attachment to the good. Importantly, the neoclassical model which is employed in Myerson and Satterthwaite (1983) cannot explain why these effects would materialize. The model by Köszegi and Rabin, however, is a natural candidate to better understand these empirical effects and their implications on trade situations from a theoretical perspective.

In this paper, we thus introduce expectations-based loss aversion into an otherwise standard mechanism-design approach to the bilateral-trade problem. More specifically, we augment the model by Myerson and Satterthwaite (1983) (henceforth MS), in which both agents have quasi-linear utility over ownership of the good and money, by allowing for both agents to have reference-dependent preferences as modeled in Köszegi and Rabin (2006, 2007) (henceforth KR). We call the standard utility from ownership of the good and money *material utility*, and, in addition, introduce *gain-loss utility* with respect to both, money and ownership of the good, separately. The reference point, relative to which agents evaluate an outcome, is formed endogenously as the rational expectations

¹There is a substantial empirical 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). In particular, see Ericson and Fuster (2014) for an excellent review on the role of loss aversion in explaining behavioral effects in exchange situations.

over the outcome.² We introduce the formal framework in detail in Section 2, where we also characterize incentive compatible mechanisms.

We then begin our analysis in Section 3 by identifying the theoretical counterparts of the endowment and attachment effect. In Proposition 1, we show that in any incentive compatible mechanism, loss aversion, by means of the attachment and the endowment effect, reduces the information rent of the buyer and increases the information rent of the seller, respectively. To better understand this and to fix ideas, consider the mechanism in which trade takes place whenever the buyer values the good more than the seller, i.e., whenever trade is *materially efficient*. In the absence of loss aversion, the buyer has an incentive to imitate a lower type, that is, pretend that she does not value the good as much as she actually does, in order to drive down the price she has to pay for it. The flip side of this behavior, is that by doing so, she reduces the probability of trade actually taking place. This is where expectations-based loss aversion kicks in. The possibility of getting the good induces an attachment to the good, which, if trade was to not take place, gives rise to a feeling of loss. In order to avoid this loss, which is felt more strongly than a commensurate gain, the buyer is less eager to shade her valuation than in the absence of loss aversion. Consequently, it is easier to induce truthful behavior from the buyer and thus her information rent decreases due to the attachment effect. Turning to the seller, we find that the endowment effect plays out in a similar fashion, but with the opposite result. In the absence of loss aversion, the seller wants to imitate a higher type, in order to receive a higher transfer. Loss aversion reinforces this behavior, as reporting a higher type increases the chance of trade not taking place and hence keeping the good the seller is endowed with. Thus, it becomes even harder to induce truthful behavior from the seller and her information rent increases.

The result on the effect of loss aversion on the agents' information rent in Proposition 1 is of interest for two reasons. First, as we've already noted, it formally identifies the theoretical counterparts of the attachment and endowment effect. Second, it suggests an interesting connection to Myerson and Satterthwaite's impossibility result. The standard interpretation of the impossibility result is that the gains from trade cannot cover the information rents that accrue to the agents in order to ensure incentive compatibility given the participation constraints and budget balance. Since loss aversion reduces the buyer's information rent, it could mitigate the severity of the impossibility problem or even reverse it, thus enabling the implementation of materially efficient trade. Indeed, Proposition 2 shows that the presence of a loss-averse buyer can mitigate the impossibility

²Ericson 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. In contrast, see Hefetz and List (2014) and Gneezy, Goette, Sprenger, and Zimmermann (2017) for papers that show the limits of this.

result in the sense that a lower subsidy would be needed to induce materially efficient trade. However, a reversal is beyond reach, as loss aversion not only reduces the buyer's information rent, but also her participation constraint becomes harder to satisfy, due to the ex-post variation in payoffs, which lower expected utility.

We would like to note that the robustness of the impossibility result in the present context is in stark contrast to other papers with non-standard preferences, which show that the impossibility result can be reversed. In the case of intentions-based social preferences the reversal is driven by the fact that 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 (Bierbrauer and Netzer, 2016). Similarly, 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 (Kucuksenel, 2012). Thus, in contrast to the present framework, the channel alleviating the impossibility problem does not conflict with the incentive compatibility or the participation constraints, meaning that a reversal is possible.

In Section 4 we turn to the problem of designing optimal mechanisms and begin with the problem of maximizing the designer's revenue. We show that in the presence of loss aversion any revenue-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. This reduces ex-post variations in payoffs, thereby making loss-averse agents better off. Turning to the optimal trade rule, we impose the assumption that types are drawn from the uniform distribution to keep the model tractable. In spite of this assumption it is not possible to obtain the optimal trade rule using pointwise maximization because the agents' expected utilities endogenously depend on the mechanism through the reference point. In order to nevertheless derive the optimal trade rule we make use of the reduced-form approach. Border (1991) characterized which interim allocation probabilities are implementable by some ex-post allocation rule in the case of single-unit auctions. Che, Kim, and Mierendorff (2013) substantially generalized this result to multi-unit auctions, and also extended the reduced-form approach from auctions to the bilateral-trade setting. Thus, instead of maximizing over the ex-post trade rule, we can maximize directly over the interim trade probabilities subject to some feasibility constraints. The resulting problem is one of optimal control with pure state constraints, which, however, is not concave in the state variable, so that standard sufficiency results are not applicable. Relying on work by Sorger (1989) for non-convex control problems with state constraints, we can derive the globally optimal trade rule and show that the designer optimally induces less trade in the presence of loss aversion. Thus, the designer eliminates all ex-post variation in the agents' transfers, thereby fully insur-

ing 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. Intuitively, as the stakes become larger, it becomes too costly to induce loss-averse agents to take on any uncertainty.

Besides maximizing the designer’s revenue, another natural question to ask is to how the designer can maximize the gains arising from trade. However, in the presence of loss aversion one needs to clarify what the relevant welfare criterion is and how to handle gain-loss utility. The literature on behavioral welfare economics provides some guidance and allows us to distinguish between model-based and model-less approaches (Manzini and Mariotti, 2014). In a model-based approach (e.g. Rubinstein and Salant (2012), Benkert and Netzer (2018)) the welfare criterion is developed based on an underlying theory (or, a model) of mistakes. In contrast, in a model-less approach (e.g., Bernheim and Rangel (2009), Apesteguia and Ballester (2015)) multiple inconsistent preferences are being aggregated into a welfare criterion solely on the basis of observed choices. We find that the distinction does matter in the present setting, but becomes irrelevant for high degrees of loss aversion. In any case, irrespective of the criterion at hand, the designer optimally provides agents with insurance by fully eliminating ex-post variation in transfers and reducing the trade frequency.

1.1 Related literature

Most closely related to our paper is the literature on mechanism design with loss-averse agents. Eisenhuth (2019) 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. This result corresponds to our finding that transfers are interim deterministic in optimal mechanisms and, as one can show, extends beyond the auction and bilateral trade setting. Duraj (2018) considers mechanism design problems with agents who are loss averse on news utility, that is, agents’ utility depends on changes in their beliefs over the outcome as in Kőszegi and Rabin (2009). In an application to bilateral trade he shows the robustness of the impossibility result in this setting.³

Also related is the (increasingly large) literature on behavioral industrial organization

³In an older version of that paper, which was made available by personal communication, Duraj showed that the impossibility result can be reversed under some conditions in the presence of news utility (Duraj, 2015). We thank Niccolò Lomys for making the connection.

with loss-averse agents.⁴ Rosato (2017) considers a sequential bargaining model with a risk-neutral seller and a loss-averse buyer.⁵ Also within the framework of KR, but assuming wide bracketing, he shows that the buyer’s loss aversion softens the rent-efficiency trade off for the seller. 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.⁶ In contrast to the present paper, neither Rosato (2017) (nor Eisenhuth (2019) above) feature loss-averse sellers, but only loss-averse buyers. Heidhues and Kőszegi (2014) and Rosato (2016) consider models with a monopolist selling to expectations-based loss-averse consumers. In both papers the monopolist uses random prices to induce the attachment effect, increasing the consumers willingness to pay and thus profits. In contrast, in the present paper agents are already confronted with uncertainty due to the private nature of types and there is no need to further “inject” randomness to induce the attachment or endowment effect. Indeed, the designer optimally insures agents fully against any variation in transfers and partially in the trade dimension in order to reduce ex-post variation in payoffs.

Finally, our paper also relates to the large literature on the bilateral trade problem, which has followed Myerson and Satterthwaite (1983). Arguably, the departure from the classical setting most closely related to our paper, is to consider risk-averse agents. However, in contrast to loss aversion, risk aversion cannot explain the endowment and attachment effect. Early on, Chatterjee and Samuelson (1983) showed that when agents “become infinitely risk averse” all material gains from trade can be realized using a double-auction. More recently, Garratt and Pycia (2020) examine the bilateral trade problem relaxing the assumption that the agents have quasi-linear utility.⁷ Allowing for risk aversion and wealth effects, they provide conditions for the possibility of realizing all gains of trade. The impossibility result can be reversed in this setting, because the presence of risk aversion and wealth effects give rise to additional gains from trade, which then suffice to cover the agents’ information rents. In contrast to Garratt and Pycia (2020) we do not attempt to establish whether efficient trade with respect to the total gains from trade can be achieved, but approach the problem as one of finding the trade mechanism which maximizes the gains from trade from an ex-ante perspective, finding that it matters whether one wants to maximize total or only material gains from trade, unless loss

⁴See for instance Karle and Möller (2020) and the references therein.

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

⁶The attachment effect also plays a role in a number of other papers, among others Karle and Schumacher (2017) in a model of advertisement or in Rosato (2021) who proposes expectations-based loss aversion as an explanation for the “afternoon effect” observed in sequential auctions.

⁷See also the references in Garratt and Pycia (2020) for more work on the bilateral trade problem in the classic framework with quasi-linear utility following Myerson and Satterthwaite (1983). Moreover, see Wolitzky (2016) and Crawford (2021) for analyses of the bilateral trade problem with maxmin and level- k agents, respectively.

aversion is sufficiently strong.

2 Model

2.1 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 $\Theta = \Theta_S \times \Theta_B$ and assume that Θ_S and Θ_B have a non-trivial intersection. We interpret the type of an agent as her valuation of the good.⁸ 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.⁹ 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 and $\mathbf{r}_i = \{r_i^1, r_i^2\} \in \mathbb{R}^2$ are the so-called riskless reference levels. 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,

⁸We 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.

⁹We follow the literature by abstracting from diminishing sensitivity.

respectively.

We adopt the following assumption from Herweg, Müller, and Weinschenk (2010):¹⁰

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. In particular, KR show that this condition ensures that agents will not choose stochastically dominated options. We will maintain this assumption throughout the paper.¹¹ 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.¹²

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 Γ 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 θ_i as given.

2.2 Equilibrium Concept and Revelation Principle

We use the concept of an (interim) choice-acclimating personal equilibrium (CPE) introduced in Kőszegi and Rabin (2007).¹³ 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 Γ , agent i 's

¹⁰This condition is commonly imposed, see for instance de Meza and Webb (2007), Eisenhuth and Grunewald (2018), Eisenhuth (2019), Karle and Peitz (2014), and Rosato (2021).

¹¹We discuss the implications of relaxing the assumption after confirming the impossibility result in Section 3.

¹²The 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. We briefly discuss relaxing the assumption in the conclusion.

¹³KR also introduce the unacclimating personal equilibrium (UPE). In the UPE the agent "maximizes expected utility taking the reference point as given", whereas in the CPE the agent "maximizes expected utility given that it determines both the reference lottery and the outcome lottery". KR note that the CPE is more appropriate when the uncertainty is resolved after the agent's decision. We thus believe that the CPE is the more natural equilibrium concept in our context, as the report of an agent determines the uncertainty she feels about the outcome given her beliefs about the other agent's type.

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 θ_S , in the mechanism Γ , when playing action $m \in M_B$, given that the buyer plays strategy s_B as

$$\begin{aligned}
U_S(m, s_B, \Gamma | \theta_S) = & \int_{a_B}^{b_B} (1 - y^g(m, s_B(\theta_B))) \theta_S + t_S^g(m, s_B(\theta_B)) dF_B(\theta_B) \\
& + \int_{a_B}^{b_B} \int_{a_B}^{b_B} \eta_S^1 \mu_S^1 (y^g(m, s_B(\theta'_B)) \theta_S - y^g(m, s_B(\theta_B)) \theta_S) dF_B(\theta'_B) dF_B(\theta_B) \quad (3) \\
& + \int_{a_B}^{b_B} \int_{a_B}^{b_B} \eta_S^2 \mu_S^2 (t_S^g(m, s_B(\theta_B)) - t_S^g(m, s_B(\theta'_B))) dF_B(\theta'_B) dF_B(\theta_B) \\
= & \theta_S \int_{a_B}^{b_B} (1 - y^g(m, s_B(\theta_B))) dF_B(\theta_B) + \int_{a_B}^{b_B} t_S^g(m, s_B(\theta_B)) dF_B(\theta_B) \\
& + \theta_S \eta_S^1 \int_{a_B}^{b_B} \int_{a_B}^{b_B} \mu_S^1 (y^g(m, s_B(\theta'_B)) - y^g(m, s_B(\theta_B))) dF_B(\theta'_B) dF_B(\theta_B) \\
& + \eta_S^2 \int_{a_B}^{b_B} \int_{a_B}^{b_B} \mu_S^2 (t_S^g(m, s_B(\theta_B)) - t_S^g(m, s_B(\theta'_B))) dF_B(\theta'_B) dF_B(\theta_B).
\end{aligned}$$

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 last line containing the money gain-loss utility. Fix any θ_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 is induced endogenously and corresponds to the distribution of possible transfers. Thus, for every θ'_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 μ_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 θ_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 θ_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

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

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 (2019).¹⁴

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 Γ 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 Γ^d .

¹⁴In later work than Eisenhuth (2019), Dato, Grunewald, Müller, and Strack (2017) have developed a framework to extend the equilibrium concepts in Kőszegi and Rabin (2006, 2007) to study strategic interaction in finite games. The equilibrium concept they define for the CPE coincides with the one in Eisenhuth (2019) and here. Interestingly, they show that in a CPE players are unwilling to randomize over pure strategies, implying that existence may fail and that restriction to pure strategies is without loss.

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 Γ if and only if f is CPEIC.*

The standard proof of the revelation principle goes through in spite of the presence of an endogenous reference point. To see this, note that the reference point is determined as the rational expectations over outcomes. Starting from an arbitrary mechanism which induces some distribution of outcomes, the corresponding direct mechanism induces the same distribution of outcomes and therefore also the same reference point. Henceforth, we focus on direct mechanisms and no longer explicitly list the mechanism as an argument in the utility function.

2.3 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. Further, we introduce our notion of an interim deterministic mechanism.

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 standard and therefore omitted.¹⁵ 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 *seem* similar as those in the absence of loss aversion, they are not and thus the set of incentive-compatible SCF need not coincide either. We say that a SCF is individually rational if for both agents $i \in I$

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

¹⁵In contrast to Carbajal and Ely (2016), who consider price discrimination using a different model of loss aversion than the one here, the standard integral representation obtains in our setting. This is driven by the fact that, in contrast to Carbajal and Ely (2016), the report of an agent and not her type determines her reference point. For instance, a high buyer type does not expect to get the good with the probability corresponding to her true type when misreporting. Rather, she is aware that reporting a lower type changes the probability of getting the good and this is reflected in her reference point.

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.¹⁶ 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.

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.

3 Attachment, Endowment and Information Rents

As noted in the introduction, the attachment and endowment effect have been empirically documented in bilateral trade situations. However, the classical model with quasi-linear utility as in Myerson and Satterthwaite (1983) cannot explain such effects, motivating the inclusion of expectations-based loss aversion in the present paper. Our first step is thus to formally identify these effects and their implications in our model.

Proposition 3 *In any CPEIC mechanism, the information rent of the seller is increasing in Λ_S and the information rent of the buyer is decreasing in Λ_B .*

Put differently, the presence of loss aversion in the trade dimension increases the information rent of the seller and decreases the information rent of the buyer. The proof is straightforward as it suffices to take the derivatives with respect to Λ_S and Λ_B from equations (4) and (5), respectively. The key step is to note that

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

¹⁶Recall that we are considering net utility and have thus already taken care of the seller's type-dependent outside option.

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

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

Thus, the attachment effect is captured by $-\Lambda_B \int_{a_B}^{\theta_B} y_B(\theta_B)(1 - y_B(\theta_B)) d\theta_B$ and the attachment effect by $\Lambda_S \int_{\theta_S}^{b_S} y_S(\theta_S)(1 - y_S(\theta_S)) d\theta_S$. As already noted, the respective decrease and increase in the rents stem from the loss aversion in the trade dimension. As one would expect, loss aversion on the money dimension plays no role here. To simplify exposition, we will use the term loss aversion as referring to loss aversion in the trade dimension unless stated differently.

Having formally identified the two effects as the impact of loss aversion on the information rents, we can conduct an interesting bit of comparative statics. Does loss aversion affect all types in the same way?

Corollary 1 *The strength of the endowment and attachment effect is increasing in the agents' type.*

The result follows immediately as one takes the derivative with respect to Λ_S and θ_S from equation (4) and with respect to Λ_B and θ_B from equation (5). To gain some intuition for this result, consider the case of the buyer. Recall from the introduction that misreporting to a lower type in order to pay a lower price increases the risk of not getting the good at all, in which case the loss of not getting the good is felt due to loss aversion. A high buyer type values the good more and thus feels the loss of not getting something more valuable more strongly than a lower type who does not care as much about the good in the first place. The analogous force strengthens the endowment effect for higher seller types, who care about the good more strongly.

Our finding about the impact of the endowment and attachment effect on the information rents suggests that the presence of a loss-averse buyer could enable the designer to implement materially efficient trade subject to ex-post budget balance and the agents' participation constraints, that is, to “reverse” the impossibility result by Myerson and Satterthwaite (1983). To see this, recall the interpretation of the result, stating that the gains from trade do not suffice to cover the agents' information rents. Thus, seller loss aversion, which increases the information rent, will make the problem only harder, while buyer loss aversion could make it easier. However, there is a countervailing effect even for the buyer. Loss aversion not only affects information rents as stated in Proposition 3, but also decreases expected utility and hence makes satisfying the participation constraints harder, too.¹⁷ Nevertheless, it suffices to consider buyer loss aversion to check whether

¹⁷Loss aversion on the money dimension does not affect information rents but also reduces expected utility, thus only making it harder to reverse the impossibility result.

the impossibility result can be reversed. Making use of this insight, we can proceed analogously to the proof in Myerson and Satterthwaite (1983). 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

$$\begin{aligned}
U_B(a_B) + U_S(b_S) = & \\
& - \int_{\max\{a_B, a_S\}}^{\min\{b_S, b_B\}} (1 - F_B(x))F_S(x)(1 - \Lambda_B(1 - F_S(x))) + \Lambda_B(1 - F_S(x))F_S(x)xf_B(x) \, dx \\
& < 0,
\end{aligned} \tag{6}$$

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

Proposition 4 *Given CPEIC, IR and BB, it is impossible to realize all material gains from trade for any degree of loss aversion in the money or trade dimension.*

The minimal subsidy needed to induce materially efficient trade under CPEIC and IR (see equation (6)) 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 agents’ types. Indeed, taking the derivative of the minimal subsidy in equation (6) with respect to Λ_B , we can see that the attachment effect mitigates the impossibility problem by dominating the diminishing effect of loss aversion on the participation constraints 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 (6) 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.

The assumption that $\Lambda_i \leq 1$ is commonly imposed in the literature for conceptual as well as technical reasons. In particular, KR showed that the assumption ensures that agents do not choose stochastically dominated options. 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.¹⁸ 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 Λ_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 (6) 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. 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

¹⁸In 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.

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 counterexamples 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 total 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 all the total gains from trade.¹⁹

4 Optimal Mechanisms

The preceding section has formally identified the endowment and the attachment effect in an otherwise standard bilateral trade setting. In particular, we have seen how loss aversion impacts the agents' information rents and the participation constraints, allowing us to show that the impossibility of implementing materially efficient trade extends to the setting with loss-averse agents. We now turn to the problem of designing optimal mechanisms. We begin by considering the problem of maximizing the designer's revenue and then turn to the (conceptually) more nuanced question of maximizing the gains from trade.

In both cases, we begin by maintaining our previous framework with 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 $[0, 1]$. The preceding, more general analysis of the impossibility result suggests that the symmetry of the type spaces is not too restrictive an assumption, as loss aversion does not matter for the extreme types for which trade is interim deterministic. We focus on the uniform distribution for tractability and because it allows us to derive the trade rule explicitly.

¹⁹In 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 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.

4.1 Maximizing the Designer's Revenue

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.²⁰ 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

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

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

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

²⁰The complete derivations and of this section are contained in Appendix A.2.

$$= (1 - \lambda_S^2) \int_{a_B}^{b_B} \int_{a_B}^{b_B} \left(t_S^f(\theta_S, \theta'_B) - t_S^f(\theta_S, \theta_B) \right) \mathbb{1}[t_S^f(\theta_S, \theta'_B) > t_S^f(\theta_S, \theta_B)] dF_B(\theta'_B) dF_B(\theta_B),$$

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 θ_B and θ'_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 5 *Any solution to the revenue maximization problem (RM) entails interim-deterministic transfers.*

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 absence of loss aversion. However, in the presence of loss aversion interim-deterministic transfers are the *only* solution.²¹

For the remainder of this section we will assume that the seller and buyer types are distributed uniformly on $[0, 1]$ with and explicitly derive the optimal trade rule. The assumption allows us to rewrite the maximization problem to

$$\begin{aligned} \max_{y^f \in \mathcal{Y}} & \int_0^1 (2\theta_B - 1) y_B(\theta_B) (1 + \Lambda_B [y_B(\theta_B) - 1]) d\theta_B \\ & - \int_0^1 2\theta_S y_S(\theta_S) (1 - \Lambda_S [y_S(\theta_S) - 1]) d\theta_S, \end{aligned} \tag{RM'}$$

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

²¹Eisenhuth (2019) proved an analogous result for the case of auctions. In fact, one can show that Proposition 5 extends beyond the bilateral trade and auction setting. Further, the result is reminiscent of the optimal mechanism found in Herweg et al. (2010), who augment a principal-agent setting with moral hazard by assuming the agent is expectations-based loss averse as in the present paper. They find that the principal optimally employs a binary payment scheme instead of a fully contingent contract in the presence of loss aversion. Hence, loss aversion drastically reduces the ex-post variation in payments, too, but, in contrast to the present setting, does not eliminate it fully to preserve incentives.

where $y_B(\theta_B) = \int_0^1 y^f(\theta_S, \theta_B) d\theta_S$ and $y_S(\theta_S) = \int_0^1 y^f(\theta_S, \theta_B) d\theta_B$ 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 $2\theta_B - 1 \geq 0$. Clearly, any optimal mechanism will therefore only induce trade for buyer types $\theta_B \geq 1/2$. 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 a 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 the size of the endowment and attachment effect. Thus, there are in some sense externalities between the outcomes of different types. Indeed, because the agents' expected utilities endogenously depend 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, resulting in an optimal control problem with pure state constraints, where y_B is the state and y'_B the control variable. The Hamiltonian of the problem reads

$$H = (2t - 1)y_B(t) (1 - \Lambda_B(1 - y_B(t))) - 2y_B(t)(1 - t) (1 + \Lambda_{St}) y'_B(t) + \lambda(t)y'_B(t),$$

where $\lambda(t)$ is the adjoint variable. Notably, because of the presence of loss aversion, the Hamiltonian is not concave in the state, so not even relatively weak Arrow-type sufficiency

conditions²² are applicable. However, a general result by Sorger (1989) provides sufficiency conditions for control problems with state constraints even in the absence of concavity, so that we can derive for the global optimum.

Proposition 6 *The revenue-maximizing trade rule is given by*

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

where δ^{RM} is non-decreasing in θ_B and non-increasing in the parameters Λ_S and Λ_B .

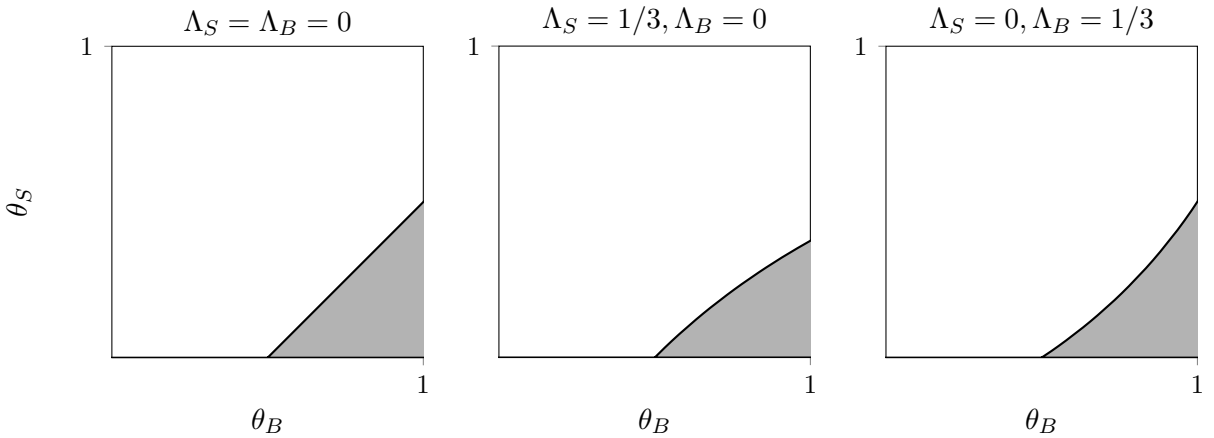


Figure 1: Illustration of the optimal trade rules. The shaded area indicates for which pairs of types trade is taking place. As we can see in the panels in the middle and on the right-hand side, less trade is induced in the presence of loss aversion than without loss aversion as in the panel on the left-hand side.

This result, for which the explicit expression can be found in the proof of the result in Appendix A.2, 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 Myerson and Satterthwaite (1983) given by $\delta^{RM}(\theta_B) = \theta_B - 1/2$. 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.²³ This may come as a surprise in view of how the endowment and attachment effect affect information rents. 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

²²See Arrow (1968) or Seierstad and Sydsaeter (1987).

²³With $\Lambda_B = 1$ no trade takes place, whereas even with $\Lambda_S = 1$ some trade still takes place.

types. However, when types are distributed uniformly, the attachment effect does not mitigate the impossibility problem. Moreover, as we saw in Corollary 1, the attachment and endowment effect are stronger for higher types, as these types value the gain-loss utility more strongly than low types. Further, as we already noted above, the designer wants trade to take place in particular with high buyer types and low seller types. Hence, among the types of agents which are attractive to the designer, the effect of loss aversion is more pronounced for the buyers than the sellers. Put more succinctly, the adverse effect of loss aversion is increasing in the type of the agents and 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, 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.

As we have noted at the beginning of this section, restricting attention to the present case with types distributed uniformly on $[0, 1]$ allows us to derive the optimal trade rule explicitly and renders the problem (more) tractable. Allowing for general distributions should not change the qualitative results as long as we assume some standard regularity conditions on the distributions ensuring that the virtual valuations are increasing. If we are willing to assume that agents are not too loss averse we can derive the optimal mechanism for the case when types are uniformly distributed on $[a, a + 1]$.²⁴ This allows for interesting comparative statics, as the optimal mechanism depends on the parameter a , which, in the context of loss aversion, 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 valuation 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 participation, 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. Contrary to conventional wisdom, the behavioral effects of loss aversion are not mitigated when the stakes are large. Rather, loss aversion has the biggest impact precisely when the stakes are large.

²⁴To be precise, we would need to assume $\Lambda_B \leq 1/(1 + a)$ and $\Lambda_S \leq \min\{1, (1 - \Lambda_B(1 + a))/a\}$. See Appendix A.2 for the derivations.

4.2 Maximizing the Gains from Trade

In this section, we consider the problem of maximizing gains from trade. In the absence of loss aversion, the objective function is given by the sum of ex-ante expected utilities of the two agents. In the presence of loss aversion, however, it may not be clear what constitutes an appropriate objective function. Naturally, one way to go about is to mirror the case without loss aversion and to maximize the sum of ex-ante expected utilities. But what if the designer is only interested in maximizing the material gains from trade, e.g., because she considers loss aversion a mistake?

In standard welfare economics, choice reveals a preference, which in turn should guide any welfare considerations. When choices do not reveal a preference because of inconsistencies or, mistakes, the case is not so clear. Within the field of behavioral welfare economics, we can distinguish between model-based and model-less approaches (Manzini and Mariotti, 2014). In a model-based approach the welfare criterion is developed based on an underlying theory (or, a model) of mistakes. In contrast, in a model-less approach multiple inconsistent preferences are being aggregated into a welfare criterion solely on the basis of observed choices. In analogy, when maximizing the trade from gains the designer could “take loss aversion seriously” and include gain-loss utility in the objective function, or “treat loss aversion as a mistake”, thus only considering material gains from trade in the maximization problem. It is not always straightforward or uncontroversial to determine the “right” approach in such situations. As we will see, the distinction matters in the presence of loss aversion, but becomes irrelevant for high degrees of loss aversion.

In order to formulate the maximization problem, we impose a budget balance condition in addition to CPEIC and IR. 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

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

We consider two maximization problems given by

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

and

$$\max_{(y^f, t_B^f, t_S^f) \in \mathcal{F}} \int_{a_S}^{b_S} (-\theta_S y_S(\theta_S) + \bar{t}_S(\theta_S)) dF_S(\theta_S) + \int_{a_B}^{b_B} (\theta_B y_B(\theta_B) - \bar{t}_B(\theta_B)) dF_B(\theta_B),$$

subject to CPEIC, IR and AB. (MG)

In problem TG the designer includes gain-loss utility in the objective function and thus maximizes what we call total gains from trade, whereas only material gains from trade are maximized in problem MG. To solve either problem, we proceed as we did before and also obtain the result that in any mechanism maximizing total or material gains from trade agents are fully insured against any ex-post variation in transfers.

Proposition 7 *Any solution to the problem (TG) or (MG) entails interim-deterministic transfers.*

The proof is analogous to the revenue maximization problem. To make further progress we again impose that types are uniformly distributed on $[0, 1]$. 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., $\Lambda_B = \Lambda_S = \Lambda$. We then again make use of the reduced-form conditions in Che et al. (2013) and sufficiency conditions in Sorger (1989) 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 8 *The trade rules given by*

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

where δ^{TG} and δ^{MG} are non-decreasing in θ_B and non-increasing in the parameters Λ maximize the problems (TG) or (MG), respectively.

Naturally, the two trade rules coincide when $\Lambda = 0$, as the two problems are identical in that case. For intermediate degrees of loss aversion, the trade rule maximizing only material gains induces less trade and for $\Lambda = 1$ the two once more coincide. Thus, just as in the case with revenue maximization loss aversion impedes trade and the designer optimally provides the agents with partial insurance in the trade dimension and with full insurance in the money dimension. The

5 Conclusion

The theoretical and empirical literature on bilateral trade have both become quite extensive over time. However, the theoretical literature has so far failed to incorporate some findings from the empirical literature, most prominently the well-documented endowment and attachment effect. The present paper aims to fill this gap by augmenting the standard model by Myerson and Satterthwaite (1983) with expectations-based loss aversion as in Kőszegi and Rabin (2006, 2007). In doing so, we also contribute to the literature combining mechanism design and loss aversion (see Kőszegi, 2014).

We first formally identify the endowment and attachment effect and study their impact on the agents' information rents. Using these insights, we can show that buyer loss aversion can mitigate the severity of the impossibility result, but that it remains impossible to implement materially efficient trade. Turning to the design of optimal mechanisms we find that the designer optimally provides agents with insurance in order to reduce ex-post variation in payoffs. More specifically, when maximizing revenue or gains from trade, agents receive full insurance in the money dimension in the form of interim-deterministic transfers, and partial insurance in the form of reduced trade frequency.

One may wonder whether other models than the one by Kőszegi and Rabin can also explain the attachment and endowment effect and thus constitute alternatives to the present analysis. One obvious alternative is a model of loss aversion with a fixed reference point, such as classical prospect theory by Kahneman and Tversky. Indeed, with an appropriately chosen, fixed reference point, such a model can give rise to both attachment and endowment effect. However, the innovation of Kőszegi and Rabin was precisely to determine the reference point endogenously, as otherwise the question of what the appropriate reference point is, remains open. Yet, even with an endogenously determined reference point there exist alternative ways to proceed. Indeed, Kőszegi and Rabin (2007) note that the models of disappointment aversion by Bell (1985) and Loomes and Sugden (1986) are very similar except that the endogenous reference point is given by the certainty equivalent of a lottery rather than the full lottery. However, Masatlioglu and Raymond (2016) find that the intersection of preferences induced by expectations-based loss aversion with CPE and any of these disappointment-aversion models is only standard expected utility, and thus while seemingly similar, the models are actually quite different. Nevertheless, Benkert (2022) shows that the optimal mechanisms for the two types of models are equivalent across a range of mechanism design settings. In particular, the optimal mechanisms derived in the present paper are also optimal if we instead work with a model of disappointment aversion as in Bell (1985) and Loomes and Sugden (1986). This finding 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 forma-

tion process of the reference point, be it fixed, as a full lottery over outcomes or as the certainty equivalent of the lottery. There appears to be some robustness, which suggests that lacking this information may not be too much of a problem, as long as loss-averse individuals are provided with insurance as derived above.

Finally, we have assumed throughout our analysis that the degree of loss aversion is commonly known. If, instead, we assumed that these parameters are private information, a hard multi-dimensional mechanism design problem arises. Our analysis nevertheless provides some insights into this problem. We could relax the assumption that the loss-aversion parameters in the money dimension are commonly known and allow them to be distributed arbitrarily, as the designer optimally eliminates any ex-post variation in the transfers irrespective of the degree of loss aversion. We leave the question of private information regarding the degree of loss aversion in the trade dimension for further research.

A Proofs

A.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 the discussion in the main text we know that we can set the loss aversion in the money dimension to zero, as it only makes the problem harder. 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

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

Myerson and Satterthwaite (1983) 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

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

Careful inspection of the limits of the integrals shows that

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

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

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.

A.2 Maximizing the Designer's Revenue

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

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

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 the arguments in the main text, 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

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

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

which proves Proposition 5.

Step 2. We next impose that types are uniformly distributed on $[a, a + 1]$, thus proving a more general result than stated in Proposition 6 but justifying the statements in the discussion following the proposition and in particular in footnote 24, and rewrite the objective function in this reduced problem. Using integration by parts we get

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

Further, we can write

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

where $\int_a^{a+1} y^f(\theta_S, \theta_B) d\theta_S = y_B(\theta_B)$. Analogously, we can write $\tilde{v}_S(\theta_S) = y_S(\theta_S)(1 + \Lambda_S(1 - y_S(\theta_S)))$. 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

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

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.

Step 4. We next show that the optimal mechanism is such that $y'_B(\theta_B) > 0$ for θ_B with $y_B(\theta_B) > 0$. That is, once the designer induces trade with strictly positive probability for some type, the trade probability increase strictly (and not only weakly, as required by CPEIC). Suppose there is an optimal mechanism with y_B and y_S such that there exists an interval $I = [\underline{i}, \bar{i}] \subset [a, a + 1]$ with $y_B(\theta_B) = c > 0$ for all $\theta_B \in I$. We only consider $c < 1$ and will show in Step 6 that this is without loss, as inducing trade with probability 1 is never optimal. Consider the interval $\tilde{I} = [\underline{i} - \varepsilon, \bar{i} + \delta]$ for some sufficiently small $\varepsilon > 0$ and choose $\delta > 0$ such that

$$\int_{\underline{i} - \varepsilon}^{\bar{i} + \delta} y_B(t) dt = \int_{\underline{i} - \varepsilon}^{\bar{i} + \delta} \tilde{y}_B(t) dt,$$

where \tilde{y}_B is given by the straight line connecting $y_B(\underline{i} - \varepsilon)$ and $y_B(\bar{i} + \delta)$ (which has by construction a strictly positive slope). Clearly, such a δ exists for $\varepsilon > 0$ sufficiently small. To complete the construction of \tilde{y}_B let $y_B = \tilde{y}_B$ for all $\theta_B \in [a, a + 1] \setminus \tilde{I}$. The function \tilde{y}_B thus coincides with y_B except on the interval \tilde{I} where we have eliminated the constant part of the function y_B by shifting trade probability from lower values to higher values of θ_B while keeping the total trade probability constant. It follows directly, that the pair (\tilde{y}_B, y_S) satisfies the aggregate constraint. Moreover, since trade probability has been shifted from lower to higher values we have

$$\int_{\theta_B}^{a+1} y_B(t) dt = \int_{\theta_B}^{a+1} \tilde{y}_B(t) dt$$

for $\theta_B \in [a, a+1] \setminus \tilde{I}$ and

$$\int_{\theta_B}^{a+1} y_B(t) dt \geq \int_{\theta_B}^{a+1} \tilde{y}_B(t) dt$$

for $\theta_B \in \tilde{I}$ so that the allocation constraint is satisfied, too. Finally, since $(2x - 1 + a)y_B(x)(1 - \Lambda_B(1 - y_B(x)))$ is increasing in y_B we have

$$\int_0^1 (2x - 1 + a)y_B(x)(1 + \Lambda_B(y_B(x) - 1)) dx < \int_0^1 (2x - 1 + a)\tilde{y}_B(x)(1 + \Lambda_B(\tilde{y}_B(x) - 1)) dx$$

showing that the initial pair (y_B, y_S) did not constitute an optimal mechanism.

Step 5. 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 $t \in [p, \bar{p}]$, (3) $\hat{q}_S(t) \leq \tilde{q}_S(t)$ for all $t \in [\underline{p}, p)$, (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

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

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

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

If

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

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 6. We next note that no optimal mechanism will induce trade with probability 1 for some set of types. Suppose (y_B, y_S) are a optimal with $y_B(\theta_B) = 1$ for $\theta_B \in [a+1-\delta_1, a+1]$ for some $\delta_1 > 0$ such that $y_B(\theta_B) < 1$ for all $\theta_B < a+1-\delta_1$. We will now construct an alternative pair $(\tilde{y}_B, \tilde{y}_S)$ which does better, showing that the original mechanism was not optimal. Suppose y_B is continuous on $[a+1-\delta_1-\delta_2, a+1-\delta_1]$ for some $\delta_2 > 0$, so that y_B does not jump to 1, although there may be jumps before that. Let $\tilde{y}_B(\theta_B) = y_B(\theta_B)$ for $\theta_B \in [a, 1-\delta_1-\delta_2]$ and $\tilde{y}_B(\theta_B) = y_B(1-\delta_1-\delta_2) = 1-\varepsilon < 1$ for $\theta_B \in [a+1-\delta_1-\delta_2, a+1]$.²⁵ Thus, we have reduced the trade probability on the interval on which it was equal to 1 and a bit before that, thus maintaining monotonicity. Similarly, let $\tilde{y}_S(\theta_S) = y_S(\theta_S)$ for $\theta_S \in [a, 1-\varepsilon]$ and $\tilde{y}_S(\theta_S) = 0$ for $\theta_S \in [a+1-\varepsilon, a+1]$, where we have accordingly reduced trade while maintaining monotonicity.²⁶ Evaluating the objective function on the intervals where the pairs (y_B, y_S) and $(\tilde{y}_B, \tilde{y}_S)$ differ we obtain

$$\begin{aligned} & \int_{a+1-\delta_1-\delta_2}^{a+1} (2\theta_B - 1 + a)(\tilde{y}_B(\theta_B))(1 - \Lambda_B(1 - \tilde{y}_B(\theta_B))) - (2\theta_B - 1 + a)y_B(\theta_B)(1 - \Lambda_B(1 - y_B(\theta_B)))d\theta_B \\ & - \int_{a+1-\varepsilon}^{a+1} (2\theta_S + a)\tilde{y}_S(\theta_S)(1 + \Lambda_S(1 - \tilde{y}_S(\theta_S))) - (2\theta_S + a)y_S(\theta_S)(1 + \Lambda_S(1 - y_S(\theta_S)))d\theta_S \\ & \geq \int_{a+1-\delta_1-\delta_2}^{a+1} (2\theta_B - 1 + a)(1 - \varepsilon)(1 + \Lambda_B\varepsilon) - (2\theta_B - 1 + a)d\theta_B \\ & + \int_{a+1-\varepsilon}^{a+1} (2\theta_S + a)(\delta_1 + \delta_2)(1 + \Lambda_S(1 - \delta_1 - \delta_2))d\theta_S \\ & \geq (\delta_1 + \delta_2)(1 + \delta_2 + \delta_2 - \varepsilon)\varepsilon \end{aligned}$$

The last inequality follows as the previous expression is increasing in both Λ_B and Λ_S . As we are considering small δ_2 and ε the final term is greater than zero, showing that the constructed pair yields a higher revenue. One can easily verify that the above approach also works if instead of y_B being continuous on $[a+1-\delta_1-\delta_2, a+1-\delta_1]$ we had a jump to 1 at $a+1-\delta_1$. Simply construct a \tilde{y}_B which only jumps to $1-\varepsilon$ with the corresponding \tilde{y}_S .

Step 7. Having eliminated the seller's interim trade probability from the problem using the allocation and aggregate constraints, the maximization problem reads

$$\max_{y_B} \int_0^1 (2t - 1 + a)y_B(t) (1 - \Lambda_B(1 - y_B(t))) dt - \int_0^1 (2t + a)(1 - y_B^{-1}(t)) (1 + \Lambda_S y_B^{-1}(t)) dt$$

subject to y_B being non-decreasing. Steps 4 and 6 above imply that there exists an

²⁵If the original y_B is always equal to one can treat it in the same way as we would a jump, see at the end of Step 6.

²⁶Here we use that $y_S = 1 - y_B^{-1}$.

$\tilde{t} \in [a, a+1]$ such that the optimal y_B is equal to zero on $[a, \tilde{t}]$ and then strictly increasing and that it never reaches 1. Therefore, y_B has an actual inverse on $[\tilde{t}, a+1]$. We can thus eliminate y_B^{-1} from the problem and we also reverse our initial substitution from Step 3 to obtain

$$\int_{\tilde{t}}^{a+1} (2t-1-a)y_B(t)(1-\Lambda_B(1-y_B(t))) dt - \int_{\tilde{t}}^1 (2y_B(t)+a)(a+1-t)(1+\Lambda_S(t-a))y'_B(t) dt$$

Note that y_B is differentiable almost everywhere and therefore the substitution is well-defined. Thus, we can now write the maximization problem as

$$\max_{y_B} \int_{\tilde{t}}^{a+1} (2t-1-a)y_B(t)(1-\Lambda_B(1-y_B(t))) - (2y_B(t)+a)(a+1-t)(1+\Lambda_S(t-a))y'_B(t) dt.$$

Step 8. We now have a problem of optimal control

$$\max_u \int_{\tilde{t}}^1 F(x, u, t) dt$$

with state variable $x(t) = y_B(t)$, control variable $u(t) = y'_B(t) > 0$, initial condition $x(a) = 0$ and a free terminal condition subject to the pure state constraint $x(t) \in [0, 1]$. In particular, this problem is not concave in the state variable and thus standard sufficiency conditions are not applicable. We thus revert to Theorem 4.1 in Sorger (1989) who provides appropriate conditions for the present case. To simplify things, we will ignore the constraint that $x(t) \leq 1$ and verify it later.

The Hamiltonian reads

$$H(x, u, \lambda, t) = F(x, u, t) + \lambda u$$

and the Lagrangian is given by

$$L(x, u, \lambda, \eta, t) = H(x, u, \lambda, t) + \eta x.$$

Consider the candidate solution with state

$$x^*(t) = \begin{cases} 0 & t \in [a, \tilde{t}) \\ \frac{2a^2\Lambda_S + a(\Lambda_B - 2\Lambda_S t + \Lambda_S - 2) - 2\Lambda_B t + \Lambda_B + 2t - 1}{2(\Lambda_B(a - 2t + 1) + \Lambda_S(-2a + 2t - 1) + 1)} & t \in [\tilde{t}, a+1] \end{cases}$$

control $u^*(t) = x^{*'}(t)$, adjoint

$$\lambda^*(t) = \begin{cases} -a(a-t+1)(a\Lambda_S - \Lambda_S t - 1) & t \in [a, \tilde{t}) \\ \frac{(a\Lambda_B + \Lambda_B - 1)(a-2t+1)(a-t+1)(a\Lambda_S - \Lambda_S t - 1)}{-\Lambda_B(a-2t+1) + 2a\Lambda_S - 2\Lambda_S t + \Lambda_S - 1} & t \in [\tilde{t}, a+1] \end{cases}$$

and multiplier

$$\eta(t) = \begin{cases} -2a^2\Lambda_S - a(\Lambda_B - 2\Lambda_S t + \Lambda_S - 2) + (\Lambda_B - 1)(2t - 1) & t \in [a, \tilde{t}) \\ 0 & t \in [\tilde{t}, a + 1] \end{cases}$$

where

$$\tilde{t} = \frac{-1 + \Lambda_B + 2a^2\Lambda_S + a(\Lambda_B + \Lambda_S - 2)}{2(-1 + \Lambda_B + a\Lambda_S)}.$$

First, note that the Theorem is indeed applicable, that is, (H1)-(H3) hold. For condition (a), note that Q is a real-valued function in our case and thus trivially constitutes a symmetric matrix and that there are no points of discontinuity for the functions λ and Q . Further, one can directly verify that conditions (b) and (c) hold by plugging in the relevant expressions. To see that the weak concavity condition (d) holds, some more work is needed. The Hessian of the Lagrangian is given by

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} 2\Lambda_B(2t - 1 - a) & -2(a + 1 - t)(1 + \Lambda_S(t - a)) \\ -2(a + 1 - t)(1 + \Lambda_S(t - a)) & 0 \end{pmatrix}.$$

Moreover, $H_{x\lambda}(t) = H_{\lambda x}(t) = 0$, $H_{u\lambda}(t) = H_{\lambda u}(t) = 1$ so that the set \mathcal{M} (notice the typo in Sorger (1989), which mistakenly refers to some set \mathcal{U}) consists of a single matrix given by

$$M = \begin{pmatrix} Q'(t) + 2\Lambda_B(2t - 1 - a) & Q(t) - 2(a + 1 - t)(1 + \Lambda_S(t - a)) \\ Q(t) - 2(a + 1 - t)(1 + \Lambda_S(t - a)) & 0 \end{pmatrix}.$$

This matrix is non-positive definite if and only if its determinant is non-negative and if $Q'(t) + 2\Lambda_B(2t - 1 - a) \leq 0$. The determinant is non-negative if and only if $Q(t) = 2(a + 1 - t)(1 + \Lambda_S(t - a))$, implying $Q'(t) = 2(-1 + \Lambda_S - 2\Lambda_S t + 2\Lambda_S a)$. Note that M is diagonal and thus non-positive definite if and only if its first entry is weakly negative, that is, if and only if for all $t \in [a, a + 1]$ we have

$$g(t) = -1 + \Lambda_S - 2\Lambda_S t + 2\Lambda_S a + \Lambda_B(2t - 1 - a) \leq 0.$$

Given that g is linear in t it suffices to verify

$$g(a) = -1 + \Lambda_S + \Lambda_B(a - 1) \leq 0$$

and

$$g(a + 1) = -1 - \Lambda_S + \Lambda_B(a + 1) \leq 0.$$

For the result in Proposition 6 we have $a = 0$ and thus both inequalities are satisfied for $\Lambda_B, \Lambda_S \leq 1$. For the statement in the discussion following the result including footnote 24, we need to consider $a > 0$ and $\Lambda_B \leq 1/(1+a)$ and $\Lambda_S \leq \min\{1, (1 - \Lambda_B(1+a))/a\}$. We have

$$\begin{aligned} g(a+1) &= -1 - \Lambda_S + \Lambda_B(a+1) \\ &\leq -1 - \Lambda_S + 1 \\ &\leq 0, \end{aligned}$$

because $\Lambda_B \leq 1/(1+a)$. Further, for $a \leq 1$ we obviously have $g(a) \leq a$ because $\Lambda_S \leq 1$. For $a > 1$ $\Lambda_B \leq 1/(1+a)$ and $\Lambda_S \leq \min\{1, (1 - \Lambda_B(1+a))/a\}$ imply

$$\begin{aligned} g(a) &\leq -1 + \frac{1 - \Lambda_B(1+a)}{a} + \Lambda_B(a-1) \\ &\leq -1 + \frac{a-1}{a+1} \\ &\leq 0. \end{aligned}$$

Thus, condition (d) of Theorem 4.1 in Sorger (1989) is also satisfied. To see that the complementary slackness condition (e) is satisfied, note that it is linear on in t on $[a, \tilde{t})$ and that $\eta(\tilde{t}) = 0$. Further,

$$\begin{aligned} \eta(a) &= 1 + \Lambda_B(a-1) - a\Lambda_S \\ &\geq 1 + \Lambda_B(a-1) - a \min\{1, (1 - \Lambda_B(1+a))/a\} \\ &\geq 1 + \Lambda_B(a-1)(1 - \Lambda_B(1+a)) = 2a\Lambda_B \geq 0. \end{aligned}$$

To verify the transversality condition (f) let $\gamma = 0$ which then immediately implies that (f1)-(f3) are satisfied since $\lambda(a+1) = Q(a+1) = 0$ since the set of matrices N_1 is simply the scalar 0 and the set of matrices N_2 the scalar 1. Finally, since there are no jumps, condition (g) is trivially satisfied. Given that the constraint $x(t) \leq 1$ which we ignored is satisfied and all conditions from Theorem 4.1 in Sorger (1989) are satisfied, it follows that our candidate compromises a global maximizer. *Step 9.* One can easily verify that the IR constraints are satisfied. *Step 10.* What remains to be done is to recover the ex-post allocation rule. It is straightforward to obtain

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

where

$$\delta^{RM}(\theta_B) = \begin{cases} 0 & \theta_B \leq \tilde{t} \\ \frac{2a^2\Lambda_S + a(\Lambda_B - 2\Lambda_S t + \Lambda_S - 2) - 2\Lambda_B t + \Lambda_B + 2t - 1}{2(\Lambda_B(a - 2t + 1) + \Lambda_S(-2a + 2t - 1) + 1)} & \theta_B > \tilde{t}. \end{cases}$$

and in the case of Proposition 6 with $a = 0$ this yields

$$\delta^{RM}(\theta_B) = \begin{cases} 0 & \theta_B \leq 1/2 \\ \frac{(2\theta_B - 1)(\Lambda_B - 1)}{2(\Lambda_B(2\theta_B - 1) - 2\Lambda_S\theta_B + \Lambda_S - 1)} & \theta_B > 1/2. \end{cases}$$

A.3 Maximizing the Gains from Trade

The derivations of the mechanisms maximizing the total and the material gains from trade proceed analogously. We here present the derivations for the case of maximizing the total gains from trade.

Step 1. We consider the problem of maximizing the total gains from trade. The analysis for the problem of maximizing the material gains from trade is analogous. We first rewrite the problem as a function of the trade rule only. We can rewrite the objective function to (imposing $\Lambda_B = \Lambda_S = \Lambda$)

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

Note that by the budget constraint (AB) we have

$$\int_a^b \bar{t}_B(\theta_B) dF_B(\theta_B) = \int_a^b \bar{t}_S(\theta_S) dF_S(\theta_S).$$

Further, $w_B(\theta_B)$ and $w_S(\theta_S)$ enter the objective positively and both are negative. Hence, we optimally set both terms to zero by choosing interim deterministic transfers. We assume that types are uniformly distributed on $[0, 1]$. This yields

$$\begin{aligned} & \int_0^1 U_B(\theta_B, s_B^t | \theta_B) d\theta_B + \int_0^1 U_S(\theta_S, s_S^t | \theta_S) d\theta_S \\ &= \int_0^1 \theta_B y_B(\theta_B)(1 + \Lambda(y_B(\theta_B) - 1)) d\theta_B - \int_0^1 \theta_S y_S(\theta_S)(1 - \Lambda(y_S(\theta_S) - 1)) d\theta_S. \end{aligned}$$

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

$$\begin{aligned} & \int_0^1 (2\theta_B - 1)y_B(\theta_B) (1 - \Lambda (1 - y_B(\theta_B))) d\theta_B \\ &= \int_0^1 2\theta_S y_S(\theta_S) (1 + \Lambda (1 - y_S(\theta_S))) d\theta_S, \end{aligned}$$

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

$$\begin{aligned} \mathcal{L}(y^f, \gamma) &= \int_0^1 (\theta_B + \gamma(2\theta_B - 1))y_B(\theta_B) (1 - \Lambda (1 - y_B(\theta_B))) d\theta_B \\ &\quad - \int_0^1 (\theta_S + 2\gamma\theta_S)y_S(\theta_S) (1 + \Lambda (1 - y_S(\theta_S))) d\theta_S. \end{aligned}$$

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, $(\theta_B + \gamma(2\theta_B - 1))$ and $(\theta_S + 2\gamma\theta_S)$ are strictly increasing in θ_B and θ_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. We can mirror the derivations for the revenue-maximizing mechanism up and including to Step 7, leaving us with the Hamiltonian given by

$$\begin{aligned} H(x, u, \lambda, t) &= (t + \gamma(2t - 1))x(t)(1 - \Lambda(1 - x(t))) - (x(t) + 2\gamma x(t))(1 - t)(1 + \Lambda t)u(t) + \lambda(t)u(t) \end{aligned}$$

We can use the Euler-Lagrange equation in order to determine our candidate solution and pin down the Lagrange multiplier γ , yielding $\gamma = (1 + \Lambda)/4$. We can now once more apply Theorem 4.1 in Sorger (1989). We ignore the constraints in $x(t)$ and verify them later. The candidate optimal path and control are given by

$$x^*(t) = \frac{((2t - 1)\Lambda + 6t - 1)(1 - \Lambda)}{2(3 - \Lambda)},$$

and $u^*(t) = x^{*'}(t)$. The adjoint is defined by

$$\lambda(t) = -\frac{(\Lambda^2 + 2\Lambda - 3)(t - 1)(\Lambda^2 t(2t - 1) + \Lambda(6t^2 + t - 1) + 6t - 1)}{4(\Lambda - 3)}.$$

Step 4. As noted above, the derivations for the maximization of the material gains from

trade proceeds in exactly the same way. The optimal trade rules are then given by

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

where

$$\delta^{TG}(\theta_B) = \begin{cases} 0 & \theta_B \leq (1 + \Lambda)/(6 + 2\Lambda) \\ \frac{((2\theta_B - 1)\Lambda + 6\theta_B - 1)(1 - \Lambda)}{2(3 - \Lambda)} & o.w. \end{cases}$$

and

$$\delta^{MG}(\theta_B) = \begin{cases} 0 & \theta_B \leq (1 + \Lambda)/(6 - 2\Lambda) \\ \frac{(1 + \Lambda - 6\theta_B + 2\Lambda\theta_B)(1 - \Lambda)}{6\Lambda - 6 - 4\Lambda^2} & o.w. \end{cases}$$

References

- ABELER, J., A. FALK, L. GOETTE, AND D. HUFFMAN (2011): “Reference Points and Effort Provision,” *American Economic Review*, 101, 470–492.
- APESTEGUIA, J. AND M. BALLESTER (2015): “A Measure of Rationality and Welfare,” *Journal of Political Economy*, 123, 1278–1310.
- ARROW, K. (1968): “Applications of Control Theory to Economic Growth,” Providence, R. I.: American Mathematical Society, Mathematics of the Decision Sciences.
- 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.
- BENKERT, J.-M. (2022): “On the equivalence of optimal mechanisms with loss and disappointment aversion,” Mimeo.
- BENKERT, J.-M. AND N. NETZER (2018): “Informational Requirements of Nudging,” *Journal of Political Economy*, 126, 2323–2355.
- BERNHEIM, B. AND A. RANGEL (2009): “Beyond Revealed Preference: Choice-Theoretic Foundations For Behavioral Welfare Economics,” *Quarterly Journal of Economics*, 124, 51–104.

- 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.
- 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.
- CRAMTON, P., R. GIBBONS, AND P. KLEMPERER (1987): “Dissolving a partnership Efficiently,” *Econometrica*, 55, 615–632.
- CRAWFORD, V. P. (2021): “Efficient Mechanisms for Level- k Bilateral Trading,” *Games and Economic Behavior*, 127, 80–101.
- 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.
- DATO, S., A. GRUNEWALD, D. MÜLLER, AND P. STRACK (2017): “xpectation-basedlossaversionandstrategicinteraction,” *Games and Economic Behavior*, 104, 681–705.
- DE MEZA, D. AND D. C. WEBB (2007): “Incentive Design under Loss Aversion,” *Journal of the European Economic Association*, 5, 66–92.
- 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.
- (2018): “Mechanism Design with News Utility,” Mimeo.
- EISENHUTH, R. (2019): “Reference-Dependent Mechanism Design,” *Economic Theory Bulletin*, 7, 77–103.
- EISENHUTH, R. AND M. GRUNEWALD (2018): “Auctions with Loss Averse Bidders,” *International Journal of Economic Theory*, 16, 129–152.

- 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 (2020): “Efficient Bilateral Trade,” Mimeo, UCSB and UZH.
- GILL, D. AND V. PROWSE (2012): “A Structural Analysis of Disappointment Aversion in a Real Effort Competition,” *American Economic Review*, 102, 469–503.
- GNEEZY, U., L. GOETTE, C. SPRENGER, AND F. ZIMMERMANN (2017): “The Limits of Expectations-Based Reference Dependence,” *Journal of the European Economic Association*, 15, 861–876.
- HEFFETZ, O. AND J. A. LIST (2014): “Is the Endowment Effect an Expectations Effect?” *Journal of the European Economic Association*, 12, 1396–1422.
- HEIDHUES, P. AND B. KŐSZEGI (2014): “Regular Prices and Sales,” *Theoretical Economics*, 9, 217–251.
- 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. MÖLLER (2020): “Selling in Advance to Loss Averse Consumers,” *International Economic Review*, 61, 441–468.
- KARLE, H. AND M. PEITZ (2014): “Competition under consumer loss aversion,” *The RAND Journal of Economics*, 45, 1–31.

- KARLE, H. AND H. SCHUMACHER (2017): “Advertising and Attachment: Exploiting Loss Aversion through Pre-Purchase Information,” *RAND Journal of Economics*, 48, 875–1135.
- 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.
- MANZINI, P. AND M. MARIOTTI (2014): “Welfare economics and bounded rationality: the case for model-based approaches,” *Journal of Economic Methodology*, 21, 343–360.
- MASATLIOGLU, Y. AND C. RAYMOND (2016): “A Behavioral Analysis of Stochastic Reference Dependence,” *The American Economic Review*, 106, 2760–2782.
- 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. (2016): “Selling Substitute Goods to Loss-Averse Consumers: Limited Availability, Bargains and Rip-offs,” *Rand Journal of Economics*.
- (2017): “Sequential Negotiations with Loss-Averse Buyers,” *European Economic Review*, 91, 290–304.
- (2021): “Loss aversion in sequential auctions,” Working paper, University of Technology Sydney.
- RUBINSTEIN, A. AND Y. SALANT (2012): “Eliciting Welfare Preferences from Behavioural Data Sets,” *Review of Economic Studies*, 79, 375–387.
- SEIERSTAD, A. AND K. SYDSAETER (1987): *Optimal Control Theory with Economic Applications*, Elsevier.
- SHALEV, J. (2002): “Loss Aversion and Bargaining,” *Theory and Decisions*, 52, 201–232.
- SORGER, G. (1989): “Sufficient Optimality Conditions for Nonconvex Control Problems with State Constraints,” *Journal of Optimization Theory and Applications*.
- 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.