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 (2015) : 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/10419/111245

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

March 2015

Bilateral Trade with Loss-Averse Agents

Jean-Michel Benkert*

This version: March 2015
First version: November 2014

Abstract

We study the bilateral trade problem put forward by Myerson and Satterthwaite (1983) under the assumption that agents are loss-averse. We use the model developed by Kőszegi and Rabin (2006, 2007) to find optimal mechanisms for the minimal subsidy, revenue maximization and welfare maximization problem. In both, welfare and revenue maximizing mechanisms, the designer induces less trade in the presence of loss-aversion. Intuitively, the designer is providing the agents with partial insurance. Moreover, the designer optimally provides the agents with full insurance in the money dimension, i.e. she offers deterministic transfers. Another implication of loss-aversion is that it increases the severity of the impossibility problem, that is, the minimal subsidy needed to induce materially efficient trade is higher. All results display robustness to the exact specification of the reference point. We also provide some general mechanism design results.

Keywords: Bilateral Trade, Loss-Aversion, Mechanism Design, Deterministic Transfers
JEL Classification: C78, D02, D03, D82, D84

*University of Zurich, Department of Economics, Bluemlisalpstrasse 10, CH-8006 Zurich, Switzerland and UBS International Center of Economics in Society at the University of Zurich. Email: jean-michel.benkert@econ.uzh.ch. I would like to thank Olivier Bochet, Juan Carlos Carbajal, Samuel Haefner, Fabian Herweg, Heiko Karle, Igor Letina, Shou Liu, Georg Nöldeke, Yuval Salant, Ran Spiegler, Tom Wilkening and seminar particpants in Zurich and at the ZWE 2014 for helpful comments. I am especially grateful to my supervisor Nick Netzer for his guidance as well as numerous comments and suggestions. I would like to thank the Faculty of Business and Economics at the University of Basel for their hospitality while some of this work was conducted and the UBS International Center of Economics in Society at the University of Zurich for financial support. All errors are my own.

1 Introduction

In many situations people evaluate an outcome relative to some reference point. For instance, whether a house owner is willing to sell her house at some price, may depend on whether or not that price is higher than the original purchase price (Genesove and Mayer, 2001). Relatedly, if a buyer expects a trade to go through, her willingness to pay for the good may increase (Ericson and Fuster, 2011). Evidence suggests that the most relevant type of reference-dependent preferences is loss-aversion (see DellaVigna, 2009, for a survey).[1] Kahneman and Tversky's (1979) prospect theory established the importance and relevance of loss-aversion early on and the literature on this phenomenon has grown substantially since. In this paper, we make use of the model developed by Kőszegi and Rabin (2006, 2007) (henceforth KR) to study the bilateral trade problem put forward by Myerson and Satterthwaite (1983) (henceforth MS) under the assumption that agents are loss-averse and form the reference point endogenously as the expectation over the outcome.[2]

In the bilateral trade problem a privately informed seller wants to sell one unit of an indivisible good to a privately informed buyer. The problem has become a corner stone of mechanism design theory and the central result in this context is the famous impossibility theorem by MS: it is generally impossible to have an individually rational, incentive compatible and budget balanced mechanism, in which trade takes place whenever it is materially efficient, i.e., whenever the buyer values the object more than the seller. In particular, MS calculate the minimal subsidy needed to induce materially efficient trade under incentive compatibility and individual rationality. We take this as the starting point of our analysis and pose this minimal subsidy problem when agents are loss-averse. In doing so, we employ the choice acclimating personal equilibrium (CPE) introduced by KR as our equilibrium concept. The reference point is thus formed endogenously as the expectation over the outcome, assuming that outcomes are resolved sufficiently long after decisions are made. Further, we let agents bracket narrowly, that is, they have separate gain-loss utility for different dimensions of utility, such as trade and money utility. We find that the presence of loss-aversion increases the minimal subsidy needed to induce materially efficient trade under incentive compatibility and individual rationality. Put differently, the impossibility problem becomes more severe.

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

[2] Ericson and Fuster (2011), Abeler, Falk, Goette, and Huffman (2011), Gill and Prowse (2012) and Karle, Kirchsteiger, and Peitz (forthcoming) provide evidence for the assumption that the reference point is determined by expectations in a series of experiments. In particular, the findings in Gill and Prowse (2012) suggest that the reference point is formed very quickly after an action is taken. This validates our decision to employ the choice acclimating personal equilibrium as our equilibrium concept in this paper. See Section 3.2 for more details.

Having confirmed the impossibility result in the presence of loss-aversion, we focus on the class of δ-inefficient mechanisms for the remainder of the paper. In a δ-inefficient mechanism trade takes place whenever the buyer's valuation exceeds the seller's valuation by some non-negative δ. In this natural class of trade mechanisms, which in fact contains the optimal mechanisms in MS, we consider the problem of maximizing expected revenue. As in the classic framework without loss-aversion, the designer optimally induces less trade than materially efficient, setting $\delta > 0$. In fact, the presence of loss-aversion reduces the trade frequency even more. Considering the case of a welfare maximizing designer, the picture looks the same: loss-aversion leads to less trade.

The intuition for this reduction in the trade frequency is that the designer is providing agents with partial insurance against ex post variation in payoffs in the trade dimension. By reducing the likelihood of trade, the designer affects the agents' reference points, thereby diminishing feelings of loss while increasing feelings of gains and, overall, reducing the disutility caused by ex post variation in payoffs. Thus, agents are partially insured against variations in payoffs in the trade dimension, which increases their utility as they are loss-averse. Full insurance, i.e., a complete elimination of any variation in trade utility, would correspond to trade taking place either always or never, which is not optimal in general. In contrast, the designer optimally provides the agents with full insurance in the money dimension, i.e., the transfers are deterministic in the sense that they do not depend on the other agent's type. Deterministic transfers also constitute *a* solution in the framework of MS, whereas they are the *only* solution in the presence of loss-aversion, meaning that transfers are deterministic in any optimal mechanism. Because agents are loss-averse, this reduction, or rather, elimination of variation in transfers increases the agents' utility, allowing the designer to extract more money from them or to make them better off, depending on her objective.

It turns out that this insurance property found in the optimal mechanisms is robust to the precise specification of the reference point. KR note that their equilibrium concept CPE is similar to the disappointment-aversion models introduced by Bell (1985), Loomes and Sugden (1986) and Gul (1991). The CPE specifies the reference point as the full distribution of a lottery, whereas the reference point corresponds to the certainty equivalent of the lottery in these models of disappointment-aversion. Masatlioglu and Raymond (2014) find that the intersection of preferences induced by the CPE and any of the listed disappointment-aversion models is simply expected utility. Thus, although the models seem to be very similar, the induced preferences are generally not the same. Nevertheless, one can show that the optimal mechanisms derived in this paper for CPE are also optimal for the models by Bell (1985) and Loomes and Sugden (1986).[3]

[3]The recursive formulation in Gul (1991) makes the analysis harder, but we would expect to find similar results for that case, too.

The insurance property is also consistent with the findings in Herweg, Müller, and Weinschenk (2010) as well as Eisenhuth (2013). Herweg et al. (2010) consider a modification of the principal-agent model with moral hazard. In a framework where standard preferences would predict a fully contingent contract, they find that, using the CPE by KR, the optimal contract is a binary payment scheme. This reduction in the ex post variation of the payment is analogous to the deterministic transfers in our optimal mechanisms. Also, the cost of implementing some actions is increasing in the degree of loss-aversion, mirroring corresponding results in the present paper. Eisenhuth (2013), whose work we follow in some methodical aspects, considers the design of an optimal auction using the CPE as the equilibrium concept. He finds that the optimal auction is an all-pay auction with reserve price when agents bracket narrowly, and that it is a first-price auction with reserve price in the case of wide bracketing. Further, he establishes that in any revenue maximizing mechanism, transfers are deterministic when agents bracket narrowly. This is, of course, perfectly in line with our findings. Besides considering a bilateral trade setting, the present paper also extends the result in Eisenhuth (2013) regarding deterministic transfers from revenue maximizing mechanisms to welfare maximizing mechanisms.

The appendix contains a section on general mechanisms (i.e., not limited to bilateral trade) with loss-averse agents and presents a generalization of the part with narrow bracketing in Eisenhuth (2013). The bilateral trade model considered in the main text is a special case of the model introduced in the appendix. Besides adding some generality, this allows us to state some of the results in the main text as corollaries and show that deterministic transfers being part of the solution is not due to the bilateral trade setting, but a general feature of any welfare maximizing mechanism.

This paper is organized as follows. The next section contains a more detailed discussion of the related literature. In Section 3 we present the model, solution concept and notation used throughout the paper. In Section 4 we address the minimal subsidy problem and consider the impossibility theorem. Section 5 contains the derivation of the revenue and welfare maximizing trade mechanisms. In Section 6 we show that these optimal mechanisms are robust to the exact specification of the reference point a discuss some of our assumptions. Section 7 concludes.

2 Related literature

As mentioned above, the work by Eisenhuth (2013) is closest to the present paper, but focuses on revenue maximizing mechanisms for the allocation of an object among agents who are loss-averse in the sense of KR. Besides, our paper contributes to three strands of literature: the literature on bilateral trade, theoretical applications of reference-dependent utility and behavioral mechanism design.

Cavallo (2011) considers a bilateral trade setting but differs in methodology and objective. He considers three different mechanisms in the bilateral trade setting with risk-neutral agents. He looks at what he calls a "surplus extracting", a "surplus sharing" and a "risk-minimizing" mechanism, using a non-standard ex ante individual rationality concept, which allows him to circumvent the impossibility problem. He then turns to a numerical evaluation of these three mechanisms when agents exhibit loss-aversion. Thus, while Cavallo (2011) also considers a bilateral trade setting, the focus of the paper is very different and, in particular, he does not consider the design of optimal mechanisms.

Garratt and Pycia (2014) examine the bilateral trade problem relaxing the assumption that utilities are quasi-linear. Instead, they assume that the good is normal. They show that given these assumptions ex post efficient trade is possible under some conditions. The impossibility result can be reversed in this setting because surplus is not only generated by allocating the good to the agent who values it most. Instead, the normality of the good provides another way to create surplus, which can then be used to induce incentive compatibility. Garratt and Pycia (2014), however, are not the first to consider the bilateral trade problem under the assumption of risk-averse agents. Copic and Ponsatí (2008) study the bilateral trade problem from an ex post perspective and are interested in robust mechanism design under incomplete information on traders' private reservation values, when agents are risk-neutral and risk-averse. Even earlier, Chatterjee and Samuelson (1983) extend their analysis of the double-auction to the case of risk-averse agents. They find that a sufficiently high degree of constant relative risk-aversion (in the words of Chatterjee and Samuelson (1983, 848), when agents "become infinitely risk-averse") induces an ex post efficient outcome in equilibrium.

Salant and Siegel (forthcoming) study the efficient allocation of a divisible asset for different types of reallocation costs. In a first period agents divide an asset between them, having not yet learned their valuations for the asset. In a second period, after uncertainty is resolved, the agents may reallocate the object at some cost. When the asset is fully assigned to one of the two agents in the first period, the second period corresponds to the setting in MS. Salant and Siegel (forthcoming) then consider concave reallocation costs, which allows for the following interpretation. Let the first period share of the asset be the agents' reference point with respect to which they feel losses but no gains. Thus, in contrast to the present setting the reference point is not the agents' endogenously determined expectation about the final outcome, but the fixed initial share of the asset, and the agents do not feel gains but only losses. Salant and Siegel (forthcoming) then show that the impossibility result holds in their setting. Thus, some of our results with an endogenous reference point extend to the case of a fixed reference point.

There are numerous theoretical papers working under the assumption of loss-averse agents, while applying it to different settings. To name but a few in addition to those al-

ready mentioned, de Meza and Webb (2007) consider incentive design under loss-aversion, Gill and Stone (2010) model a two-player rank-order tournament when agents are loss-averse, Carbajal and Ely (2014) study optimal price discrimination when a monopolist faces a continuum of consumers with reference-dependent preferences and Karle and Peitz (2014) investigate firm strategy in imperfect competition.

The paper is also related to the literature on behavioral mechanism design. As Kőszegi (2014) notes in his recent survey, little work has been done in this direction so far. Bierbrauer and Netzer (2014) modify the standard mechanism design framework by introducing intention-based social preferences. Like the reference point in the present paper, intentions are determined endogenously in their paper. Three of their findings are of particular interest to us. First, they also have an insurance property in the sense that the expected payoff of one agent does not depend on the other agent's type. Second, they consider the bilateral trade problem as an application and find that the impossibility result by MS is turned into a possibility result. Third, they introduce mechanisms that are robust with respect to the existence of social preferences. That is, these robust mechanisms implement an economic outcome irrespective of whether or not agents have social preferences. Thus, their notion of (behavioral) robustness is in the same spirit as the robustness to the exact specification of the reference point we find in our mechanisms. Kucuksenel (2012) considers the mechanism design problem under the assumption that agents are altruistic. He also considers the bilateral trade problem as an application and similar to Bierbrauer and Netzer (2014) he finds that the more altruistic agents are, the higher the probability of efficient trade taking place.

3 Model

3.1 Utility, Social Choice Functions and Mechanisms

The set of agents is given by $I = \{S, B\}$ where S and B denote seller and buyer, respectively. It is commonly known that the type of agent $i \in I$ is uniformly drawn from the set $\Theta_i = [0, 1] \subseteq \mathbb{R}$ and is private information. Let $\Theta = \Theta_S \times \Theta_B$. We interpret the type of the buyer as her valuation of the good to be traded and the type of the seller as her cost of producing this good.

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, riskless total utility is given by

$$u_S(\mathbf{x}, \mathbf{r}_S, \theta_S) = -y\theta_S + t_S + \eta^1\mu^1(r_S^1\theta_S - y\theta_S) + \eta^2\mu^2(t_S - r_S^2) \tag{1}$$

$$u_B(\mathbf{x}, \mathbf{r}_B, \theta_B) = y\theta_B - t_B + \eta^1\mu^1(y\theta_B - r_B^1\theta_B) + \eta^2\mu^2(r_B^2 - t_B) \tag{2}$$

where $\eta^k \geq 0$, $k = 1, 2$. The functions μ^k, $k = 1, 2$, are value functions in the sense of Kahneman and Tversky (1979) exhibiting loss-aversion with

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

for some $\lambda^k > 1$, $k = 1, 2$ but ignoring diminishing sensitivity. We can interpret the parameters η^k and λ^k, $k = 1, 2$, as the weights put on gain-loss utility and the degrees of loss-aversion, respectively. The parameter $\mathbf{r}_i = \{r_i^1, r_i^2\}$, $i = S, B$, is the riskless reference level. Following KR we will allow the reference point to be the agent's rational expectations and therefore a probability distribution over all riskless reference levels. We will refer to $-y\theta_S + t_S$ and $y\theta_B - t_B$ as material utility and to the other terms as gain-loss utility. We adopt the following assumption from Herweg et al. (2010):

Assumption 1 (No Dominance of Gain-Loss Utility) $\Lambda = \eta^1(\lambda^1 - 1) \leq 1$.

This assumption ensures that gain-loss utility does not dominate material utility and will be crucial for incentive compatibility. The utility specification in (1) and (2) corresponds to "narrow bracketing", meaning that for the two material utility dimensions, trade and money utility, there is a separate gain-loss term each. Thus, gain-loss feelings are bracketed narrowly and not widely.[4]

A social choice function (SCF) $f : \Theta \to X$ 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 $g : M_S \times M_B \to X$. 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 $s_i : \Theta_i \to M_i$. 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.

[4]See, for instance, Heidhues and Kőszegi (2014, p. 224) for a brief discussion of the narrow bracketing assumption.

3.2 Equilibrium Concept and Revelation Principle

We use the concept of a choice-acclimating personal equilibrium (CPE) introduced by Kőszegi and Rabin (2006, 2007). As mentioned earlier, the reference point, or rather lottery, of an agent will be equal to her rational expectations about the eventual outcome. When contemplating what action to take in a bilateral trade situation, it seems natural for agents to form an expectation over the consequences of an action and base their decision on this assessment. Thus, the notion of a CPE is appropriate, as there is enough time to form beliefs over the outcome between the moment in which an agent takes an action and the realization of payoffs. Put differently, the outcome is only realized after beliefs and choice have had enough time to adjust.[5] This makes expectations a natural reference point and the CPE the appropriate equilibrium concept. We discuss the alternative equilibrium concept introduced by KR, the "unacclimating personal equilibrium" (UPE), in the conclusion.

The set of all possible riskless reference levels is given by the set of all social alternatives, X. Thus, an agent compares the eventual outcome to what could have happened. As mentioned above, we allow for the reference point to be a distribution over the set X. In a mechanism Γ, this distribution is induced endogenously for each agent: conditional on the other agent playing s_{-i}, agent i induces a distribution over the set of social alternatives, X, by playing the strategy s_i. Hence, the loss-averse agent will compare any given social alternative to all possible social alternatives, allowing for gain or loss feelings in every comparison.

Moving to the interim stage and allowing 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 strategy s_S, given that the buyer plays s_B as

$$
\begin{aligned}
U_S(s_S(\theta_S),&s_B,\Gamma|\theta_S) = \\
&\int_0^1 -y^g(s_S(\theta_S),s_B(\theta_B))\theta_S + t_S^g(s_S(\theta_S),s_B(\theta_B))\,d\theta_B \\
&+ \int_0^1\int_0^1 \eta^1\mu^1\left(y^g(s_S(\theta_S),s_B(\theta'_B))\theta_S - y^g(s_S(\theta_S),s_B(\theta_B))\theta_S\right)\,d\theta'_B\,d\theta_B \qquad (3)\\
&+ \int_0^1\int_0^1 \eta^2\mu^2\left(t^g(s_S(\theta_S),s_B(\theta_B)) - t^g(s_S(\theta_S),s_B(\theta'_B))\right)\,d\theta'_B\,d\theta_B \\
&= -\theta_S\int_0^1 y^g(s_S(\theta_S),s_B(\theta_B))\,d\theta_B + \int_0^1 t_S^g(s_S(\theta_S),s_B(\theta_B))\,d\theta_B \\
&+ \theta_S\eta^1\int_0^1\int_0^1 \mu^1\left(y^g(s_S(\theta_S),s_B(\theta'_B)) - y^g(s_S(\theta_S),s_B(\theta_B))\right)\,d\theta'_B\,d\theta_B
\end{aligned}
$$

[5] Gill and Prowse (2012) conduct a lab experiment and find that a subject who competes in a sequential move contest adjusts her reference point "essentially instantaneously to her own effort choice and to that of her competitor", suggesting that there is indeed enough time for this adjustment.

$$
\begin{aligned}
&+ \eta^2 \int_0^1 \int_0^1 \mu^2 \left(t_S^g(s_S(\theta_S), s_B(\theta_B)) - t_S^g(s_S(\theta_S), s_B(\theta'_B)) \right) \, d\theta'_B \, d\theta_B \\
&= -\theta_S \tilde{v}_S(\theta_S) + \tilde{t}_S(\theta_S), \qquad (4)
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{v}_S(\theta_S) &= \int_0^1 y^g(s_S(\theta_S), s_B(\theta_B)) \, d\theta_B \\
&\quad - \eta^1 \int_0^1 \int_0^1 \mu^1 \left(y^g(s_S(\theta_S), s_B(\theta'_B)) - y^g(s_S(\theta_S), s_B(\theta_B)) \right) \, d\theta'_B \, d\theta_B, \\
\tilde{t}_S(\theta_S) &= \int_0^1 t_S^g(s_S(\theta_S), s_B(\theta_B)) \, d\theta_B \\
&\quad + \eta^2 \int_0^1 \int_0^1 \mu^2 \left(t_S^g(s_S(\theta_S), s_B(\theta_B)) - t_S^g(s_S(\theta_S), s_B(\theta'_B)) \right) \, d\theta'_B \, d\theta_B.
\end{aligned}
$$

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 perhaps unfamiliar looking double integral has a clear intuition. To illustrate, consider the third 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 $t^g(s_S(\theta_S), s_B(\theta_B))$, which she would compare to the reference point. The reference point, or rather distribution, is induced endogenously and corresponds to the distribution of possible transfers. Thus, for every θ'_B in the domain of the inner integral we get a possible transfer $t^g(s_S(\theta_S), s_B(\theta'_B))$ given the strategies and the seller's type. The seller compares the actual transfer $t^g(s_S(\theta_S), s_B(\theta_B))$ with all these other possible transfers and the value function μ^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 (uniform) probability. 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 he compares the outcome with all other possible outcomes, aggregating gains and losses in each comparison.

The compact notation in (4) 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

$$
\begin{aligned}
\bar{t}_S(\theta_S) &= \int_0^1 t_S^g(s_S(\theta_S), s_B(\theta_B)) \, d\theta_B, \\
w_S(\theta_S) &= \int_0^1 \int_0^1 \mu^2 \left(t_S^g(s_S(\theta_S), s_B(\theta_B)) - t_S^g(s_S(\theta_S), s_B(\theta'_B)) \right) \, d\theta'_B \, d\theta_B,
\end{aligned}
$$

allowing us to write $\tilde{t}_S(\theta_S) = \bar{t}_S(\theta_S) + \eta^2 w_S(\theta_S)$. Similarly, we can write the buyer's utility as $U_B(s_B(\theta_B), s_S, \Gamma|\theta_B) = \theta_B \tilde{v}_B(\theta_B) + \tilde{t}_B(\theta_B)$, where

$$\begin{aligned}
\tilde{v}_B(\theta_B) &= \int_0^1 y^g(s_S(\theta_S), s_B(\theta_B))\, d\theta_S \\
&\quad + \eta^1 \int_0^1 \int_0^1 \mu^1 \left(y^g(s_S(\theta_S), s_B(\theta_B)) - y^g(s_S(\theta'_S), s_B(\theta_B))\right)\, d\theta'_S\, d\theta_S, \\
\tilde{t}_B(\theta_B) &= -\int_0^1 t_B^g(s_S(\theta_S), s_B(\theta_B))\, d\theta_S \\
&\quad + \eta^2 \int_0^1 \int_0^1 \mu^2 \left(t_B^g(s_S(\theta'_S), s_B(\theta_B)) - t_B^g(s_S(\theta_S), s_B(\theta_B))\right)\, d\theta'_S\, d\theta_S.
\end{aligned}$$

We also define

$$\begin{aligned}
\bar{t}_B(\theta_B) &= \int_0^1 t_B^g(s_S(\theta_S), s_B(\theta_B))\, d\theta_S, \\
w_B(\theta_B) &= \int_0^1 \int_0^1 \mu^2 \left(t_B^g(s_S(\theta'_S), s_B(\theta_B)) - t_B^g(s_S(\theta_S), s_B(\theta_B))\right)\, d\theta'_S\, d\theta_S,
\end{aligned}$$

allowing us to write $\tilde{t}_B(\theta_B) = -\bar{t}_B(\theta_B) + \eta^2 w_B(\theta_B)$.

In an interim CPE, an agent behaving optimally anticipates that her action will not only affect the eventual outcome, but also determine her reference point. We can now define our equilibrium concept, which follows Eisenhuth (2013).

Definition 1 *A strategy profile $s^* = (s_S^*, s_B^*)$ is an interim CPE of the mechanism $\Gamma = (M_S, M_B, g)$ if for all $i \in I$ and $\theta_i \in \Theta_i$,*

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

Definition 2 *A mechanism Γ implements a social choice function, f, in CPE if there is a CPE strategy profile $s = (s_S, s_B)$ of Γ 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 social choice function f is CPE incentive compatibility (CPEIC) if the truthful profile $s^t = (s_S^t, s_B^t)$ is a CPE strategy in the direct mechanism Γ^d.*

Further, the revelation principle for CPE holds, which will allow us to economize notation considerably.

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 bilateral trade model introduced in this section is a special case of the more general model presented in Appendix A. Therefore, all results here are corollaries and the proofs are omitted. Henceforth, we focus on direct mechanisms and no longer explicitly list the mechanism as an argument in the utility function.

3.3 Incentive Compatibility and Efficiency

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

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

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

(ii) we can write utility as

$$U_S(\theta_S, s_B^t|\theta_S) = U_S(1, s_B^t|1) + \int_{\theta_S}^{1} \tilde{v}_S(t)\, dt, \tag{5}$$

$$U_B(\theta_B, s_S^t|\theta_B) = U_B(0, s_S^t|0) + \int_{0}^{\theta_B} \tilde{v}_B(t)\, dt. \tag{6}$$

We say that a SCF is individually rational if for both agents $i \in I$

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

and that it is ex post budget balanced if

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

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

A trade mechanism is materially efficient[6] if

$$y^f(\theta_S, \theta_B) = \begin{cases} 1 & \text{if } \theta_B > \theta_S \\ 0 & \text{if } \theta_B \leq \theta_S. \end{cases} \tag{ME}$$

In the classic framework with no loss-aversion, material efficiency and budget balance taken together are equivalent to Pareto efficiency. We will thus use these concepts as a benchmark allowing us to analyze the impact of the introduction of loss-aversion in a familiar environment and to draw a clear comparison to the classic framework.

4 The Minimal Subsidy Problem

In this section we address the minimal subsidy problem, by which we mean the following. We consider a designer who wants to induce materially efficient trade under CPEIC and IR at the lowest possible ex ante expected cost. This problem is also of interest in the classic setting with no loss-aversion. The famous impossibility result by MS states that there is no materially efficient SCF satisfying simultaneously CPEIC, IR and BB. Therefore, the natural next question is to ask at what cost a designer can achieve materially efficient trade under CPEIC and IR. The size of the subsidy needed to achieve this goal is an indicator of how big the impossibility problem is. Formally, we consider the problem

$$\min_{(y^f, t_S^f, t_B^f) \in \mathcal{F}} \int_0^1 \int_0^1 \left(t_S^f(\theta_S, \theta_B) - t_B^f(\theta_S, \theta_B) \right) \, d\theta_S \, d\theta_B,$$

subject to IR, CPEIC and ME. (MSP)

Proposition 3 *The minimal subsidy needed to induce materially efficient trade under CPEIC and IR is given by* $(1 + \Lambda)/6$.

Proof. See Appendix B.1. ■

This result has one particular implication: Since the ex ante minimal subsidy is strictly positive, the mechanism cannot be ex post budget balanced. That is, the impossibility result by MS is still valid in the present bilateral trade framework with loss-averse agents. We summarize this as a corollary.

Corollary 1 *There exists no SCF satisfying simultaneously CPEIC, IR, ME and BB.*

Notice that the minimal subsidy is monotonically increasing in Λ, which has an additional implication. The presence of loss-aversion makes it even harder, that is, more expensive, to

[6] Note that the tie-breaking rule $y^f(\theta, \theta) = 0$ is without loss of generality.

induce materially efficient trade under CPEIC and IR. Put differently, the impossibility problem is even bigger in the presence of loss-aversion. The fact that the subsidy is increasing in Λ implies that it is increasing in both, the weight put on gain-loss utility, as well as the degree of loss-aversion.

In the proof of Proposition 3 we find that the transfers leading to the minimal subsidy must satisfy

$$\bar{t}_S(\theta_S) = \frac{1}{2} + \frac{\Lambda}{6} - \frac{\theta_S^2}{2} - \left(\frac{2}{3}\theta_S^3 - \frac{1}{2}\theta_S^2\right)\Lambda,$$
$$\bar{t}_B(\theta_B) = \frac{\theta_B^2}{2} + \left(\frac{2}{3}\theta_B^3 - \frac{1}{2}\theta_B^2\right)\Lambda.$$

Interestingly, deterministic transfers satisfy these conditions and are thus optimal. That is, the transfer of either agent is independent of the type reported by the other agent. Hence, there is some insurance property on the money dimension. Intuitively, by eliminating the variation in the transfers, the designer increases the agents' utility, as they dislike this variation. This is a consequence of the first-order risk-aversion induced by loss-aversion. Thus, eliminating this variation and keeping utility of the agents constant, the designer can reduce the subsidy needed to induce materially efficient trade under CPEIC and IR. As a consequence of the deterministic transfers, the minimal subsidy does not depend on η^2 and λ^2, the loss-aversion parameters associated with the money utility, but only on $\Lambda = \eta^1(\lambda^1 - 1)$.

In the minimal subsidy problem we have fixed the trade rule by requiring materially efficient trade. Thereby, we have ruled out the possibility to provide the agents with some insurance on the trade dimension. In Sections 5.1 and 5.2, when we consider the revenue and the welfare maximization problem, respectively, the designer has this additional degree of freedom and will optimally provide the agents with some insurance in both dimensions.

5 Optimal Mechanisms

5.1 The Revenue Maximization Problem

The preceding section has confirmed the impossibility result in a framework with loss-averse agents. In particular, a designer who wants to ensure materially efficient trade while satisfying CPEIC and IR cannot make a positive profit. The best she can do is an expected loss of $(1 + \Lambda)/6$. A natural question is, thus, whether a materially *in*efficient trade mechanism satisfying incentive compatibility and individual rationality can lead to a positive profit for the designer. This section provides a positive answer to this question.

The revenue-maximizing designer's problem reads

$$\max_{(y^f,t_S^f,t_B^f)\in\mathcal{F}} \int_0^1 \int_0^1 \left(t_B^f(\theta_S,\theta_B) - t_S^f(\theta_S,\theta_B)\right)\, d\theta_S\, d\theta_B,$$

subject to CPEIC and IR. (RM)

We will first rewrite this problem and present an equivalent problem of some interest. We will then proceed by solving a constrained form of the problem (RM).

Proposition 4 *The problem* (RM) *is equivalent to the problem*

$$\begin{aligned}\max_{y^f\in\mathcal{Y}} &\int_0^1 (2\theta_B - 1)\mathbb{E}_S[y^f(\tilde{\theta}_S,\theta_B)]\left(\Lambda\left[\mathbb{E}_S[y^f(\tilde{\theta}_S,\theta_B)] - 1\right] + 1\right)\, d\theta_B \\ &+ \int_0^1 2\theta_S \mathbb{E}_B[y^f(\theta_S,\tilde{\theta}_B)]\left(\Lambda\left[\mathbb{E}_B[y^f(\theta_S,\tilde{\theta}_B)] - 1\right] - 1\right)\, d\theta_S \end{aligned} \quad \text{(RM')}$$

subject to $\tilde{v}_S$ being non-increasing, $\tilde{v}_B$ being non-decreasing and $\tilde{v}_S(1) \geq 0, \tilde{v}_B(0) \geq 0$.

Proof. See Appendix B.2. ■

The term $\mathbb{E}_S[y^f(\tilde{\theta}_S,\theta_B)]$ is the buyer's probability assessment of trade taking place when her value is θ_B, given the trade rule y^f. The analog is true for the seller and $\mathbb{E}_B[y^f(\theta_S,\tilde{\theta}_B)]$. Thus, instead of maximizing over trade rules, y^f, we could maximize directly over conditional trade probabilities, subject to the remaining monotonicity and individual rationality constraints.

Notice also, that we have eliminated transfers from the maximization problem. A key step there is the finding that deterministic transfers are optimal and, therefore, any gain-loss feelings related to money disappear. As an illustration, we present the proof that deterministic transfers for the seller are optimal. The more general result, which states that deterministic transfers are optimal in any revenue or welfare maximizing mechanism, is stated in Lemma 1 in Appendix A. Recall that we defined

$$w_S(\theta_S) = \int_0^1 \int_0^1 \mu^2\left(t_S^f(\theta_S,\theta_B) - t_S^f(\theta_S,\theta'_B)\right)\, d\theta'_B\, d\theta_B.$$

This expression collects all gain-loss feeling of the seller with respect to money and (after some rewriting) it enters the designer's maximization problem positively. We then get

$$\begin{aligned} w_S(\theta_S) &= \int_0^1 \int_0^1 \mu^2\left(t_S^f(\theta_S,\theta_B) - t_S^f(\theta_S,\theta'_B)\right)\, d\theta'_B\, d\theta_B \\ &= \int_0^1 \int_0^1 \left(t_S^f(\theta_S,\theta_B) - t_S^f(\theta_S,\theta'_B)\right) \mathbb{1}[t_S^f(\theta_S,\theta_B) - t_S^f(\theta_S,\theta'_B) > 0]\, d\theta'_B\, d\theta_B \\ &+ \int_0^1 \int_0^1 \lambda^2\left(t_S^f(\theta_S,\theta_B) - t_S^f(\theta_S,\theta'_B)\right) \mathbb{1}[t_S^f(\theta_S,\theta_B) - t_S^f(\theta_S,\theta'_B) < 0]\, d\theta'_B\, d\theta_B \end{aligned}$$

$$
\begin{aligned}
&= \int_0^1 \int_0^1 \left(t_S^f(\theta_S, \theta_B) - t_S^f(\theta_S, \theta'_B) \right) \mathbb{1}[t_S^f(\theta_S, \theta_B) - t_S^f(\theta_S, \theta'_B) > 0] \, d\theta'_B \, d\theta_B \\
&- \lambda^2 \int_0^1 \int_0^1 \left(t_S^f(\theta_S, \theta'_B) - t_S^f(\theta_S, \theta_B) \right) \mathbb{1}[t_S^f(\theta_S, \theta'_B) - t_S^f(\theta_S, \theta_B) > 0] \, d\theta'_B \, d\theta_B \\
&= (1 - \lambda^2) \int_0^1 \int_0^1 \left(t_S^f(\theta_S, \theta'_B) - t_S^f(\theta_S, \theta_B) \right) \mathbb{1}[t_S^f(\theta_S, \theta'_B) - t_S^f(\theta_S, \theta_B) > 0] \, d\theta'_B \, d\theta_B,
\end{aligned}
$$

where $\mathbb{1}$ denotes the indicator function. The key step in the above derivation lies in the last equality. Comparing the two integrands on the third and second to last lines, we notice that they look the same but that 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 performing 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 $\lambda^2 > 1$ we find $w_i(\theta_i) \leq 0$. 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 coincides with deterministic transfer for almost all types. Thus, for all that matters, deterministic transfers are the only transfers that achieve $w_S(\theta_S) = 0$.

For the remainder of the paper we will consider only a special class of mechanisms. Namely, we consider the class of trade mechanisms that are δ-inefficient:

$$
y^f(\theta) = \begin{cases} 1 & \text{if } \theta_B - \theta_S > \delta, \\ 0 & \text{else,} \end{cases} \tag{7}
$$

for some $\delta \geq 0$. Note that this class of mechanisms contains (a) the materially efficient mechanism when setting $\delta = 0$, and (b) the revenue maximizing mechanism from MS in the framework without loss-aversion when $\delta = 1/2$.

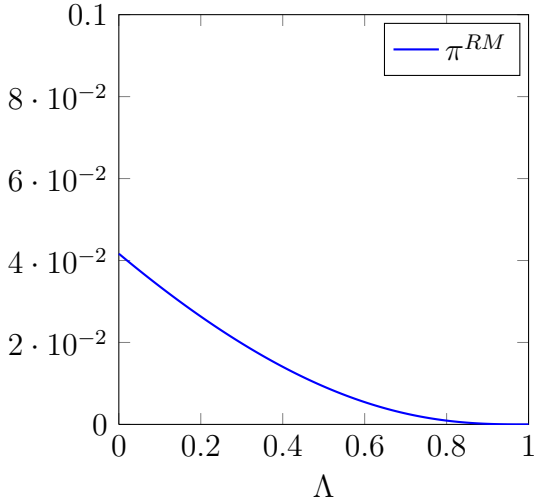
Proposition 5 *In the class of δ-inefficient mechanisms, expected revenue is maximized for $\delta^{RM} = 1/(2 - \Lambda)$. The maximal revenue is given by*

$$
\pi^{RM} = \frac{(1 - \Lambda)^3}{6(2 - \Lambda)^2} \geq 0.
$$

Proof. See Appendix B.3. ■

There are several things to note about this result. First, we already mentioned that the class of δ-inefficient mechanisms contains the revenue maximizing mechanism from MS where $\delta = 1/2$. Thus, the solution from the setting without loss-aversion is obviously the limit case when Λ goes to zero, i.e., when loss-aversion vanishes. Second, when loss-aversion is maximal, i.e., $\Lambda = 1$, no trade takes place. Third, δ^{RM} is larger than in the framework without loss-aversion, which means that less trade will take place in the

Figure 1: Maximal revenue in the class of $\delta-$inefficient mechanisms



presence of loss-aversion. Finally, the maximal revenue is monotonically decreasing in the parameter Λ and thus decreasing in the level of loss-aversion (see Figure 1).

To gain an intuition for this result we first consider the classic framework with no loss-aversion. In the absence of loss-aversion trade optimally takes place whenever $\theta_B - \theta_S \geq 1/2$. Hence, compared to the materially efficient case, there is a wedge: trade takes place less often. The function of this wedge is analogous to the role of a reserve price in an optimal auction. It increases the designers revenue by decreasing the agents' information rent.

Let us now turn to the case with loss-aversion. We have found that in this case the designer chooses a larger wedge than $1/2$. By increasing the wedge, the designer reduces the possibility of the agents feeling loss and can thereby increase her revenue. To see this, suppose $\Lambda = 8/9$ and therefore $\delta^{RM} = 0.9$. Obviously, for any $\theta_B \leq 0.9$ the buyer realizes that no trade will take place and thus has a utility of zero. In case $\theta_B > 0.9$ it is still very unlikely that trade will take place: the buyer assigns probability $\theta_B - 0.9 < 0.1$ to trade actually taking place. Consequently, her reference point is also relatively low ($0.1\theta_B < 0.1$). Thus, if no trade takes place, the loss is going to be quite small. If trade does take place, however, the gain is going to be correspondingly large. This allows the designer to extract more rent from the agent. Roughly speaking, keeping the utility level of the buyer constant, by reducing the expected loss, the designer can increase the transfer and thereby increase her revenue.

In some sense the designer is providing the agents with insurance against loss in the trade dimension by reducing the opportunities for trade. In contrast to the money dimension, where agents receive full insurance through deterministic transfers, insurance is only partial on the trade dimension. Thus, the insurance property from the minimal subsidy problem persists and even extends (partially) to the trade dimension.

Further, the result is consistent with the findings in the previous section on an additional level. There the presence of loss-aversion made it harder to satisfy CPEIC and IR with materially efficient trade, leading to a greater subsidy being required to induce materially efficient trade. The same forces lead to a smaller revenue when maximizing the profits compared to the framework in MS with no loss-aversion.

5.2 The Welfare Maximization Problem

In this section, we put ourselves in the shoes of a benevolent designer who wants to maximize ex ante welfare in a Benthamite sense. That is, the designer wants to maximize the sum of ex ante utilities. Besides CPEIC and IR, we impose one additional restriction on the set of available mechanisms. Namely, we do not want the designer to inject money in the economy. More precisely, we want budget balance on average. This is in line with the the preceding sections, where we looked at ex ante revenue maximization and the ex ante minimal subsidy. We say that a mechanism is ex ante budget balanced if

$$\int_0^1 \int_0^1 \left(t_S^f(\theta_S, \theta_B) - t_B^f(\theta_S, \theta_B) \right) \, d\theta_S \, d\theta_B = 0. \tag{AB}$$

The designer's problem reads

$$\max_{(y^f, t_B^t, t_S^f) \in \mathcal{F}} \int_0^1 U_S(\theta_S, s_B^t | \theta_S) \, d\theta_S + \int_0^1 U_B(\theta_B, s_S^t | \theta_B) \, d\theta_B,$$
$$\text{subject to CPEIC, IR and AB.} \tag{WM}$$

We will proceed as in the preceding section. The key difference to the revenue maximization problem is that we have to set up a Lagrangian to take care of the AB constraint.

Proposition 6 *The problem WM is equivalent to the problem*

$$\begin{aligned} \max_{\{(y^f, t_S^f, t_B^f, \gamma)\}} \quad & (1-\gamma) U_S(1, s_B^t | 1) + (1-\gamma) U_B(0, s_S^t | 0) \\ & + (1-2\gamma) \int_0^1 \theta_S \mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)] \left(1 - \Lambda \left(\mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)] - 1 \right) \right) \, d\theta_S \\ & + (1-\gamma) \int_0^1 (1-\theta_B) \mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)] \left(1 + \Lambda \left(\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)] - 1 \right) \right) \, d\theta_B \\ & + \gamma \int_0^1 \theta_B \mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)] \left(1 + \Lambda \left(\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)] - 1 \right) \right) \, d\theta_B \end{aligned}$$
$$\textit{subject to } \tilde{v}_S \textit{ being non-increasing}, \tilde{v}_B \textit{ being non-decreasing and IR.} \tag{WM'}$$

Proof. See Appendix B.4. ■

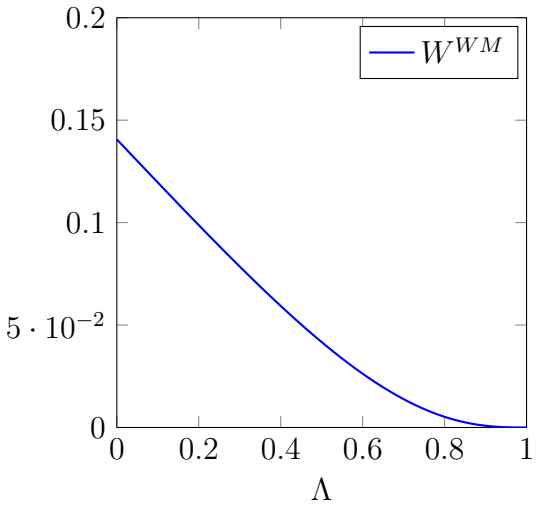
Mirroring the section on the revenue maximization problem we focus on the class of δ-inefficient mechanisms.

Proposition 7 *In the class of δ-inefficient mechanisms expected welfare is maximized for $\delta^{WM} = (1+\Lambda)/(2(2-\Lambda))$. The maximal expected welfare is given by*

$$W^{WM} = \frac{9(\Lambda-1)^3}{8(\Lambda-2)^3} \geq 0.$$

Proof. See Appendix B.5. ■

Figure 2: Maximal welfare in the class of $\delta-$inefficient mechanisms



This solution mirrors the solution to the revenue maximization problem. MS find that in the classic framework without loss-aversion expected welfare is maximized for $\delta = 1/4$, which is the limit case of δ^{WM} when loss-aversion vanishes. Thus, the designer optimally provides the agents with partial insurance in the trade dimension. Moreover, when loss-aversion is maximal, i.e., $\Lambda = 1$, no trade takes place. Finally, loss-aversion monotonically decreases welfare (see Figure 2) and the amount of trade taking place.

As in the minimal subsidy and revenue maximization problem, optimal transfers are deterministic. Thus, the designer provides the agents with complete insurance on the money dimension. This result extends the findings in Eisenhuth (2013), who finds that in any revenue maximizing mechanism transfers are deterministic, to the case of welfare maximization. In the general mechanism design approach in the appendix, we show that this result is not specific to the bilateral trade setting, but applies to *any* welfare maximizing mechanism.

6 Robust Mechanisms

The literature on robust mechanism design (pioneered by Bergemann and Morris, 2005) focuses on relaxing the common knowledge assumption about the environment. In the context of bilateral trade this kind of robustness analysis has been conducted by Copic and Ponsatí (2008). The type of robustness we have in mind here, however, is closer to the one mentioned in Bierbrauer and Netzer (2014) who introduce intention-based preferences into a mechanism design framework. One of their contributions is the introduction of mechanisms that are robust with respect to the existence of social preferences. That is, these robust mechanisms implement an economic outcome irrespective of whether or not agents have social preferences. In this section, we will show that our results display robustness to the exact specification of the reference point.

KR note that the equilibrium concepts in the models on disappointment-aversion by Bell (1985) and Loomes and Sugden (1986) are closely related to the CPE. The CPE specifies the reference point as the full distribution of a lottery, whereas the reference point corresponds to the certainty equivalent of the lottery in these models of disappointment-aversion. However, Masatlioglu and Raymond (2014) find that the intersection of preferences induced by the CPE and any of these disappointment-aversion models is simply expected utility. Thus, although the models seem to be very similar, the induced preferences do generally not coincide. Nevertheless, the optimal mechanisms derived in Section 5 and the results in Section 4 remain valid if we specify the reference point as the certainty equivalent of the lottery as in Bell (1985) and Loomes and Sugden (1986). To keep the analysis concise, we focus on the seller only. The arguments are essentially the same for the buyer. Under the alternative specification of the reference point the utility of the seller reads

$$\begin{aligned} U_S(\theta_S, s_B^t|\theta_S) = &\int_0^1 \left(-y^f(\theta_S,\theta_B)\theta_S + t_S^f(\theta_S,\theta_B)\right)\, d\theta_B \\ &+ \int_0^1 \left(\eta^1\mu^1\left(\mathbb{E}_B[y^f(\theta_S,\tilde{\theta}_B)]\theta_S - y^f(\theta_S,\theta_B)\theta_S\right) + \eta^2\mu^2\left(t_S^f(\theta_S,\theta_B) - \mathbb{E}_B[t_S^f(\theta_S,\tilde{\theta}_B)]\right)\right)\, d\theta_B. \end{aligned}$$

Comparing this alternative expression to the expected utility we worked with (see equation (3)), we notice that the material utility on the first line remains unchanged, while the gain-loss utility in the second line takes a new form. Indeed, instead of comparing the induced outcome to every single potential outcome in the reference lottery, the agent now compares the outcome only to the certainty equivalent of the reference lottery, which enters the value function directly. Two observations about the alternative gain-loss utility yield the robustness result. Consider the money dimension first and recall that μ^2 is a concave function. Thus, by Jensen's inequality we get

$$\int_0^1 \eta^2\mu^2\left(t_S^f(\theta_S,\theta_B) - \mathbb{E}_B[t_S^f(\theta_S,\tilde{\theta}_B)]\right)\, d\theta_B$$

$$\leq \eta^2\mu^2 \left(\int_0^1 \left(t_S^f(\theta_S, \theta_B) - \mathbb{E}_B[t_S^f(\theta_S, \tilde{\theta}_B)] \right) \, d\theta_B \right) = 0,$$

as $\int_0^1 t_S^f(\theta_S, \theta_B) \, d\theta_B = \mathbb{E}_B[t_S^f(\theta_S, \tilde{\theta}_B)]$ by definition. Therefore, the result in Lemma 1 carries through to this specification. Hence, irrespective of which of the two specifications of the reference point we use, the gain-loss utility in the money dimension is non-positive, making deterministic transfers optimal under this alternative specification.

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

$$\begin{aligned}
&\int_0^1 \eta^1\mu^1 \left(\mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)]\theta_S - y^f(\theta_S, \theta_B)\theta_S \right) \, d\theta_B \\
&= \theta_S\eta^1 \int_0^1 \left(\lambda^1 y^f(\theta_S, \theta_B) \left(\mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)] - 1 \right) + (1 - y^f(\theta_S, \theta_B))\mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)] \right) \, d\theta_B \\
&= \theta_S\eta^1 \int_0^1 \int_0^1 \left(\lambda^1 y^f(\theta_S, \theta_B)(y^f(\theta_S, \theta'_B) - 1) + (1 - y^f(\theta_S, \theta_B))y^f(\theta_S, \theta'_B) \right) \, d\theta'_B \, d\theta_B \\
&= \theta_S\eta^1 \int_0^1 \int_0^1 \mu^1(y^f(\theta_S, \theta'_B) - y^f(\theta_S, \theta_B)) \, d\theta'_B \, d\theta_B,
\end{aligned}$$

where the final line is the very expression of gain-loss utility in the trade dimension under the specification used in the paper, that is, with the reference point being the lottery. Thus, regarding gain-loss utility in the trade dimension the two different specifications of the reference point are equivalent.[7] Consequently, all of our results continue to hold under the alternative specification of the reference point, as the two specifications are equivalent conditional on deterministic transfers.

7 Conclusion

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

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

[7] Notice that this finding does not hinge on the piece-wise linearity of μ^1, but is solely due to the binary nature of trade.

ing institutions. We find in all three cases that the impossibility problem becomes more severe in the presence of loss-aversion. The common theme in all three problems is that of insurance. In both, welfare and revenue maximizing mechanisms, the designer induces less trade in the presence of loss-aversion. The intuition for this result is that the designer provides the agents with partial insurance in the trade dimension. Moreover, the designer optimally provides agents with full insurance in the money dimension in any revenue or welfare maximizing mechanisms. That is, deterministic transfers are optimal.

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

Throughout the main text we have upheld two symmetry assumptions. We have assumed that agents have symmetric type spaces and that the loss-aversion parameters are symmetric. Both assumptions are made for tractability and are without loss of generality. In the case of asymmetric type spaces, we can distinguish between favorable and unfavorable asymmetry, in the sense that they increase or decrease the likelihood of trade taking place, respectively. By favorable asymmetry we mean the case when the largest buyer type is larger than the largest seller type and/or the smallest buyer type is larger than the smallest seller type. An unfavorable asymmetry captures the reversed cases. One can show that all our results are robust to both types of asymmetry and thus the we can restrict the analysis to the more tractable symmetric case. Relaxing the symmetry assumption on the loss-aversion parameters would change the precise value of the optimal δ in the revenue and welfare maximizing mechanisms, but not change the fact, that less trade is induced or that optimal transfers are deterministic. The minimal subsidy would change, too, but it would still be increasing in the degree of loss-aversion and bigger than in the classic framework.

KR introduce the concept of an unacclimating personal equilibrium (UPE), as an alternative to the CPE we used. The difference lies in the timing of the formation of the expectation. The CPE is suited for situations when there is sufficient time between the taking a decision and the realization of the payoffs. The UPE, however, should be used when this is not the case, i.e., when the agent first forms expectations and therefore the reference point and only subsequently takes an action, taking expectations as given. Thus, in such a setting there is not enough time for the expectations to acclimate to the action. In order to guarantee internal consistency, the UPE requires the agent to only form expectations which he will also want to comply with. Thus, expectations are met in equilibrium and expected utility takes the same form under UPE as it does under CPE.

Since the aversion to ex post variation in payoffs is what drives the results, we expect the partial and full insurance results in the trade and money dimension, respectively, to be robust to the use of UPE.

A question of some importance is how an optimal mechanism can be implemented. In the classic framework without loss-aversion, Chatterjee and Samuelson (1983) show that there exists an equilibrium in linear strategies in the double auction, which implements the welfare maximizing outcome. Moreover, they show that this equilibrium extends to the case of risk-averse agents and find that a sufficiently high degree of constant relative risk-aversion induces an ex post efficient outcome. One can show, however, that in the present case with loss-averse agents there does not exist such an equilibrium in linear strategies.

Going forward, there are many open questions, especially in the context of general mechanism design. Most prominently, the existence of an expected externality mechanism is not certain and the appropriate formulations in the context of equilibria in dominant strategies are yet to be made. Moreover, further advances on wide bracketing of losses could be possible and are of interest.

References

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011): "Reference Points and Effort Provision," *American Economic Review*, 101, 470–492.

Bell, D. E. (1985): "Disappointment in Decision Making under Uncertainty," *Operations Research*, 33, 1–27.

Bergemann, D. and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 73, 1771–1813.

Bierbrauer, F. and N. Netzer (2014): "Mechanism Design and Intentions," Working paper, University of Zurich.

Carbajal, J. C. and J. C. Ely (2014): "A Model of Price Discrimination under Loss Aversion and State-Contingent Reference Points," Mimeo.

Cavallo, R. (2011): "Efficient Mechanisms with Risky Participation," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, ed. by T. Walsh, AAAI Press/International Joint Conferences on Artificial Intelligence.

Chatterjee, K. and W. Samuelson (1983): "Bargaining under Incomplete Information," *Operations Research*, 31, 835–851.

Copic, J. and C. Ponsatí (2008): "Ex-Post Constrained-Efficient Bilateral Trade with Risk-Averse Traders," Mimeo.

Crawford, V. P. and J. Meng (2011): "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 101, 1912–1932.

de Meza, D. and D. C. Webb (2007): "Incentive Design under Loss Aversion," *Journal of the European Economic Association*, 5, 66–92.

DellaVigna, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–372.

Eisenhuth, R. (2013): "Reference Dependent Mechanism Design," Mimeo.

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.

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.

Garratt, R. and M. Pycia (2014): "Efficient Bilateral Trade," Mimeo, FRBNY and UCLA.

Genesove, D. and C. Mayer (2001): "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *The Quarterly Journal of Economics*, 116, 1233–1260.

Gill, D. and V. Prowse (2012): "A Structural Analysis of Disappointment Aversion in a Real Effort Competition," *American Economic Review*, 102, 469–503.

Gill, D. and R. Stone (2010): "Fairness and desert in tournaments," *Games and Economic Behavior*, 69, 346–364.

Gul, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59, 667–686.

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 (forthcoming): "Loss Aversion and Consumption Choice: Theory and Experimental Evidence," *American Economic Journal: Microeconomics*.

Karle, H. and M. Peitz (2014): "Competition under consumer loss aversion," *The RAND Journal of Economics*, 45, 1–31.

Kőszegi, B. (2014): "Behavioral Contract Theory," *Journal of Economic Literature*, 52, 1075–1118.

Kőszegi, B. and M. Rabin (2006): "A Model of Reference-Dependent Preferences," *The Quarterly Journal of Economics*, 121, 1133–1165.

——— (2007): "Reference-Dependent Risk Attitudes," *The American Economic Review*, 97, 1047–1073.

Kucuksenel, S. (2012): "Behavioral Mechanism Design," *Journal of Public Economic Theory*, 14, 767–789.

Loomes, G. and R. Sugden (1986): "Disappointment and Dynamic Consistency in Choice under Uncertainty," *The Review of Economic Studies*, 53, 271–282.

Masatlioglu, Y. and C. Raymond (2014): "A Behavioral Analysis of Stochastic Reference Dependence," mimeo, University of Michigan and University of Oxford.

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.

Salant, Y. and R. Siegel (forthcoming): "Reallocation Costs and Efficiency," *American Economic Journal:Microeconomics*.

A A General Mechanism Design Approach

In this section we consider general mechanisms, that is, we do not limit ourselves to the bilateral trade problem. This presents a generalization of the narrow bracketing part of Eisenhuth (2013). He considers an auction framework only and assumes that material utility is linear from the beginning.

A.1 Utility, Social Choice Functions and Mechanisms

An environment $E = [I, X, (\Theta_i, \pi_i)_{i \in I}, F_i]$ is characterized by the following components. There is a finite set of N agents denoted by $I = \{1, \ldots, N\}$. The set of social alternatives is given by X. We consider an independent private values setting. Hence, the type of agent i is private information and is independently drawn from a distribution F_i with bounded support $\Theta_i \subset \mathbb{R}_+$. Throughout, we use the conventional notation $\Theta = \prod_{i=1}^{N} \Theta_i$, with typical element θ, and $\Theta_{-i} = \prod_{j \neq i} \Theta_j$, with typical element θ_{-i}. The agents and the principal have identical prior beliefs.

Following KR, agents' utility is additively separable in material utility and in gain-loss utility. The function $\pi_i : X \times \Theta_i \to \mathbb{R}$ is the material utility function. The riskless total utility function of agent i is given by $u_i : X \times \mathbb{R} \times \Theta_i \to \mathbb{R}$ and is defined as

$$u_i(x, r_i, \theta_i) = \pi_i(x, \theta_i) + \eta\mu(\pi_i(x, \theta_i) - \pi_i(r_i, \theta_i)), \tag{8}$$

with some $\eta \geq 0$ and where

$$\mu(s) = \begin{cases} s & s \geq 0, \\ \lambda s & s < 0, \end{cases}$$

is a value function in the sense of Kahneman and Tversky (1979), with $\lambda > 1$, thereby capturing loss-aversion.

The parameter r_i is the riskless reference level. We allow for the reference point to be stochastic, i.e., to be a reference lottery over all riskless reference levels. Following KR the reference point will be equal to the agent's rational expectations.

A social choice function (SCF) $f : \Theta \to X$ assigns a collective choice $f(\theta_1, \ldots, \theta_N) \in X$ to each possible profile of the agents' types $(\theta_1, \ldots, \theta_N) \in \Theta$. We denote the set of all SCFs $\mathcal{F}$.

A mechanism $\Gamma = (M_1, \ldots, M_N, g)$ is a collection of N message sets $(M_1, \ldots, M_N)$ and an outcome function $g : M_1 \times \ldots \times M_N \to X$. We denote the direct mechanism by $\Gamma^d = (\Theta_1, \ldots, \Theta_N, f)$. Since agents privately observe their types, they can condition their

message on their type. Consequently, a pure strategy for agent i in a mechanism Γ is a function $s_i : \Theta_i \to M_i$. Note that $g(s_1(\theta_1), \ldots, s_N(\theta_N)) = x \in X$. Let S_i denote the set of all pure strategies of agent i. Further, we denote the truthful strategy $s_i^t(\theta_i) = \theta_i$.

A.2 Equilibrium Concept and Revelation Principle

KR introduce different equilibrium concepts. The concept used here is based on their choice-acclimating personal equilibrium (CPE). The set of all possible reference levels is given by the set of all social alternatives, X. Thus, an agent compares the eventual outcome to what could have happened. As mentioned above, we allow for the reference point to be a distribution over the set X. In a mechanism Γ, this distribution is induced endogenously for each agent: conditional on the other agents playing s_{-i}, agent i induces a distribution over the set of social alternatives, X, by playing the strategy s_i. Hence, the loss-averse agent will compare any given social alternative to all possible social alternatives, allowing for gain or loss feelings in every comparison. Moving to the interim stage and allowing for a reference lottery, we can define the interim expected utility of agent i with type θ_i, in the mechanism Γ, when playing strategy s_i, given that the other agents play s_{-i} as

$$U_i(s_i(\theta_i), s_{-i}, \Gamma|\theta_i) = \int_{\Theta_{-i}} \int_{\Theta_{-i}} [\pi(s_i(\theta_i), s_{-i}(\theta_{-i}), g) + \eta\mu\left(\pi(s_i(\theta_i), s_{-i}(\theta_{-i}), g) - \pi(s_i(\theta_i), s_{-i}(\theta'_{-i}), g)\right)] \; dF_{-i}(\theta'_{-i}) \; dF_{-i}(\theta_{-i}).$$

Defining the reference point this way, we keep with the spirit of the CPE in KR, as the strategy determines both the reference point and the eventual outcome. The outer integral corresponds to taking the expectation over all possible types of the buyer and yields standard interim expected utility. The inner integral corresponds to the reference point. Recall that an outcome is compared to all social alternatives and that the reference point, or rather distribution, is induced endogenously. Thus, for any θ_{-i} in the domain of integration of the outer integral, the inner integral allows us to compare the induced outcome to all other potential outcomes. We can now define our equilibrium concept, which follows Eisenhuth (2013).

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

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

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

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

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

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

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

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

Proof. Suppose f was CPEIC. Then, by definition the strategy profile s^t a CPE in the direct mechanism Γ^d and thus, again by definition, the direct mechanism implements f in CPE. Conversely, suppose there is a mechanism $\Gamma = (M_1, \ldots, M_N, g)$ that implements f in CPE. If $s^* = (s_1^*, \ldots, s_N^*)$ is a CPE, then for all $i, m_i' \in M_i$ and θ_i

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

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

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

Since Γ implements f in CPE we have

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

implying

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

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

A.3 Incentive Compatibility

We now move closer to the setting in the first part of Eisenhuth (2013) by assuming $X = Y \times T$ with typical element $\mathbf{x} = (y, t_1, t_2, \ldots, t_N)$. The (general) set Y is the set

of projects and the set $T \subset \mathbb{R}^N$ the set of transfers. A social choice function in this environment takes the form $f = (y^f, t_1^f, \ldots, t_N^f)$. Thanks to the revelation principle we can limit attention to direct mechanisms and henceforth simplify notation by dropping the argument referring to the mechanism in the utility function. We further assume

$$\pi_i(\mathbf{x}, \theta_i) = \theta_i v_i(y) + t_i,$$

for some $v_i : Y \to \mathbb{R}$, that is, material utility is linear in the type. Moreover, we let $\Theta_i = [0, 1]$. Interim expected utility of playing action m_i when all other agents play according to the truthful strategy becomes

$$\begin{aligned} U_i(m_i, s_{-i}^t|\theta_i) = \theta_i &\int_{\Theta_{-i}} v_i(y^f(m_i, \theta_{-i}))\, dF(\theta_{-i}) + \int_{\Theta_{-i}} t_i^f(m_i, \theta_{-i})\, dF(\theta_{-i}) \\ &+ \eta^1 \theta_i \int_{\Theta_{-i}} \int_{\Theta_{-i}} \mu^1 \left(v_i(y^f(m_i, \theta_{-i})) - v_i(y^f(m_i, \theta'_{-i})) \right) dF_i(\theta'_{-i})\, dF(\theta_{-i}) \\ &+ \eta^2 \int_{\Theta_{-i}} \int_{\Theta_{-i}} \mu^2 \left(t_i^f(m_i, \theta_{-i}) - t_i^f(m_i, \theta'_{-i}) \right) dF_i(\theta'_{-i})\, dF(\theta_{-i}). \end{aligned}$$

This utility specification corresponds to narrow bracketing, meaning that for the two material utility dimensions, consumption and money utility, there is a separate gain-loss term each. Thus, gain-loss feelings are bracketed narrowly and not widely, as is for instance the case in the second part of Eisenhuth (2013). We allow for the weight parameters η^1 and η^2 to differ across the two dimensions and also allow for the value functions μ^1 and μ^2 to differ in the parameters λ^1 and λ^2. Further, notice that the proof of the revelation principle goes through unchanged. In order to economize notation we define

$$\begin{aligned} \tilde{v}_i(m_i) = &\int_{\Theta_{-i}} v_i(y^f(m_i, \theta_{-i}))\, dF(\theta_{-i}) \\ &+ \eta^1 \int_{\Theta_{-i}} \int_{\Theta_{-i}} \mu^1 \left(v_i(y^f(m_i, \theta_{-i})) - v_i(y^f(m_i, \theta'_{-i})) \right) dF_i(\theta'_{-i})\, dF(\theta_{-i}), \\ \tilde{t}_i(m_i) = &\int_{\Theta_{-i}} t_i^f(m_i, \theta_{-i})\, dF(\theta_{-i}) \\ &+ \eta^2 \int_{\Theta_{-i}} \int_{\Theta_{-i}} \mu^2 \left(t_i^f(m_i, \theta_{-i}) - t_i^f(m_i, \theta'_{-i}) \right) dF_i(\theta'_{-i})\, dF(\theta_{-i}). \end{aligned}$$

This allows us to write $U_i(m_i, s_{-i}^t|\theta_i) = \theta_i \tilde{v}_i(m_i) + \tilde{t}_i(m_i)$. It will turn out to be useful to further define

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

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

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

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

Proposition 9 *The social choice function $f = (y^f, t^f_1, \ldots, t^f_N)$ is CPEIC if and only if, for all $i \in I$,*

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

(ii) $U_i(\theta_i, s^t_{-i}|\theta_i) = U_i(0, s^t_{-i}|0) + \int_0^{\theta_i} \tilde{v}_i(s)\, ds$ for all $\theta_i \in \Theta_i$.

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

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

and analogously

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

Thus,

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

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

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

and so

$$U_i(\theta_i, s^t_{-i}|\theta_i) = U_i(0, s^t_{-i}|0) + \int_0^{\theta_i} \tilde{v}_i(s)\, ds$$

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

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

$$= (\theta_i - \hat{\theta}_i)\tilde{v}_i(\hat{\theta}_i).$$

Hence,

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

and similarly

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

Consequently, f is CPEIC. ■

A.4 Deterministic Transfers

In this subsection we present two additional results that hold in general. The first was already stated in Eisenhuth (2013) and says that in any revenue maximizing mechanism transfers are deterministic. The second extends this finding to welfare maximizing mechanisms.

We say that an SCF is individually rational if for all agents $i \in I$

$$U_i(\theta_i, s^t_{-i}|\theta_i) \geq 0, \quad \forall \theta_i \in \Theta_i, \tag{IR}$$

that a mechanism is ex ante budget balanced if

$$\sum_{i=1}^{N} \int_0^1 \bar{t}_i(\theta_i)\, dF_i(\theta_i) = 0, \tag{AB}$$

and that transfers are deterministic if for all $i \in I$, $\theta_i \in \Theta_i$ we have $t^f_i(\theta_i, \theta_{-i}) = t^f_i(\theta_i, \theta'_{-i})$ for all $\theta_{-i}, \theta'_{-i} \in \Theta_{-i}$ with $\theta_{-i} \neq \theta'_{-i}$.

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

$$\min_{(y^f, t^f_1, \ldots, t^f_N) \in \mathcal{F}} \sum_{i=1}^{N} \int_0^1 \bar{t}_i(\theta_i)\, dF_i(\theta_i),$$

subject to CPEIC and IR.

We first prove a lemma which we will use repeatedly.

Lemma 1 *We have $w_i(\theta_i) \leq 0$ for all i and $\theta_i \in \Theta_i$.*

Proof. Recall that we defined

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

We can rewrite these expressions as follows

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

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

Note that any transfers achieve $w_i(\theta_i) = 0$ if and only if the transfers coincide with deterministic transfers for almost all types. Thus, for all that matters, deterministic transfers are the only transfers that achieve $w_i(\theta_i) = 0$.

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

$$\min_{(y^f, t_1^f, \ldots, t_N^f) \in \mathcal{F}} \sum_{i=1}^{N} \int_0^1 \left(\bar{t}_i(0) + \eta^2 w_i(0) - \theta_i \tilde{v}_i(\theta_i) - \eta^2 w_i(\theta_i) + \int_0^{\theta_i} \tilde{v}_i(s) \, ds \right) \, dF_i(\theta_i),$$

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

By Lemma 1 we have $w_i(\theta_i) \leq 0$ for all i and $\theta_i \in \Theta_i$. Note that these terms enter the problem negatively. Since we want to minimize the objective function, we optimally choose transfers such that $w_i(\theta_S) = 0$ for all $\theta_i \in [0, 1]$ to minimize the integrands pointwise and therefore minimize the integrals. Doing so does not contradict the IR constraint, on the contrary, it relaxes it. Thus, choosing deterministic transfers is optimal and part of a solution to the problem. ■

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

$$\min_{(y^f,t_1^f,\ldots,t_N^f)\in\mathcal{F}} \sum_{i=1}^{N} \int_0^1 U_i(\theta_i, s_{-i}^t|\theta_i)\, dF_i(\theta_i),$$

subject to CPEIC, IR and AB.

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

$$\sum_{i=1}^{N} \left(U_i(0, s_{-i}^t|0) + \int_0^1 \int_0^{\theta_i} \tilde{v}_i(s)\, ds\, dF_i(\theta_i) \right). \tag{9}$$

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

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

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

$$\sum_{i=1}^{N} \int_0^1 \bar{t}_i(\theta_i)\, dF_i(\theta_i) = 0.$$

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

$$\sum_{i=1}^{N} \int_0^1 \left(\eta^2 w_i(\theta_i) + \theta_i \tilde{v}_i(\theta_i) - U_i(0, s_{-i}^t|0) - \int_0^{\theta_i} \tilde{v}_i(t)\, dt \right) dF_i(\theta_i) = 0. \tag{10}$$

Using the rewritten objective function in (9) and using the form of the AB constraint in (10), we can set up a Lagrangian:

$$\begin{aligned}\mathcal{L}(y^f, t_1^f, \ldots, t_N^f, \gamma) = &\sum_{i=1}^{N} (1-\gamma) U_i(0, s_{-i}^t|0) + \sum_{i=1}^{N} (1-\gamma) \int_0^1 \int_0^{\theta_i} \tilde{v}_i(s)\, ds\, dF_i(\theta_i) \\ &+ \gamma \sum_{i=1}^{N} \int_0^1 \eta^2 w_i(\theta_i)\, dF_i(\theta_i) + \gamma \sum_{i=1}^{N} \int_0^1 \theta_i \tilde{v}_i(\theta_i)\, dF_i(\theta_i),\end{aligned}$$

where γ is the Lagrange multiplier. By Lemma 1 we have $w_i(\theta_i) \leq 0$ for $i \in I$, which enter the Lagrangian positively. In order to maximize the Lagrangian, we can choose deterministic transfers which result in $w_i(\theta_i) = 0$ for $i \in I$. Note that this is in line with the remaining constraints given by condition (i) from Proposition 9 and the IR constraint.

■

B Proofs

B.1 Proof of Proposition 3

Step 1. We first find conditions for the mechanism to be CPEIC, i.e., we need (i) and (ii) in Proposition 2 to be satisfied. Making use of the functional form of the materially efficient SCF we get

$$\begin{aligned}\tilde{v}_B(\theta_B) &= \int_0^1 y^f(\theta_S,\theta_B)\,d\theta_S + \eta^1\int_0^1\int_0^1 \mu^1\left(y^f(\theta_S,\theta_B)-y^f(\theta'_S,\theta_B)\right)\,d\theta'_S\,d\theta_S \\ &= \theta_B + \eta^1\int_0^{\theta_B}\int_{\theta_B}^1 1\,d\theta'_S\,d\theta_S - \eta^1\lambda^1\int_{\theta_B}^1\int_0^{\theta_B} 1\,d\theta'_S\,d\theta_S \\ &= \theta_B - (1-\theta_B)\theta_B\Lambda \end{aligned} \tag{11}$$

and

$$\begin{aligned}\tilde{v}_S(\theta_S) &= \int_0^1 y^f(\theta_S,\theta_B)\,d\theta_B - \eta^1\int_0^1\int_0^1 \mu^1\left(y^f(\theta_S,\theta'_B)-y^f(\theta_S,\theta_B)\right)\,d\theta'_B\,d\theta_B \\ &= 1-\theta_S + \eta^1\lambda^1\int_0^{\theta_S}\int_{\theta_S}^1 1\,d\theta'_B\,d\theta_B - \eta^1\int_{\theta_S}^1\int_0^{\theta_S} 1\,d\theta'_B\,d\theta_B \\ &= 1-\theta_S+\theta_S(1-\theta_S)\Lambda. \end{aligned} \tag{12}$$

Taking the respective derivatives it is easy to see that $\tilde{v}_B$ and $\tilde{v}_S$ are non-decreasing and non-increasing, respectively, given our assumption $\Lambda \leq 1$. Thus, condition (i) is satisfied. For condition (ii) to be satisfied we need there to exist transfer functions t_S^f and t_B^f such that we can write utility as

$$U_S(\theta_S, s_B^t|\theta_S) = U_S(1, s_B^t|1) + \int_{\theta_S}^1 (1-\theta_S+\theta_S(1-\theta_S)\Lambda)\;d\theta_S \tag{13}$$

$$= \bar{t}_S(1) + \eta^2 w_S(1) + \frac{1}{2} + \frac{\Lambda}{6} - \theta_S + \frac{\theta_S^2}{2} - \left(\frac{\theta_S^2}{2} - \frac{\theta_S^3}{3}\right)\Lambda \tag{14}$$

$$U_B(\theta_B, s_S^t|\theta_B) = U_B(0, s_S^t|0) + \int_0^{\theta_B} (\theta_B - (1-\theta_B)\theta_B\Lambda)\;d\theta_B \tag{15}$$

$$= -\bar{t}_B(0) + \eta^2 w_B(0) + \frac{\theta_B^2}{2} - \left(\frac{\theta_B^2}{2} - \frac{\theta_B^3}{3}\right)\Lambda. \tag{16}$$

Step 2. We next find conditions on the mechanism to satisfy IR. From equations (13) and (15) we know that utility of the "worst" types of the buyer and seller (0 and 1, respectively) is given by

$$U_S(1, s_B^t|1) = \bar{t}_S(1) + \eta^2 w_S(1),$$
$$U_B(0, s_S^t|0) = -\bar{t}_B(0) + \eta^2 w_B(0),$$

which we both need to be greater equal zero for IR. From equations (11) and (12) we know that $\tilde{v}_B(0) = \tilde{v}_S(1) = 0$. Moreover, from CPEIC we know that $\tilde{v}_B$ and $\tilde{v}_S$ are non-decreasing and non-increasing, respectively. Thus, the integrand in (13) is positive on the complete domain of integration, as the function $\tilde{v}_S$ is zero at the upper-bound of the domain and, because it is non-increasing, it is positive on the rest of the domain of integration. Similarly, the integrand in (15) is positive because the function $\tilde{v}_B$ is equal to zero at the lower-bound and, because it is non-decreasing, it is positive on the rest of the integration domain. Hence, since the integral of a positive function is positive, the utility of all the types of both agents is greater than zero if and only if

$$\bar{t}_S(1) + \eta^2 w_S(1) \geq 0 \text{ and } -\bar{t}_B(0) + \eta^2 w_B(0) \geq 0. \tag{17}$$

Step 3. Recall from Section 3.1 that utility can be written as

$$U_S(m_S, s_B^t|\theta_S) = -\theta_S \tilde{v}_S(m_S) + \bar{t}_S(m_S) + \eta^2 w_S^2(m_S)$$

and

$$U_B(m_B, s_S^t|\theta_B) = \theta_B \tilde{v}_B(m_B) - \bar{t}_B(m_B) + \eta^2 w_B(m_B).$$

Using equations (13) to (16) we can then write expected transfers as

$$\begin{aligned}\bar{t}_S(\theta_S) &= \theta_S \tilde{v}_S(\theta_S) - \eta^2 w_S(\theta_S) + \bar{t}_S(1) + \eta^2 w_S(1) + \frac{1}{2} + \frac{\Lambda}{6} - \theta_S + \frac{\theta_S^2}{2} - \left(\frac{\theta_S^2}{2} - \frac{\theta_S^3}{3}\right)\Lambda \\ &= -\eta^2 w_S(\theta_S) + \bar{t}_S(1) + \eta^2 w_S(1) + \frac{1}{2} + \frac{\Lambda}{6} - \frac{\theta_S^2}{2} - \left(\frac{2}{3}\theta_S^3 - \frac{1}{2}\theta_S^2\right)\Lambda \end{aligned} \tag{18}$$

and

$$\begin{aligned}\bar{t}_B(\theta_B) &= \theta_B \tilde{v}_B(\theta_B) + \eta^2 w_B(\theta_B) + \bar{t}_B(0) - \eta^2 w_B(0) - \frac{\theta_B^2}{2} + \left(\frac{\theta_B^2}{2} - \frac{\theta_B^3}{3}\right)\Lambda \\ &= \eta^2 w_B(\theta_B) + \bar{t}_B(0) - \eta^2 w_B(0) + \frac{\theta_B^2}{2} + \left(\frac{2}{3}\theta_B^3 - \frac{1}{2}\theta_B^2\right)\Lambda. \end{aligned} \tag{19}$$

We can now rewrite the minimal subsidy problem in (MSP) to

$$\max_{\{t_S,t_B\}} \int_0^1 \left(\eta^2 w_B(\theta_B) + \bar{t}_B(0) - \eta^2 w_B(0) + \frac{\theta_B^2}{2} + \left(\frac{2}{3}\theta_B^3 - \frac{1}{2}\theta_B^2\right)\Lambda\right) d\theta_B$$
$$+ \int_0^1 \left(\eta^2 w_S(\theta_S) - \bar{t}_S(1) - \eta^2 w_S(1) - \frac{1}{2} - \frac{\Lambda}{6} + \frac{\theta_S^2}{2} + \left(\frac{2}{3}\theta_S^3 - \frac{1}{2}\theta_S^2\right)\Lambda\right) d\theta_S. \tag{MSP'}$$

subject to (17). All the constraints in (MSP) are still respected: The constraint that y^f be ME is taken care of by explicitly plugging in the right functional form when we derived $\tilde{v}_B$ and $\tilde{v}_S$ in (11) and (12). The CPEIC constraint is taken care of implicitly in the functional form of the transfers in (18) and (19). The IR constraint is respected jointly by the constraint that y^f be ME and the remaining constraint in (17).

Step 4. We will now maximize the integrands in (MSP') pointwise, thereby maximizing the integrals. By Lemma 1 we get that $w_S(\theta_S) \leq 0$ and $w_B(\theta_B) \leq 0$ for all $\theta_B, \theta_S \in [0,1]$. Moreover, by the constraint (17) we know that

$$\bar{t}_S(1) + \eta^2 w_S(1) \geq 0 \text{ and } -\bar{t}_B(0) + \eta^2 w_B(0) \geq 0,$$

but both expressions enter the integrands negatively. Moreover, w_B and w_S enter the integrands positively (but are negative). Hence, in order to maximize the integrands pointwise we need transfers such that

$$\begin{aligned} \bar{t}_S(1) + \eta^2 w_S(1) &= 0 \\ -\bar{t}_B(0) + \eta^2 w_B(0) &= 0 \\ w_B(\theta_B) &= 0, \ \forall \theta_B \in [0,1] \\ w_S(\theta_S) &= 0, \ \forall \theta_S \in [0,1]. \end{aligned}$$

One can easily check that the deterministic transfers

$$\begin{aligned} t_S^f(\theta_S, \theta_B) &= \frac{1}{2} + \frac{\Lambda}{6} - \frac{\theta_S^2}{2} - \left(\frac{2}{3}\theta_S^3 - \frac{1}{2}\theta_S^2\right)\Lambda, \\ t_B^f(\theta_S, \theta_B) &= \frac{\theta_B^2}{2} + \left(\frac{2}{3}\theta_B^3 - \frac{1}{2}\theta_B^2\right)\Lambda, \end{aligned}$$

satisfy these conditions, thus maximize the integrands pointwise and therefore maximize the objective function in (MSP) while satisfying all the constraints. The value of

$$\int_0^1 \left(\frac{\theta_B^2}{2} + \left(\frac{2}{3}\theta_B^3 - \frac{1}{2}\theta_B^2\right)\Lambda\right) d\theta_B - \int_0^1 \left(\frac{1}{2} + \frac{\Lambda}{6} - \frac{\theta_S^2}{2} - \left(\frac{2}{3}\theta_S^3 - \frac{1}{2}\theta_S^2\right)\Lambda\right) d\theta_S$$

is $-(1+\Lambda)/6$. Therefore, the minimal subsidy needed is $(1+\Lambda)/6$.

B.2 Proof of Proposition 4

Step 1. We begin by simplifying the problem. 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 (5) and (6) from condition (ii), we can rewrite the objective function in the problem (RM) to

$$\int_0^1 \left(\eta^2 w_B(\theta_B) + \theta_B \tilde{v}_B(\theta_B) - U_B(0, s_S^t|0) - \int_0^{\theta_B} \tilde{v}_B(t)\, dt\right)\, d\theta_B$$
$$+ \int_0^1 \left(\eta^2 w_S(\theta_S) - \theta_S \tilde{v}_S(\theta_S) - U_S(1, s_B^t|1) - \int_{\theta_S}^1 \tilde{v}_S(t)\, dt\right)\, d\theta_S.$$

From the IR constraint we have $U_B(0, \theta_S|0) \geq 0$ and $U_S(1, \theta_B|1) \geq 0$, which enter the objective function negatively. Since we are maximizing the objective function, we choose transfers such that $U_B(0, \theta_S|0) = 0$ and $U_S(1, \theta_B|1) = 0$. If the expected utility of these "worst" types was not equal to zero in the optimal mechanism, we could modify the transfers by adding lump-sum transfers and reduce their expected utility to zero without affecting CPEIC. Moreover, w_B and w_S, which are negative by Lemma 1, enter positively. Thus, we impose an additional restriction on transfers, namely that they are deterministic, which leads to $w_B(\theta_B) = w_S(\theta_S) = 0$ for all $\theta_B, \theta_S \in [0, 1]$. Note that these two restrictions on transfers do not contradict each other. Given this, the problem reduces to

$$\max_{\{(y^f, t_S^f, t_B^f)\}} \int_0^1 \left(\theta_B \tilde{v}_B(\theta_B) - \int_0^{\theta_B} \tilde{v}_B(t)\, dt\right)\, d\theta_B$$
$$+ \int_0^1 \left(-\theta_S \tilde{v}_S(\theta_S) - \int_{\theta_S}^1 \tilde{v}_S(t)\, dt\right)\, d\theta_S$$

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

Step 2. We next rewrite the objective function in this reduced problem. Using integration by parts we get

$$\int_0^1 \left(\theta_B \tilde{v}_B(\theta_B) - \int_0^{\theta_B} \tilde{v}_B(t)\, dt\right)\, d\theta_B + \int_0^1 \left(-\theta_S \tilde{v}_S(\theta_S) - \int_{\theta_S}^1 \tilde{v}_S(t)\, dt\right)\, d\theta_S$$
$$= \int_0^1 \tilde{v}_B(\theta_B)(2\theta_B - 1)\, d\theta_B - \int_0^1 2\theta_S \tilde{v}_S(\theta_S)\, d\theta_S.$$

Further, the first integral can be rewritten as

$$\int_0^1 \tilde{v}_B(\theta_B)(2\theta_B - 1)\, d\theta_B$$
$$= \int_0^1 \left(\int_0^1 y^f(\theta_S, \theta_B)\, d\theta_S + \eta^1 \int_0^1 \int_0^1 \mu^1 \left(y^f(\theta_S, \theta_B) - y^f(\theta'_S, \theta_B)\right)\, d\theta'_S\, d\theta_S\right)(2\theta_B - 1)\, d\theta_B$$

$$
\begin{aligned}
&= \int_0^1 \int_0^1 y^f(\theta_S, \theta_B)(2\theta_B - 1)\, d\theta_S\, d\theta_B \\
&\quad + \int_0^1 \int_0^1 \int_0^1 \left(\eta^1 \left[y^f(\theta_S, \theta_B)(1 - y^f(\theta'_S, \theta_B)) - \lambda^1(1 - y^f(\theta_S, \theta_B))y^f(\theta'_S, \theta_B)\right]\right)(2\theta_B - 1)\, d\theta'_S\, d\theta_S\, d\theta_B \\
&= \int_0^1 \mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)](2\theta_B - 1)\, d\theta_B \\
&\quad + \int_0^1 \left(\eta^1 \left[\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)](1 - \mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)]) - \lambda^1(1 - \mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)])\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)]\right]\right)(2\theta_B - 1)\, d\theta_B \\
&= \int_0^1 (2\theta_B - 1)\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)] \left(1 + \eta^1 \left[(1 - \mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)]) - \lambda^1(1 - \mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)]\right]\right)\, d\theta_B \\
&= \int_0^1 (2\theta_B - 1)\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)] \left(\Lambda \left[\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)] - 1\right] + 1\right)\, d\theta_B.
\end{aligned}
$$

Proceeding analogously the second integral can be rewritten to

$$
\begin{aligned}
&\int_0^1 2\theta_S \tilde{v}_S(\theta_S)\, d\theta_S \\
&= \int_0^1 2\theta_S \mathbb{E}_B[y^f(\theta_S, \theta_B)] \left(1 - \Lambda \left(\mathbb{E}_B[y^f(\theta_S, \theta_B)] - 1\right)\right)\, d\theta_S.
\end{aligned}
$$

Thus, in summary we have

$$
\begin{aligned}
&\int_0^1 \tilde{v}_B(\theta_B)(2\theta_B - 1)\, d\theta_B - \int_0^1 2\theta_S \tilde{v}_S(\theta_S)\, d\theta_S \\
&= \int_0^1 (2\theta_B - 1)\mathbb{E}_S[y^f(\theta_S, \theta_B)] \left(\Lambda \left[\mathbb{E}_S[y^f(\theta_S, \theta_B)] - 1\right] + 1\right)\, d\theta_B \\
&+ \int_0^1 2\theta_S \mathbb{E}_B[y^f(\theta_S, \theta_B)] \left(\Lambda \left(\mathbb{E}_B[y^f(\theta_S, \theta_B)] - 1\right) - 1\right)\, d\theta_S.
\end{aligned}
$$

Step 3. Finally, we can replace the constraint IR by $\tilde{v}_S(1) \geq 0$ and $\tilde{v}_B(0) \geq 0$ by the following argument. Condition (ii) in Proposition 2 allows us to write utility as

$$
U_S(\theta_S, s_B^t|\theta_S) = U_S(1, s_B^t|1) + \int_{\theta_S}^1 \tilde{v}_S(t)\, dt, \tag{20}
$$

$$
U_B(\theta_B, s_S^t|\theta_B) = U_B(0, s_S^t|0) + \int_0^{\theta_B} \tilde{v}_B(t)\, dt. \tag{21}
$$

Recall that in Step 1 we chose transfers such that $U_S(1, \theta_B|1) = U_B(0, \theta_S|0) = 0$. Moreover, by condition (i) in Proposition 2 the functions $\tilde{v}_S$ and $\tilde{v}_B$ are non-increasing and non-decreasing, respectively. Thus, if $\tilde{v}_S(1) \geq 0$ and $\tilde{v}_B(0) \geq 0$ the integrands in equations (20) and (21) are positive on the domain of integration and therefore the integrals are positive. Thus, $\tilde{v}_S(1) \geq 0$ and $\tilde{v}_B(0) \geq 0$ jointly with CPEIC are equivalent to IR.

B.3 Proof of Proposition 5

Step 1. Given

$$y^f(\theta) = \begin{cases} 1 & \text{if } \theta_B - \theta_S > \delta, \\ 0 & \text{else,} \end{cases}$$

we find $\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)] = \max\{\theta_B - \delta, 0\}$ and $\mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)] = \max\{1 - \theta_S - \delta, 0\}$. Then, plugging in these expectations into the objective function in the problem (RM') the problem becomes (note the limits of integration)

$$\begin{aligned} \max_{\delta \geq 0} \ \pi(\Lambda) = & \int_{\delta}^{1} (2\theta_B - 1)(\theta_B - \delta) \left(\Lambda \left[(\theta_B - \delta) - 1\right] + 1\right) \, d\theta_B \\ & + \int_{0}^{1-\delta} 2\theta_S(1 - \theta_S - \delta) \left(\Lambda \left[(1 - \theta_S - \delta) - 1\right] - 1\right) \, d\theta_S. \end{aligned}$$

subject to $\tilde{v}_S$ being non-increasing, $\tilde{v}_B$ being non-decreasing, $\tilde{v}_S(1) \geq 0$ and $\tilde{v}_B(0) \geq 0$.

Step 2. We will now maximize the unconstrained problem and check the constraints afterwards. Solving the integrals yields

$$\begin{aligned} & \int_{\delta}^{1} (2\theta_B - 1)(\theta_B - \delta)(\Lambda(\theta_B - \delta - 1) + 1) \, d\theta_B \\ = & \int_{\delta}^{1} (2\theta_B^2 - 2\theta_B\delta - \theta_B + \delta) + \Lambda(2\theta_B^3 - 4\theta_B^2\delta - 3\theta_B^2 + 4\delta\theta_B + 2\theta_B\delta^2 - \delta^2 + \theta_B - \delta) \, d\theta_B \\ = & \ \frac{2}{3}\theta_B^3 - \theta_B^2\delta - \frac{\theta_B^2}{2} + \delta\theta_B + \Lambda\left(\frac{1}{2}\theta_B^4 - \frac{4}{3}\theta_B^3\delta - \theta_B^3 + 2\theta_B^2\delta + \theta_B^2\delta^2 - \delta^2\theta_B + \frac{1}{2}\theta_B^2 - \delta\theta_B\right)\Bigg|_{\delta}^{1} \\ = & \ \frac{1}{6} - \frac{1}{3}\Lambda\delta - \left(\frac{1}{2}\delta^2 - \frac{1}{3}\delta^3\right) - \Lambda\left(\frac{1}{6}\delta^4 - \frac{1}{2}\delta^2\right) \\ = & \ \frac{1}{6} - \frac{1}{2}\delta^2 + \frac{1}{3}\delta^3 - \Lambda\left(\frac{1}{6}\delta^4 - \frac{1}{2}\delta^2 + \frac{1}{3}\delta\right), \end{aligned}$$

and

$$\begin{aligned} & \int_{0}^{1-\delta} 2\theta_S(1 - \theta_S - \delta)(\Lambda(-\theta_S - \delta) - 1) \, d\theta_S \\ = & \int_{0}^{1-\delta} 2\delta\theta_S - 2\theta_S + 2\theta_S^2 + \Lambda\left(-2\delta\theta_S + 2\delta^2\theta_S - 2\theta_S^2 + 4\delta\theta_S^2 + 2\theta_S^3\right) \, d\theta_S \\ = & \ \delta\theta_S^2 - \theta_S^2 + \frac{2}{3}\theta_S^3 + \Lambda\left(-\delta\theta_S^2 + \theta_S^2\delta^2 - \frac{2}{3}\theta_S^3 + \frac{4}{3}\delta\theta_S^3 + \frac{1}{2}\theta_S^4\right)\Bigg|_{0}^{1-\delta} \\ = & \ (1 - \delta)^2\left(-\frac{1}{3} + \frac{1}{3}\delta\right) + \Lambda(1 - \delta)^2\left(-\frac{1}{6} + \frac{1}{6}\delta^2\right). \end{aligned}$$

Summing these two expressions yields

$$-\frac{1}{6}(\delta-1)^2(1+2\delta(\Lambda-2)+\Lambda),$$

and taking the derivative with respect to δ thereof we get

$$\delta^2(2-\Lambda)+\delta(\Lambda-3)+1.$$

Setting this derivative equal to zero yields a quadratic equation with two solutions. The first, $\delta=1$, is a minimum and the second, $\delta^{RM}=1/(2-\Lambda)$, is the maximum we are looking for, which is monotonically increasing in Λ.

Step 3. We proceed by showing that all constraints are satisfied when $\delta=\delta^{RM}$. First, we prove that $\tilde{v}_S$ is non-increasing and $\tilde{v}_S(1)\geq 0$. Recall that

$$\begin{aligned}\tilde{v}_S(\theta_S) &= \int_0^1 y^f(\theta_S,\theta_B)\,d\theta_B - \eta^1\int_0^1\int_0^1 \mu^1\left(y^f(\theta_S,\theta'_B)-y^f(\theta_S,\theta_B)\right)\,d\theta'_B\,d\theta_B\\ &= \begin{cases}1-\theta_S-\delta^{RM}+\Lambda(\theta_S+\delta^{RM})(1-\theta_S-\delta^{RM}) & \text{if } \theta_S\leq 1-\delta^{RM}\\ 0 & \text{else.}\end{cases}\end{aligned}$$

In particular, $\tilde{v}_S(1)=0$, proving $\tilde{v}_S(1)\geq 0$. Thus, what remains to be done is to check whether $\tilde{v}_S$ is non-increasing when $\theta_S\leq 1-\delta^{RM}$. We find that the derivative with respect to θ_S is given by $-1+\Lambda(1-2\theta_S-2\delta^{RM})$. We have

$$-1+\Lambda(1-2\theta_S-2\delta^{RM})\leq 0 \Leftrightarrow (\Lambda-1)(2-\Lambda)\leq 2\Lambda$$

and thus $\tilde{v}_S$ is indeed non-increasing since $\Lambda\leq 1$. Second, we show $\tilde{v}_B$ is non-decreasing and $\tilde{v}_B(0)\geq 0$. We have

$$\begin{aligned}\tilde{v}_B(\theta_B) &= \int_0^1 y^f(\theta_S,\theta_B)\,d\theta_S + \eta^1\int_0^1\int_0^1 \mu^1\left(y^f(\theta_S,\theta_B)-y^f(\theta'_S,\theta_B)\right)\,d\theta'_S\,d\theta_S\\ &= \begin{cases}\theta_B-\delta^{RM}-\Lambda(\theta_B-\delta^{RM})(1-\theta_B+\delta^{RM}) & \text{if } \theta_B\geq\delta^{RM}\\ 0 & \text{else.}\end{cases}\end{aligned}$$

In particular, $\tilde{v}_B(0)=0$, proving $\tilde{v}_B(0)\geq 0$. We still need to show that $\tilde{v}_B$ is non-decreasing when $\theta_B\geq\delta^{RM}$. Taking the derivative with respect to θ_B yields $1-\Lambda(1-2(\theta_B-\delta^{RM}))$. We then get

$$1-\Lambda(1-2(\theta_B-\delta^{RM})) = \underbrace{1-\Lambda}_{\geq 0}+2\Lambda\underbrace{(\theta_B-\delta^{RM})}_{\geq 0}\geq 0,$$

because we are considering the case $\theta_B \geq \delta^{RM}$.
Step 4. The revenue at the optimal value $\delta^{RM} = 1/(2-\Lambda)$ is given by

$$\begin{aligned}\pi^{RM}(\Lambda) &= -\frac{1}{6}\left(\frac{1}{2-\Lambda} - 1\right)^2 \left(1 + 2\frac{\Lambda - 2}{2-\Lambda} + \Lambda\right) \\ &= -\frac{1}{6}\left(\frac{\Lambda - 1}{2-\Lambda}\right)^2 \left(1 - 2\frac{2-\Lambda}{2-\Lambda} + \Lambda\right) \\ &= -\frac{(\Lambda-1)^3}{6(2-\Lambda)^2}.\end{aligned}$$

The derivative of this with respect to Λ is given by

$$\pi^{RM\prime}(\Lambda) = -\frac{(\Lambda-1)^2}{2(\Lambda-2)^2} + \frac{(\Lambda-1)^3}{3(\Lambda-2)^3},$$

and we find

$$\begin{aligned}\pi^{RM\prime}(\Lambda) = -\frac{(\Lambda-1)^2}{2(\Lambda-2)^2} + \frac{(\Lambda-1)^3}{3(\Lambda-2)^3} \leq 0 &\Leftrightarrow -3(\Lambda-2) + 2(\Lambda-1) \geq 0 \\ &\Leftrightarrow 4 - \Lambda \geq 0,\end{aligned}$$

where the last inequality is obviously true by the assumption $\Lambda \leq 1$, proving that the maximal revenue is decreasing in Λ.

B.4 Proof of Proposition 6

Step 1. 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 (5) and (6) from condition (ii), we can rewrite the objective function in the problem WM to

$$U_S(1, s_B^t|1) + U_B(0, s_S^t|0) + \int_0^1 \int_{\theta_S}^1 \tilde{v}_S(t)\, dt\, d\theta_S + \int_0^1 \int_0^{\theta_B} \tilde{v}_B(t)\, dt\, d\theta_B. \tag{22}$$

We still have to keep condition (i) as a separate constraint.
Step 2. Recall that we can write expected transfers as

$$\bar{t}_S(\theta_S) = U_S(\theta_S, \theta_B|\theta_S) + \theta_S \tilde{v}_S(\theta_S) - \eta^2 w_S(\theta_S), \tag{23}$$

$$\bar{t}_B(\theta_B) = -U_B(\theta_B, \theta_S|\theta_B) + \theta_B \tilde{v}_B(\theta_B) + \eta^2 w_B(\theta_B), \tag{24}$$

and note that we can rewrite the (AB) constraint as

$$\int_0^1 \bar{t}_S(\theta_S)\, d\theta_S - \int_0^1 \bar{t}_B(\theta_B)\, d\theta_B = 0. \tag{25}$$

Thus, given the CPEIC constraint and using equations (23) to (25) we can rewrite the (AB) constraint to

$$\begin{aligned} &\int_0^1 \left(\eta^2 w_B(\theta_B) + \theta_B \tilde{v}_B(\theta_B) - U_B(0, s_S^t|0) - \int_0^{\theta_B} \tilde{v}_B(t)\, dt\right) d\theta_B \\ &+ \int_0^1 \left(\eta^2 w_S(\theta_S) - \theta_S \tilde{v}_S(\theta_S) - U_S(1, s_B^t|1) - \int_{\theta_S}^1 \tilde{v}_S(t)\, dt\right) d\theta_S \\ &= 0. \end{aligned} \tag{AB'}$$

The rewritten objective function in (22) and (AB') allow us to set up a Lagrangian given by

$$\begin{aligned} \mathcal{L}(y^f, t_S^f, t_B^f, \gamma) = {} & (1-\gamma) U_S(1, s_B^t|1) + (1-\gamma) U_B(0, s_S^t|0) \\ & + (1-\gamma) \int_0^1 \int_{\theta_S}^1 \tilde{v}_S(t)\, dt\, d\theta_S + (1-\gamma) \int_0^1 \int_0^{\theta_B} \tilde{v}_B(t)\, dt\, d\theta_B \\ & + \gamma \int_0^1 \theta_B \tilde{v}_B(\theta_B)\, d\theta_B - \gamma \int_0^1 \theta_S \tilde{v}_S(\theta_S)\, d\theta_S \\ & + \gamma \int_0^1 \eta^2 w_B(\theta_B)\, d\theta_B + \gamma \int_0^1 \eta^2 w_S(\theta_S)\, d\theta_S, \end{aligned} \tag{26}$$

where γ is the Lagrange multiplier.

Step 3. By Lemma 1 we have $w_i(\theta_i) \leq 0$ for $i = S, B$, which enter the Lagrangian positively. In order to maximize the Lagrangian, we choose deterministic transfers which result in $w_i(\theta_i) = 0$ for $i = S, B$. Note that this is in line with the other constraints. Using integration by parts we can rewrite the Lagrangian to

$$\begin{aligned} \mathcal{L}(y^f, t_S^f, t_B^f, \gamma) = {} & (1-\gamma) U_S(1, s_B^t|1) + (1-\gamma) U_B(0, s_S^t|0) \\ & + (1-\gamma) \int_0^1 \theta_S \tilde{v}_S(\theta_S)\, d\theta_S + (1-\gamma) \int_0^1 (1-\theta_B) \tilde{v}_B(\theta_B)\, d\theta_B \\ & + \gamma \int_0^1 \theta_B \tilde{v}_B(\theta_B)\, d\theta_B - \gamma \int_0^1 \theta_S \tilde{v}_S(\theta_S)\, d\theta_S. \end{aligned}$$

Note that we still need to maximize with respect to t_S^f and t_B^f, as the transfer functions still show up in the terms $U_S(1, \theta_B|1)$ and $U_B(0, \theta_S|0)$.

Step 4. Mirroring Step 2 in the proof of Proposition 4 in Appendix B.2, we can extensively

rewrite the Lagrangian. As the steps are very similar we omit them this time around. The Lagrangian eventually reads

$$\begin{aligned}\mathcal{L}(y^f, t_S^f, t_B^f, \gamma) = (1-\gamma)U_S(1, s_B^t|1) + (1-\gamma)U_B(0, s_S^t|0) \\ + (1-2\gamma)\int_0^1 \theta_S \mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)]\left(1 - \Lambda\left(\mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)] - 1\right)\right)\, d\theta_S \\ + (1-\gamma)\int_0^1 (1-\theta_B)\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)]\left(1 + \Lambda\left(\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)] - 1\right)\right)\, d\theta_B \\ + \gamma\int_0^1 \theta_B \mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)]\left(1 + \Lambda\left(\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)] - 1\right)\right)\, d\theta_B.\end{aligned}$$

Thus, the problem is now to maximize this Lagrangian subject to condition (i) of Proposition 2 and IR.

B.5 Proof of Proposition 7

Step 1. We find $\mathbb{E}_S[y^f(\tilde{\theta}_S, \theta_B)] = \max\{\theta_B - \delta, 0\}$ and $\mathbb{E}_B[y^f(\theta_S, \tilde{\theta}_B)] = \max\{1 - \theta_S - \delta, 0\}$. We can then rewrite the objective function in the problem (WM'), i.e., the Lagrangian, to

$$\begin{aligned}\mathcal{L}(y^f, t_S^f, t_B^f, \gamma) = (1-\gamma)U_S(1, s_B^t|1) + (1-\gamma)U_B(0, s_S^t|0) \\ + (1-2\gamma)\int_0^{1-\delta} \theta_S(1-\theta_S-\delta)\left(1 - \Lambda\left((1-\theta_S-\delta) - 1\right)\right)\, d\theta_S \\ + (1-\gamma)\int_\delta^1 (1-\theta_B)(\theta_B-\delta)\left(1 + \Lambda\left((\theta_B-\delta) - 1\right)\right)\, d\theta_B \\ + \gamma\int_\delta^1 \theta_B(\theta_B-\delta)\left(1 + \Lambda\left((\theta_B-\delta) - 1\right)\right)\, d\theta_B.\end{aligned}$$

Next, note that by the definition of the class of δ-inefficient mechanisms we have $U_S(1, s_B^t|1) = \bar{t}_S(1)$ and $U_B(0, s_S^t|0) = -\bar{t}_B(0)$. Thus, instead of maximizing over t_S^f and t_B^f, we only have to maximize over $\bar{t}_S(1)$ and $-\bar{t}_B(0)$. We next calculate the value of the three integrals in the Lagrangian. The first one is given by

$$\begin{aligned}&\int_0^{1-\delta} \theta_S(1-\theta_S-\delta)\left(1 - \Lambda\left((1-\theta_S-\delta) - 1\right)\right)\, d\theta_S \\ &= \int_0^{1-\delta} \theta_S - \theta_S\delta - \theta_S^2 + \Lambda\left(\theta_S\delta - \theta_S\delta^2 + \theta_S^2\delta - 2\theta_S^2\delta - \theta_S^3\delta\right)\, d\theta_S \\ &= \frac{1}{2}\theta_S^2 - \frac{1}{2}\theta_S^2\delta - \frac{1}{3}\theta_S^3 + \Lambda\left(\frac{1}{2}\theta_S^2\delta - \frac{1}{2}\theta_S^2\delta^2 + \frac{1}{3}\theta_S^3\delta - \frac{2}{3}\theta_S^3\delta - \frac{1}{4}\theta_S^4\delta\right)\Bigg|_0^{1-\delta} \\ &= -\frac{1}{12}(\delta-1)^3(2 + \Lambda(1+\delta)).\end{aligned}$$

The second integral reads

$$\begin{aligned}
&\int_\delta^1 (1-\theta_B)(\theta_B-\delta)\left(1+\Lambda\left((\theta_B-\delta)-1\right)\right)\, d\theta_B \\
&= \int_\delta^1 \theta_B - \theta_B^2 - \delta + \theta_B\delta + \Lambda\left(-\theta_B + 2\theta_B^2 - \theta_B^3 + \delta - 3\theta_B\delta + 2\theta_B^2\delta + \delta^2 - \theta_B\delta^2\right)\, d\theta_B \\
&= \frac{1}{2}\theta_B^2 - \frac{1}{3}\theta_B^3 - \delta\theta_B + \frac{1}{2}\theta_B^2\delta + \Lambda\left(-\frac{1}{2}\theta_B^2 + \frac{2}{3}\theta_B^3 - \frac{1}{4}\theta_B^4 + \delta\theta_B - \frac{3}{2}\theta_B^2\delta + \frac{2}{3}\theta_B^3\delta + \delta^2\theta_B - \frac{1}{2}\theta_B^2\delta^2\right)\Big|_\delta^1 \\
&= \frac{1}{12}(\delta-1)^3(-2+\Lambda(1+\delta)).
\end{aligned}$$

Finally, the third integral yields

$$\begin{aligned}
&\int_\delta^1 \theta_B(\theta_B-\delta)\left(1+\Lambda\left((\theta_B-\delta)-1\right)\right)\, d\theta_B \\
&= \int_\delta^1 \theta_B^2 - \theta_B\delta + \Lambda\left(-\theta_B^2 + \theta_B^3 + \theta_B\delta - 2\theta_B^2\delta + \theta_B\delta^2\right)\, d\theta_B \\
&= \frac{1}{3}\theta_B^3 - \frac{1}{2}\theta_B^2\delta + \Lambda\left(-\frac{1}{3}\theta_B^3 + \frac{1}{4}\theta_B^4 + \frac{1}{2}\theta_B^2\delta - \frac{2}{3}\theta_B^3\delta + \frac{1}{2}\theta_B^2\delta^2\right)\Big|_\delta^1 \\
&= -\frac{1}{12}(\delta-1)^2(-4-2\delta+\Lambda(1+4\delta+\delta^2)).
\end{aligned}$$

Summing the values of the three integrals the Lagrangian reads

$$\begin{aligned}
\mathcal{L}(\delta, \bar{t}_S(1), \bar{t}_B(0), \gamma) = &-\frac{1}{6}(\delta-1)^2(-2+\gamma+\Lambda\gamma+2\delta(1+(\Lambda-2)\gamma)) \\
&+ (1-\gamma)\bar{t}_S(1) - (1-\gamma)\bar{t}_B(0).
\end{aligned}$$

We proceed by including the IR constraints using the Kuhn-Tucker method. By the same argument as in the proof to Proposition 5 in Appendix B.3 we can replace the IR constraint by $\bar{t}_S(1) \geq 0$ and $-\bar{t}_B(0) \geq 0$, because we still keep condition (i) from Proposition 2 and have $\tilde{v}_S(1) = 0$ and $\tilde{v}_B(0) = 0$. The Lagrangian then reads

$$\begin{aligned}
\mathcal{L}(\delta, \bar{t}_S(1), \bar{t}_B(0), \gamma) = &-\frac{1}{6}(\delta-1)^2(-2+\gamma+\Lambda\gamma+2\delta(1+(\Lambda-2)\gamma)) \\
&+ (1-\gamma)\bar{t}_S(1) - (1-\gamma)\bar{t}_B(0) + \alpha\bar{t}_S(1) - \beta\bar{t}_B(0)
\end{aligned} \tag{27}$$

where α and β are the multipliers of the constraints $\bar{t}_S(1) \geq 0$ and $-\bar{t}_B(0) \geq 0$, respectively.

Step 2. The respective derivatives of (27) are given by

$$\frac{\partial \mathcal{L}}{\partial \delta} = (1-\delta)(-1+\gamma+\delta(1+\gamma(\Lambda-2))) \tag{28}$$

$$\frac{\partial \mathcal{L}}{\partial \gamma} = -\frac{1}{6}(\delta - 1)^2(1 + 2\delta(\Lambda - 2) + \Lambda) - \bar{t}_S(1) + \bar{t}_B(0) \tag{29}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \bar{t}_S(1) \tag{30}$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = -\bar{t}_B(0) \tag{31}$$

$$\frac{\partial \mathcal{L}}{\partial \bar{t}_S(1)} = 1 - \gamma + \alpha \tag{32}$$

$$\frac{\partial \mathcal{L}}{\partial \bar{t}_B(0)} = \gamma - 1 - \beta \tag{33}$$

and we get the usual Kuhn-Tucker conditions. Now, suppose $\bar{t}_S(1) > 0$. Then, by the complementary slackness condition we get $\alpha = 0$. From (32) we then have $\gamma = 1$. Jointly with (28) this implies $\delta = 1$ or $\delta = 0$. Now, if $\delta = 1$, (29) implies $\bar{t}_S(1) = \bar{t}_B(0) > 0$ yielding a contradiction with $-\bar{t}_B(0) \geq 0$. If instead $\delta = 0$ equation (29) implies $\bar{t}_B(0) = 1/6(1 + \Lambda) + \bar{t}_S(1) > 0$, again yielding a contradiction. Thus, $\bar{t}_S(1) = 0$.

Next, suppose $\bar{t}_B(0) < 0$. Then, by the complementary slackness condition we get $\beta = 0$. From (33) we then have $\gamma = 1$. Jointly with (28) this implies $\delta = 1$ or $\delta = 0$. Now, if $\delta = 1$, (29) implies $\bar{t}_S(1) = \bar{t}_B(0) < 0$ yielding a contradiction with $\bar{t}_S(1) \geq 0$. If instead $\delta = 0$ equation (29) implies $\bar{t}_B(0) - 1/6(1 + \Lambda) = \bar{t}_S(1) < 0$, again yielding a contradiction.Thus, $\bar{t}_B(0) = 0$.

This finding simplifies the problem considerably, as we can restrict attention to the arguments δ and γ when maximizing. Setting equations (28) and (29) equal to zero and using $\bar{t}_S(1) = \bar{t}_B(0) = 0$, yields a system of two equations in two unknowns. The solution $\delta = 1$ (with any γ) is a minimum. The maximum we are after is achieved by the pair

$$(\delta^{WM}, \gamma^{WM}) = \left(\frac{1 + \Lambda}{2(2 - \Lambda)}, \frac{3}{2 - \Lambda} \right).$$

Further, plugging in the pair $(\delta^{WM}, \gamma^{WM})$ in the Lagrangian we get

$$\mathcal{L}(\delta^{WM}, \gamma^{WM}) = \frac{9}{8} \frac{(\Lambda - 1)^3}{(\Lambda - 2)^3},$$

which is decreasing in Λ.

Step 3. We still need to check whether the remaining constraints hold at this solution. Namely, we need to check condition (i) form Proposition 2. First, the function $\tilde{v}_S$ is non-increasing. Recall that

$$\begin{aligned} \tilde{v}_S(\theta_S) &= \int_0^1 y^f(\theta_S, \theta_B)\, d\theta_B - \eta^1 \int_0^1 \int_0^1 \mu^1 \left(y^f(\theta_S, \theta'_B) - y^f(\theta_S, \theta_B) \right) d\theta'_B\, d\theta_B \\ &= \begin{cases} 1 - \theta_S - \delta^{WM} + \Lambda(\theta_S + \delta^{WM})(1 - \theta_S - \delta^{WM}) & \text{if } \theta_S \leq 1 - \delta^{WM} \\ 0 & \text{else.} \end{cases} \end{aligned}$$

In particular, we have $\tilde{v}_S(1) = 0$. What remains to be done is to check whether $\tilde{v}_S$ is non-increasing when $\theta_S \leq 1 - \delta^{WM}$. In this case, the derivative with respect to θ_S is given by $-1 + \Lambda(1 - 2\theta_S - 2\delta^{WM})$. We find

$$-1 + \Lambda(1 - 2\theta_S - 2\delta^{WM}) \leq 0 \Leftrightarrow \Lambda(1 - \Lambda) \leq 1$$

and thus $\tilde{v}_S$ is indeed non-increasing since $\Lambda \leq 1$. Second, the function $\tilde{v}_B$ is non-decreasing. We have

$$\begin{aligned}\tilde{v}_B(\theta_B) &= \int_0^1 y^f(\theta_S, \theta_B)\, d\theta_S + \eta^1 \int_0^1 \int_0^1 \mu^1 \left(y^f(\theta_S, \theta_B) - y^f(\theta'_S, \theta_B) \right) d\theta'_S\, d\theta_S \\ &= \begin{cases} \theta_B - \delta^{WM} - \Lambda(\theta_B - \delta^{WM})(1 - \theta_B + \delta^{WM}) & \text{if } \theta_B \geq \delta^{WM} \\ 0 & \text{else.} \end{cases}\end{aligned}$$

In particular, $\tilde{v}_B(0) = 0$. We still need to show that $\tilde{v}_B$ is non-decreasing when $\theta_B \geq \delta^{WM}$. We find that the derivative with respect to θ_B is given by $1 - \Lambda(1 - 2(\theta_B - \delta^{WM}))$. We then get

$$1 - \Lambda(1 - 2(\theta_B - \delta^{WM})) = \underbrace{1 - \Lambda}_{\geq 0} + 2\Lambda \underbrace{(\theta_B - \delta^{WM})}_{\geq 0} \geq 0,$$

because we are considering the case $\theta_B \geq \delta^{WM}$, which completes the proof.