Bilateral Trade with Loss-Averse Agents*

Jean-Michel Benkert†

This version: July 2023
First version: November 2014

Abstract

We introduce expectations-based loss aversion, which can explain the empirically well-documented endowment and attachment effect, into the classical bilateral-trade setting (Myerson and Satterthwaite, 1983). We derive optimal mechanisms for different objectives and find that relative to no loss aversion, the designer optimally provides agents with partial insurance in the ownership dimension and with full insurance in the money dimension. Notably, the former is achieved either by increasing or decreasing the trade frequency, depending on the distribution of types. Finally, we show that the impossibility of inducing materially efficient trade persists with loss aversion.

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

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

*This paper is a revised version of the first chapter of my Ph.D. thesis submitted at the University of Zurich. I would like to thank Zoltán Balogh, Olivier Bochet, Juan Carlos Carbajal, Eddie Dekel, Jeff Ely, Samuel Häfner, Fabian Herweg, Heiko Karle, Botond Kőszegi, René Leal Vizcaíno, Igor Letina, Shangen Li, Shou Liu, Daniel Martin, Simon Martin, Konrad Mierendorff, Marc Möller, Oleg Muratov, Wojciech Olszewski, Marek Pycia, Anne-Katrin Roesler, Yuval Salant, Aleksei Smirnov, Gerhard Sorger, Ran Spiegler, Egor Starkov, Severin Wildhaber, Tom Wilkening, Peio Zuazo Garin, and seminar participants in Bern, Zurich, at the Workshop on Mechanism Design and Behavioural Economic in Glasgow, at the ZWE 2014 and ESEM 2016 for helpful comments. I am especially grateful to my supervisors Nick Netzer and Georg Nöldeke for their guidance as well as numerous comments and suggestions. I would like to thank the University of Basel and Northwestern University for their hospitality while some of this work was conducted and the UBS International Center of Economics in Society at the University of Zurich as well as the Swiss National Science Foundation (Doc.Mobility Grant P1ZHP1_161810) for financial support. All errors are my own.

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

1 Introduction

The bilateral-trade setting describes a simple, yet economically important situation, the study of which goes back at least to Coase (1960). Applications abound and range from friendly haggling at flea markets, over professional trading on the stock market, to inter-country negotiations intended to avoid wars. In its simplest formulation, there is a seller who owns a good and might be willing to sell it, and a buyer who might be interested in buying it. Different parts of the economic literature have approached this setting differently. In the field of mechanism design, we assume that the agents' valuation of the good is private information and study what outcomes a designer can achieve through different institutions. Famously, Myerson and Satterthwaite (1983) (henceforth, MS) have shown that some loss of efficiency is unavoidable under standard constraints. This impossibility result in efficiency constitutes a cornerstone within economics overall and highlights the restrictive nature of Coase's assumptions (Myerson, 1989).

In the empirical literature, especially through experiments, we have studied how people behave in such trade situations and how the institution affects behavior. Here, two notable effects have been documented: the endowment effect, going back to Thaler (1980), and, more recently, the attachment effect (Ericson and Fuster, 2011). The endowment effect tells us that ownership of the good drives up the seller's valuation of the good.[1] The attachment effect tells us that a buyer can get attached to a good she does not own (yet) and that this attachment drives up her valuation for it. Importantly, in models of bilateral trade with quasi-linear utility (e.g., Myerson and Satterthwaite, 1983) or risk aversion (e.g., Garratt and Pycia, 2023), these (behavioral) effects do not exist, and thus their implications cannot be studied. In contrast, both of these empirical findings can be explained by the model of expectations-based loss aversion by Kőszegi and Rabin (2006, 2007) (henceforth, KR), which builds on the seminal work by Kahneman and Tversky (1979).[2] In their model, essentially, people compare an outcome to some reference point given by their initial, rational expectations of the outcome. In the case of the buyer, for instance, the expectation that she will buy the good leads to her attachment to the good. The model by KR, thus, is a natural candidate to understand better these observed effects and their implications on trade situations theoretically.

This paper introduces expectations-based loss aversion into an otherwise standard mechanism-design approach to the bilateral-trade problem. More specifically, we augment the model by Myerson and Satterthwaite (1983), in which both agents have quasi-

[1] Recent anecdotal evidence suggests that the endowment effect may play a role in the trade of non-fungible tokens (NFTs), see Adar (2021).

[2] There is substantial empirical evidence of loss aversion, e.g., Fehr and Goette (2007), Post, van den Assem, Baltussen, and Thaler (2008), Crawford and Meng (2011) and Pope and Schweitzer (2011). In particular, see Ericson and Fuster (2014) for an excellent review on the role of loss aversion in explaining behavioral effects in exchange situations.

linear utility over ownership of the good and money, by allowing for both agents to have reference-dependent preferences as modeled in Kőszegi and Rabin (2006, 2007), employing their "choice-acclimating personal equilibrium" (CPE). We call the standard utility from ownership of the good and money *material utility*, and, in addition, introduce *gain-loss utility* for both, money and ownership of the good separately. The reference point, relative to which agents evaluate an outcome, is formed endogenously as the rational expectation over the outcome.[3] We introduce the formal framework in detail in Section 2, where we prove the revelation principle and characterize incentive-compatible mechanisms (Propositions 1 and 2).

We then begin our analysis in Section 3 by identifying the theoretical counterparts of the endowment and attachment effect. In Proposition 3, we show that in any incentive-compatible mechanism, loss aversion, through the attachment and the endowment effect, reduces the information rent of the buyer and increases the information rent of the seller, respectively. The fact that the information rents change is not surprising, given that we are changing the agents' preferences by introducing loss aversion. What is noteworthy, though, is that the difference in the information rent goes in opposite directions for the buyer and the seller.

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

[3]Ericson and Fuster (2011), Abeler, Falk, Goette, and Huffman (2011), Crawford and Meng (2011), Gill and Prowse (2012), Karle, Kirchsteiger, and Peitz (2015), and Bartling, Brandes, and Schunk (2015) provide evidence for the assumption that expectations determine the reference point. In contrast, see Heffetz and List (2014) and Gneezy, Goette, Sprenger, and Zimmermann (2017) for papers that show the limits of this. Heffetz (2021) provides a nuanced discussion of some conflicting evidence.

The result on the effect of loss aversion on the agents' information rent in Proposition 3 is of interest for two reasons. First, as noted, it formally identifies the theoretical counterparts of the attachment and endowment effect. Second, it suggests a connection to Myerson and Satterthwaite's impossibility result. The standard interpretation of the impossibility result is that the gains from trade cannot cover the information rents that accrue to the agents to ensure incentive compatibility, given the participation constraints and budget balance. Loss aversion's reduction of the buyer's information rent suggests a potential to mitigate the severity of the impossibility problem or even reverse it, thus enabling the implementation of materially-efficient trade. Indeed, Proposition 4 shows that the presence of a loss-averse buyer can mitigate the impossibility result in the sense that a lower subsidy would be needed to induce materially-efficient trade. However, a reversal is beyond reach.

In Section 4, we turn to the problem of designing optimal mechanisms and begin with the problem of maximizing the designer's revenue. In this application, the designer acts as a broker or intermediary, thus providing the agents with a platform to trade. The central theme is providing agents with insurance to reduce the ex-post variation of payoffs, which they dislike due to their aversion to losses. We show that in the presence of loss aversion, any revenue-maximizing mechanism features what we call *interim-deterministic transfers*, that is, the transfer of an agent is independent of the other agent's report and is thus deterministic given her type. This eliminates any ex-post variation in transfers, thus fully insuring agents in the money dimension and making loss-averse agents better off.

Turning to the optimal allocation rule, we first observe that it is not possible to obtain the optimal allocation rule by pointwise maximization as in MS because the agents' expected utilities endogenously depend on the mechanism through the reference point. We thus first show that the optimal allocation rule must take a particular form. Namely, holding fixed the buyer's type, if trade optimally takes place for some seller type, then the trade should also occur for all lower seller types (Lemma 1). This captures the intuitively appealing notion that trade should occur for buyers with high valuations and sellers with low valuations.

With this intermediate result in hand, we can reformulate the objective function such that pointwise maximization is once more applicable, and the problem boils down to comparing *modified* virtual types of the buyer and seller. However, even given the assumption of increasing virtual types, these modified virtual types need not be increasing. We can, however, iron the modified virtual types and thus derive the optimal mechanism.

We find that the optimal allocation rule reduces but does not necessarily eliminate ex-post variation in agents' payoffs, thus providing them partial insurance. Interestingly, this reduction in ex-post variation of payoffs is sometimes achieved by *increasing* the trade frequency and sometimes by *decreasing* the trade frequency relative to the benchmark

of no loss aversion. To gain some intuition, note that ex-post variation for an agent is the largest when the probability of trade is 1/2, as both outcomes are equally (un)likely. Thus, uncertainty reduction can be achieved by either moving towards more trade or less trade. It turns out that the optimal allocation rule increases the trade probability for high buyer types as buyer loss aversion increases and for low seller types as seller loss aversion increases. Thus, in the presence of loss aversion, the designer extracts rents from agents by reducing ex-post variation in payoffs and optimally does it in a way that aligns with material interests by inducing more trade for those types, leading to a higher profit margin. For an extreme example, consider the case of no buyer loss aversion and sufficiently high seller loss aversion. Then, for some distributions, the optimal allocation rule induces no trade unless the seller type is sufficiently low, in which case trade takes place with probability 1, yielding a bang-bang allocation rule.[4]

We close the revenue-maximizing mechanism analysis with some comparative statics (Corollary 2). To obtain a closed-form solution of the allocation rule, we consider uniformly distributed types on symmetric support, in which case the introduction of loss aversion leads to a reduction of the trade frequency for all types. One can then show that the designer reduces the trade probability as the stakes increase and provides agents full insurance by eliminating trade for sufficiently high stakes. Intuitively, as the stakes become larger, it becomes too costly to satisfy the participation constraints of loss-averse agents. Relatedly, in a study of online bargaining on eBay, Backus, Blake, Larsen, and Tadelis (2020, p. 1342) find that the probability of a sale taking place decreases with the listing price, suggesting that less trade occurs when the stakes are higher.[5]

Besides maximizing the designer's revenue, another natural question is how the designer can maximize the agents' gains arising from trade. In this application, the designer is still providing agents with a platform to trade but does so to maximize the resulting gains from trade. In the presence of loss aversion, one needs to clarify what the relevant welfare criterion is and how to handle gain-loss utility. The literature on behavioral welfare economics provides some guidance and allows us to distinguish between model-based and model-less approaches (Manzini and Mariotti, 2014). In a model-based approach (e.g., Benkert and Netzer, 2018; Rubinstein and Salant, 2012), the welfare criterion is developed based on an underlying theory (or a model) of mistakes. In contrast, in a model-less approach (e.g., Apesteguia and Ballester, 2015; Bernheim and Rangel, 2009), multiple inconsistent preferences are being aggregated into a welfare criterion solely based

[4]Note that such an allocation rule (illustrated in Figure 2b on page 24 below) yields a constant, strictly positive trade probability for buyers, who in this example are not loss averse and thus do not mind the ex-post variation in payoffs.

[5]Note that the dataset in Backus et al. (2020) encompasses a broad range of products of arguably very heterogenous value. Thus, unless sellers set disproportionally higher ask prices for more valuable products, the effect seems unlikely to be purely the result of a downward-sloping demand. Nevertheless, we cannot say whether the effect is driven by loss aversion.

on observed choices. Thus, the designer may take different stances on how to treat gain-loss utility when aiming to maximize gains from trade. Proceeding analogously as for the revenue-maximizing mechanism, we can derive the optimal mechanism for both when the designer wants to maximize only material gains from trade or total gains from trade (including gain-loss utility). In general, the optimal mechanisms for these two distinct objectives need not coincide. It turns out, however, that for the case of uniformly distributed types and symmetric degrees of loss aversion, the optimal mechanisms coincide, so it does not matter whether the designer considers loss aversion a mistake.

Overall, our findings suggest that trade platform designers should strive to take into account loss aversion. For instance, consider the goal of inducing materially-efficient trade by subsidizing the traders. If loss aversion is not adequately factored in, the subsidy may fall short, so the designer does not reach her goal, or the subsidy may be too high, leaving money on the table. Turning to the design of optimal trade platforms, our results highlight the importance of providing traders with insurance. This can be implemented relatively easily for the buyer as the "buy now" option on eBay illustrates: The seller offers to sell the product with probability one at a fixed price. This posted-price mechanism leaves no room for expectation-based loss aversion for the buyer, offering full insurance in both the money and the ownership dimension. Conversely, such a posted-price mechanism only partially insures the seller, as trade and payment depend on whether a buyer is found. However, the platform could leverage its data on past transactions to advise the seller on what posted price or what reserve price to set in an auction, thus providing the seller with insurance by trying to deliberately affect the seller's expectations and trade probabilities that way.[6] To get these decisions right, the platform could estimate the distribution of traders' valuations (see, e.g., Larsen, 2021) and their degrees of loss aversion (see, e.g., Karle et al., 2015).

1.1 Related literature

The literature on mechanism design with loss-averse agents is most closely related to our paper. Eisenhuth (2019) considers the problem of a risk-neutral seller who wants to maximize revenue by selling a good to loss-averse buyers. Using the framework of KR, he finds that the optimal auction is an all-pay auction with a reserve price when agents bracket narrowly. This result corresponds to our finding that transfers are interim deterministic in optimal mechanisms and, as one can show, extends beyond the auction and bilateral trade setting. Duraj (2018) considers mechanism design problems with agents who are loss averse to news utility; that is, agents' utility depends on changes

[6] The platform could even take a more active role as an intermediary and buy the good from the seller. Indeed, this is, for instance, what Apple and Samsung do when they offer customers the possibility to "trade-in" their old phones.

in their beliefs over the outcome as in Kőszegi and Rabin (2009). In an application to bilateral trade, he shows the robustness of the impossibility result in this setting.[7]

Also related is the (increasingly large) literature on behavioral industrial organization with loss-averse agents.[8] Rosato (2017) considers a sequential bargaining model with a risk-neutral seller and a loss-averse buyer.[9] Also within the framework of KR, but assuming wide bracketing, he shows that the buyer's loss aversion softens the rent-efficiency trade-off for the seller. As in the present paper, this is driven by the attachment effect: the buyer is willing to accept lower offers to avoid the risk of a breakdown of the negotiations.[10] In contrast to the present paper, neither Rosato (2017) (nor Eisenhuth, 2019, above) feature loss-averse sellers, but only loss-averse buyers.[11] Heidhues and Kőszegi (2014) and Rosato (2016) consider models with a monopolist selling to expectations-based loss-averse consumers. In both papers, the monopolist uses random prices to induce the attachment effect, increasing the consumers' willingness to pay and thus profits. In contrast, in the present paper, agents are already confronted with uncertainty due to the private nature of types, and there is no need to further "inject" randomness to induce the attachment or endowment effect. Indeed, the designer optimally insures agents fully against any variation in transfers and partially in the ownership dimension to reduce ex-post variation in payoffs.[12]

Our paper also relates to the extensive literature on the bilateral-trade problem, which has followed Myerson and Satterthwaite (1983). Arguably, the departure from the classical setting most closely related to our paper is to consider risk-averse agents. However, in contrast to loss aversion, risk aversion cannot explain the endowment and attachment effect. Early on, Chatterjee and Samuelson (1983) showed that when agents "become infinitely risk averse", all material gains from trade can be realized using a double-auction. More recently, Garratt and Pycia (2023) examine the bilateral trade problem relaxing the assumption that the agents have quasi-linear utility.[13] Allowing for risk aversion and wealth effects, they provide conditions for realizing all gains from trade. The impossibility

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

[8] See for instance Karle and Möller (2020) and the references therein.

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

[10] The attachment effect also plays a role in several other papers, among others Karle and Schumacher (2017) in a model of advertisement or in Rosato (2023) who proposes expectations-based loss aversion as an explanation for the "afternoon effect" observed in sequential auctions.

[11] Note that we make no symmetry assumptions on loss aversion across agents or the money and ownership dimensions, allowing for substantial generality in that regard.

[12] These two papers rely on the so-called unacclimating personal equilibrium and its refinement, the preferred personal equilibrium (PPE), an alternative equilibrium concept by KR, to the choice-acclimating personal equilibrium (CPE) we employ here. We discuss the alternative concept of PPE in the conclusion.

[13] See also the references in Garratt and Pycia (2023) for more work on the bilateral trade problem in the classic framework with quasi-linear utility following Myerson and Satterthwaite (1983).

result can be reversed in this setting because risk aversion and wealth effects give rise to additional gains from trade, which then suffice to cover the agents' information rents.[14] In contrast to Garratt and Pycia (2023) we do not attempt to establish whether efficient trade concerning the total gains from trade can be achieved but approach the problem as one of finding the mechanism which maximizes the gains from trade from an ex-ante perspective, finding that it generally matters whether one wants to maximize total or only material gains from trade.

Moving to more behavioral models of the bilateral-trade problem, Crawford (2021) and Kneeland (2022) consider level-k agents. They show that depending on the assumptions regarding the observability of the agents' levels, the impossibility result can be reversed. Wolitzky (2016) considers max-min preferences and also obtains a reversal of the impossibility result. Neither of these papers considers the problem of deriving revenue-maximizing mechanisms as we do here, although Crawford (2021) derives second-best mechanisms for the maximization of gains of trade for a restricted type space. Finally, in the case of intentions-based social preferences (Bierbrauer and Netzer, 2016) as well as for altruism (Kucuksenel, 2012), the impossibility result can be reversed, too.

2 Model

2.1 Utility, Social Choice Functions and Mechanisms

The set of agents is given by $I = \{S, B\}$ where S and B denote seller and buyer, respectively. It is commonly known that the type of agent $i \in I$ has distribution F_i with full support on the set $\Theta_i = [a_i, b_i] \subset \mathbb{R}_+$, and is private information. Let $\Theta = \Theta_S \times \Theta_B$ and assume that Θ_S and Θ_B have a non-trivial intersection. We interpret the type of an agent as her valuation of the good.[15] A social alternative is given by $\mathbf{x} = (y, t_S, t_B) \in X = \{0, 1\} \times \mathbb{R}^2$, where y indicates whether or not trade takes place and t_S and t_B denote the respective transfers of the seller and buyer.

Following KR, we allow the agents to be loss averse in the ownership and money dimensions. That is, the buyer, for example, derives the standard material utility from

[14] In contrast to Garratt and Pycia (2023), we obtain quasi-linear utility due to narrow-bracketing of gain-loss utility and having piece-wise linear value functions. Thus, the relaxation of quasi-linear utility, which gives rise to the possibility result in their paper, is not present in our framework. The narrow-bracketing assumption also sets the current setting apart from that in Gershkov, Moldovanu, Strack, and Zhang (2021), who study optimal auction design when agents have constant relative risk aversion. They find that agents' utility is, in the language of the present paper, interim deterministic, i.e., does not depend on other agents' reports. At the same time, we only obtain interim deterministic transfers with narrow bracketing. The setting in Gershkov et al. (2021) is more closely related to the part in Eisenhuth (2019) with wide bracketing.

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

obtaining and paying for the good, and additionally, the buyer feels weighted gain-loss utility concerning getting the good as well as weighted gain-loss utility concerning paying for the good. Loss aversion is captured by value functions in the sense of Kahneman and Tversky (1979) given by

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

for some $\lambda_i^k > 1$, which reflect the degree of loss aversion in the dimensions of ownership and money, respectively.[16] Thus, the riskless total utility is given by

$$u_S(\mathbf{x}, \mathbf{r}_S, \theta_S) = \underbrace{(1-y)\theta_S + t_S}_{\text{material utility}} + \underbrace{\eta_S^1 \mu_S^1 \left(r_S^1 \theta_S - y\theta_S\right)}_{\text{gain-loss utility in ownership}} + \underbrace{\eta_S^2 \mu_S^2 (t_S - r_S^2)}_{\text{gain-loss utility in money}} \tag{1}$$

$$u_B(\mathbf{x}, \mathbf{r}_B, \theta_B) = \underbrace{y\theta_B - t_B}_{\text{material utility}} + \underbrace{\eta_B^1 \mu_B^1 (y\theta_B - r_B^1 \theta_B)}_{\text{gain-loss utility in ownership}} + \underbrace{\eta_B^2 \mu_B^2 (r_B^2 - t_B)}_{\text{gain-loss utility in money}} \tag{2}$$

where $\eta_i^k \geq 0$ are the weights put on gain-loss utility and $\mathbf{r}_i = \{r_i^1, r_i^2\} \in \mathbb{R}^2$ are the so-called riskless reference levels. Following KR, we will allow the reference point to be the agent's rational expectations and, therefore a probability distribution over all riskless reference levels (see more below).

The model by KR has several moving parts, so we devote the following paragraph to discussing several (implicit) assumptions. First, as noted in equations (1) and (2), we distinguish between material utility and gain-loss utility in the dimensions of ownership and money. We follow most of the literature working with the model by KR and adopt the following assumption by Herweg, Müller, and Weinschenk (2010) in the ownership dimension.[17]

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

As KR noted, this condition ensures that agents will not choose stochastically dominated options. Essentially, we need the assumption to ensure incentive compatibility and will

[16] We follow the literature by abstracting from diminishing sensitivity. This assumption is not needed for gain-loss utility in the money dimension. For instance, all the proofs go through directly if we assume $\mu_i^2(x) = g(x)$ if $x \geq 0$, and $\mu_i^2(x) = -\lambda_i^2 g(-x)$ if $x < 0$, for some concave function g. In the ownership dimension, however, we cannot dispense of the piece-wise linearity, as this ensures that the expected utility remains linear in the agents' type in the presence of loss aversion.

[17] Eisenhuth and Grunewald (2018) and Banerji and Gupta (2014) estimate the parameter Λ_i for bidders in auctions at 0.42 and 0.283, respectively, statistically different from 0 and 1 at all conventional significance levels. Further, this condition is commonly imposed, see for instance de Meza and Webb (2007), Eisenhuth and Grunewald (2018), Eisenhuth (2019), Karle and Peitz (2014), and Gershkov et al. (2021). For examples not adopting the assumption, see Meisner and von Wangenheim (2021), Dreyfuss, Heffetz, and Rabin (2022), and Rosato (2023).

discuss its role when stating our results.[18] Further, we follow KR by assuming "narrow bracketing" i.e., we assume a separate gain-loss term for each of the two material utility dimensions, ownership, and money utility. Without this assumption, the endowment and attachment effect that motivate this paper, could not materialize, making it conceptionally crucial to our analysis. Further, the assumption is very well-supported empirically.[19] Beyond this conceptual point and the empirical evidence for it, the assumption is important in our setting, as it allows us to maintain quasi-linearity in the presence of loss aversion.[20] Finally, the assumption that the loss aversion parameters are commonly known may seem restrictive. However, we are essentially assuming that the functional form of the utility function is common knowledge and that all private information pertains to the agents' valuation of the good. We are following, for instance, Maskin and Riley (1984) who assume in their study of optimal auctions with risk-averse buyers that the buyers' parameter of risk aversion is commonly known. We briefly discuss relaxing the assumption in the conclusion.

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 current 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.

2.2 Equilibrium Concept and Revelation Principle

We use the concept of an (interim) choice-acclimating personal equilibrium (CPE) introduced in Kőszegi and Rabin (2007). However, as outlined in the conclusion, several of our results also hold when employing KR's alternative solution concept, the preferred personal equilibrium (PPE).[21] The set of all riskless reference levels is given by the set of all social

[18] Note that the assumption applies only to gain-loss utility in the ownership dimension, while no restrictions are placed on the money dimension.

[19] Narrow bracketing, or mental accounting, as it is also called, goes beyond the endowment effect Thaler (see e.g., 1999). See also Ellis and Freeman (2021) for more recent and very compelling evidence for narrow bracketing.

[20] As discussed in the literature review above (see footnote 14 in particular), this is a key distinction to the models considering risk aversion, which implicitly corresponds to wide bracketing.

[21] In the PPE, the agent "maximizes expected utility taking the reference point as given", whereas in the CPE, the agent "maximizes expected utility given that it determines both the reference lottery

alternatives X. Essentially, the set X captures all the outcomes that could materialize at the end of the agents' interaction. In a mechanism Γ, agent i's action induces a distribution over the set of social alternatives X, conditional on the other agent playing s_{-i}. This endogenously generated distribution over X forms the agents' reference point, or rather, reference distribution in a CPE. Effectively, when an agent evaluates an outcome, she compares it to all other possible social alternatives that could have materialized given the distribution induced over them. Moreover, when the agent takes an action in a CPE, she takes the action anticipating that it will not only determine the mechanism's outcome, but also the distribution over the set X and, therefore, the reference point.

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

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

The expression in (3) may require some explanation. The first line corresponds to material utility, the second to gain-loss utility in the ownership dimension, and the third to gain-loss utility in the money dimension. The double integral has a clear intuition. To illustrate, consider the last line containing the money gain-loss utility. Fix any θ_B in the domain of integration of the outer integral and suppose this was the actual realization of the buyer's type. The seller would then receive a transfer of $t_S^g(m, s_B(\theta_B))$, which she would compare to the reference point. The reference point is induced endogenously and corresponds to the distribution of possible transfers. Thus, for every θ'_B in the domain of the inner integral, we get a possible transfer $t_S^g(m, s_B(\theta'_B))$ given the buyer's strategy and the seller's message. The seller compares the actual transfer $t_S^g(m, s_B(\theta_B))$ with all these

and the outcome lottery". KR note that the CPE is more appropriate when the uncertainty is resolved after the agent's decision. We thus believe that the CPE is the more natural equilibrium concept in our context, as an agent's report determines the uncertainty she feels about the outcome given her beliefs about the other agent's type. In the conclusion, we discuss how much our results extend to the PPE.

other possible transfers, and the value function μ_S^2 weights these comparisons differently, depending on whether they result in a loss or a gain. The inner integral then aggregates the gains and losses weighted by the induced probability distribution. Next, integrate over all the values θ_B in the domain of the outer integral to get the familiar interim expected utility. In summary, the seller aggregates over each possible realization of transfers, and for each of these possible realizations, she compares the outcome with all other possible outcomes, aggregating gains and losses in each comparison.

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

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

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

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

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

We can now define our equilibrium concept, which follows Eisenhuth (2019).[22]

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

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

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

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

As a first result, we note that the revelation principle for CPE holds in our setting.[23]

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

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

2.3 Incentive Compatibility and Efficiency

In this section, we characterize the set of all CPEIC social choice functions and introduce some familiar concepts, such as individual rationality and ex-post budget balance. Further, we introduce our notion of an interim deterministic mechanism.[24]

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

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

(ii) we can write utility as

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

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

[23] Proofs are relegated to the appendix unless noted otherwise.

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

Recall that the functions $\tilde{v}_B$ and $\tilde{v}_S$ contain terms of gain-loss utility. However, even in the presence of loss aversion, utility is linear in the type, which means that this standard result goes through directly. Note, however, that loss aversion changes the agents' preferences compared to the case with no loss aversion, so the set of incentive-compatible SCF need not be the same with loss aversion as without.

Definition 4 *We say that an SCF is individually rational if for both agents $i \in I$*

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

Setting the outside option in (IR) equal to zero is without loss of generality.[25] An agent could 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 occur if she walks away. Consequently, there would be no feelings of gain or loss and zero material utility. Finally, the following definition will be necessary to state our results.

Definition 5 *We say that a mechanism has interim-deterministic transfers when, given her type, an agent's transfer does not depend on almost all types of the other agent, that is, for all $\theta_i \in \Theta_i$, $t_i(\theta_i, \theta_j)$ is constant for almost all $\theta_j \in \Theta_j$.*

3 Attachment, Endowment and Information Rents

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

$$IR_B(\theta_B) := \int_{a_B}^{\theta_B} \tilde{v}_B(t)\, dt \tag{6}$$

$$IR_S(\theta_S) := \int_{\theta_S}^{b_S} \tilde{v}_S(t)\, dt, \tag{7}$$

and obtain the following result.

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

Proposition 3 *In any CPEIC mechanism, the seller's information rent $IR_S(\theta_S)$ is increasing in Λ_S and the buyer's information rent $IR_B(\theta_B)$ is decreasing in Λ_B.*

The proof is straightforward. It suffices to take the derivatives with respect to Λ_S and Λ_B from equations (6) and (7), respectively. The key step is to note that

$$\begin{aligned}
&\tilde{v}_B(\theta_B) \\
&= \int_{a_S}^{b_S} y^f(\theta_S, \theta_B)\, dF_S(\theta_S) + \eta_B^1 \int_{a_S}^{b_S} \int_{a_S}^{b_S} \mu_B^1 \left(y^f(\theta_S, \theta_B) - y^f(\theta'_S, \theta_B) \right)\, dF_S(\theta'_S)\, dF_S(\theta_S), \\
&= y_B(\theta_B) \\
&+ \eta_B^1 \int_{a_S}^{b_S} \int_{a_S}^{b_S} y^f(\theta_S, \theta_B)(1 - y^f(\theta'_S, \theta_B)) - \lambda_B^1 (1 - y^f(\theta_S, \theta_B)) y^f(\theta'_S, \theta_B)\, dF_S(\theta'_S)\, dF_S(\theta_S), \\
&= y_B(\theta_B)(1 - \Lambda_B(1 - y_B(\theta_B))) \qquad (8)
\end{aligned}$$

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

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

Motivated by the empirical evidence, we interpret the increase in the seller's information rent as the endowment effect and to decrease in the buyer's information rent as the attachment effect. As noted, the respective decrease and increase in the information rents stem from the loss aversion in the ownership dimension. As expected, loss aversion on the money dimension plays no role here. To simplify exposition, we will use the term loss aversion as referring to loss aversion in the ownership dimension unless stated differently.

We want to emphasize that in the present model of expectations-based loss aversion, agents will typically anticipate both feelings of loss *and* gain at the interim stage. This can be seen nicely in the derivation of equation (8). From the buyer's perspective, reporting a type will give rise to a trade probability so that with some probability, trade will take place (resulting in a feeling of gain), and with the converse probability, no trade materializes (resulting in a feeling of loss). Importantly, loss aversion means that this uncertainty over the eventual payoff decreases expected utility in the ownership dimension on average, as the potential loss is felt more strongly than the potential gain.

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

Corollary 1 *The strength of the endowment and attachment effect is increasing and decreasing in the type of buyer and seller, respectively.*

The result follows immediately as one takes the derivative with respect to Λ_B and

θ_B from equation (6) and with respect to Λ_S and θ_S from equation (7) so that a proof is omitted. The finding about the impact of the endowment and attachment effect on the information rents suggests that the presence of a loss-averse buyer could enable the designer to implement materially-efficient trade subject to ex-post budget balance and the agents' participation constraints,[26] that is, to "reverse" the impossibility result by Myerson and Satterthwaite (1983). To see this, recall the interpretation of the result, stating that the gains from trade do not suffice to cover the agents' information rents. Thus, seller loss aversion, which increases the information rent, will make the problem only harder, while buyer loss aversion could make it easier. However, there is a countervailing effect even for the buyer. As noted above, loss aversion decreases expected utility in the ownership dimension as agents dislike ex-post variation in payoffs and hence makes satisfying the participation constraints harder, too.[27] Thus, loss aversion makes it harder to satisfy the participation constraints of both agents and the seller's incentive-compatibility constraint but slackens the buyer's incentive-compatibility constraint. Consequently, it suffices to consider buyer loss aversion to check whether the impossibility result can be reversed. Using this insight, we can proceed analogously to the proof in Myerson and Satterthwaite (1983). That is, impose budget balance and incentive compatibility to obtain an expression for the sum of utilities of the "worst" buyer and seller types in the materially-efficient mechanism and show that it is strictly negative. Indeed, we obtain

$$\begin{aligned} &U_B(a_B) + U_S(b_S) = \\ &- \int_{\max\{a_B,a_S\}}^{\min\{b_S,b_B\}} (1 - F_B(x))F_S(x)(1 - \Lambda_B(1 - F_S(x))) + \Lambda_B(1 - F_S(x))F_S(x)x f_B(x)\, dx \qquad (9) \\ &< 0, \end{aligned}$$

which violates individual rationality for any $\Lambda_B \leq 1$, proving our next result.

Proposition 4 *Given CPEIC, individual rationality, and ex-post budget balance, realizing all material gains from trade for any degree of buyer and seller loss aversion in the money or ownership dimension is impossible.*

The minimal subsidy needed to induce materially efficient trade under CPEIC and IR in equation (9) can be interpreted as a measure of the severity of the impossibility problem and will generally depend on the degree of loss aversion and the distribution of the agents' types. Indeed, taking the derivative of the minimal subsidy in equation (9) with

[26] Ex-post budget balance corresponds to the condition $t_S^f(\theta_S, \theta_B) = t_B^f(\theta_S, \theta_B), \quad \forall(\theta_S, \theta_B) \in \Theta$.

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

respect to Λ_B, we can see that the attachment effect mitigates the impossibility problem by dominating the diminishing effect of loss aversion on the participation constraints whenever

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

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

Another noteworthy point is that loss aversion does not matter for the extreme types, i.e., those who lie outside the intersection of the intervals, loss aversion does not matter. This finding is very intuitive: observe that trade is interim deterministic for these types and hence there is no gain-loss utility as there is no room for ex-post variations in payoffs. Put differently, expectations-based loss aversion only has bite when there is unresolved uncertainty, which is only the case for types lying strictly in the intersection of the type spaces.[28]

The fact that the impossibility result is not reversed is linked to the assumption that $\Lambda_B \leq 1$, i.e., that gain-loss utility does not dominate for the buyer. For instance, when types are drawn from $[0,1]$ with distributions $F_S(x) = x$ and $F_B(x) = x^{10}$ the subsidy in equation (9) turns into a surplus for $\Lambda_B \geq 13/3$. However, in this example, $\Lambda_B \leq 1$ is necessary for the materially-efficient mechanism to be incentive compatible for the buyer. Hence, incentive compatibility limits the feasible degree of loss aversion and, consequently, on the strength of the attachment effect, meaning that the impossibility result cannot be reversed. Yet, as we will discuss next, $\Lambda_B \leq 1$ is, in general only a sufficient condition for incentive compatibility and only sometimes necessary.

The assumption that $\Lambda_i \leq 1$ is commonly imposed in the literature for conceptual as well as technical reasons and appears supported empirically (see footnote 17). In particular, KR showed that the assumption ensures that agents do not choose stochastically

[28] One should note, however, that the presence of types strictly outside of the intersection affects the reference points of those within the intersection.

dominated options. In the present context, it is easy to show that the assumption is sufficient for the materially efficient allocation rule to be incentive compatible in the presence of loss aversion. Moreover, whenever $F_S(a_B) = 0$, the assumption is sufficient and necessary. Whenever the smallest buyer type has a zero probability of trading, the materially efficient trading rule is CPEIC if and only if $\Lambda_B \leq 1$. In particular, this is true when the types of both agents are drawn from the same support. It turns out, however, that when $F_S(a_B) > 0$, the assumption is no longer necessary.[29] Indeed, when $F_S(a_B) < 1/2$ the necessary condition reads $\Lambda_B \leq 1/(1 - 2F_S(a_B))$ and when $F_S(a_B) \geq 1/2$ no restrictions need to be put on Λ_B. In light of the above result, the question thus arises whether the impossibility result persists when $F_S(a_B) > 0$ and the assumption is relaxed, as this would allow us to strengthen the attachment effect and possibly set the required subsidy in equation (9) equal to zero.

To this end, one can show that the impossibility result continues to hold for $\Lambda_B \leq 1/(1 - F_S(a_B))$. This condition ensures that the lowest buyer type a_B is, in fact, the "worst" buyer type. For $\Lambda_B > 1/(1 - F_S(a_B))$, the worst buyer type is some intermediate type, and the above approach to proving the impossibility result fails: if the lowest buyer type is no longer the worst type, satisfying individual rationality for the lowest buyer type does no longer guarantee satisfying individual rationality for all types. The observation that an intermediate type is the worst type is reminiscent of the related model of partnership dissolution (Cramton, Gibbons, and Klemperer, 1987; Fieseler, Kittsteiner, and Moldovanu, 2003). In this model, the good is initially not exclusively owned by one agent only, but by several agents. As a result, the worst type of an agent may be an intermediate type. However, despite this similarity, the approach taken in that model cannot be extended to the present context due to the endogeneity of the reference point. In sum, although counterexamples have proved elusive, a reversal of the impossibility for when $\Lambda_B > 1/(1 - F_S(a_B))$ cannot be ruled out. Note, however, that for sufficiently strong loss aversion, the total gains from trade disappear completely. Thus, even if the buyer's information rent can be reduced using the attachment effect, impossibility will arise for sufficiently strong loss aversion because it will eliminate all the total gains from trade.[30]

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

[30] In the above, we have only discussed the degree of loss aversion of the buyer. Analogous arguments regarding the necessity and sufficiency of $\Lambda_S \leq 1$ for incentive compatibility of the seller apply. However, as loss aversion on the seller's side makes the impossibility problem only harder, it does not enter our result.

4 Optimal Mechanisms

The preceding section has formally identified the endowment and the attachment effect in an otherwise standard bilateral trade setting. In particular, we have seen how loss aversion impacts the agents' information rents and the participation constraints, allowing us to show that the impossibility of implementing materially efficient trade extends to the setting with loss-averse agents. We now turn to the problem of designing optimal mechanisms. We begin by considering the problem of maximizing the designer's revenue and then turn to the (conceptually) more nuanced question of maximizing the gains from trade. In contrast to the previous section, we assume symmetric support for the distributions of buyer and seller types to simplify notation.[31] We will continue to allow for arbitrary distributions but will make the standard regularity assumption of increasing virtual types.

4.1 Maximizing the Designer's Revenue

The revenue-maximizing designer's problem reads

$$\max_{(y^f,t_S^f,t_B^f)\in\mathcal{F}} \int_a^b \int_a^b \left(t_B^f(\theta_S,\theta_B) - t_S^f(\theta_S,\theta_B)\right)\, dF_S(\theta_S)\, dF_B(\theta_B),$$

subject to CPEIC and IR. (RM)

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

$$\begin{aligned}&\int_a^b \left(\eta_B^2 w_B(\theta_B) + \theta_B \tilde{v}_B(\theta_B) - \int_a^{\theta_B} \tilde{v}_B(t)\, dt\right)\, dF_B(\theta_B)\\ &+ \int_a^b \left(\eta_S^2 w_S(\theta_S) - \theta_S \tilde{v}_S(\theta_S) - \int_{\theta_S}^b \tilde{v}_S(t)\, dt\right)\, dF_S(\theta_S). \qquad (10)\end{aligned}$$

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

$$w_S(\theta_S) = \int_a^b \int_a^b \mu_S^2 \left(t_S^f(\theta_S,\theta_B) - t_S^f(\theta_S,\theta_B')\right)\, dF_B(\theta_B')\, dF_B(\theta_B).$$

[31] All arguments go through analogously for the case with asymmetric supports.

This expression and its analog for the buyer collect all gain-loss utility with respect to money and thus the objective function still depends on transfers. Nevertheless, the problem can be reduced to only choosing the optimal allocation rule, because in any optimal mechanism, the transfers of the seller will be interim deterministic and thus not depend on the buyer's type, and vice versa, so that $w_i(\theta_i) = 0$.

Proposition 5 *Any solution to the revenue maximization problem* (RM) *entails interim-deterministic transfers.*

Intuitively, loss-averse agents dislike ex-post variations in their payoffs. By making the transfers independent of the other agent's type, the designer completely insures the agents from any ex-post variation in the transfers. Thus, starting from any mechanism with non-interim-deterministic transfers, the designer can extract more surplus from the agents by choosing appropriate interim-deterministic transfers, effectively selling insurance to the agents. Note that interim-deterministic transfers are also *a* solution in the absence of loss aversion. However, in the presence of loss aversion interim-deterministic transfers are the *only* solution.[32]

Proposition 5 allows us to rewrite the maximization problem as

$$\begin{aligned}\max_{y^f \in \mathcal{Y}} & \int_a^b J_B(\theta_B) y_B(\theta_B)(1 - \Lambda_B\,(1 - y_B(\theta_B))\, f_B(\theta_B) d\theta_B \\ & - \int_a^b J_S(\theta_S) y_S(\theta_S)(1 + \Lambda_S(1 - y_S(\theta_S))) f_S(\theta_S) d\theta_S \qquad \text{(RM')} \\ & \text{subject to } y_B(\theta_B) \text{ being non-decreasing and } y_S(\theta_S) \text{ being non-increasing,}\end{aligned}$$

where $y_B(\theta_B) = \int_a^b y^f(\theta_S, \theta_B)\ dF_S(\theta_S)$ and $y_S(\theta_S) = \int_a^b y^f(\theta_S, \theta_B)\ dF_B(\theta_B)$ denote the interim trade probabilities of the buyer and seller, respectively, and

$$J_B(\theta_B) = \theta_B - \frac{1 - F_B(\theta_B)}{f_B(\theta_B)}, \quad J_S(\theta_S) = \theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}$$

denote the buyer's and the seller's virtual types. We make the following standard assumption.

Assumption 2 (Regularity) *The virtual types J_B and J_S are strictly increasing.*

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

The designer faces the trade-off that inducing trade comes at a cost in the form of the payment due to the seller and with a benefit in the form of the payment from the buyer. Further, the form of the objective function in (RM') suggests that even in the presence of loss aversion the designer wants to induce trade between high buyer and low seller types in particular. Put differently, the designer wants to buy the good from a low-value seller and sell it to a high-value buyer, as this yields a large profit margin. However, as a consequence of expectations-based loss aversion, it matters for an agent's utility whether trade takes place with only a few or many types of the other agent, as this affects her expectations, which in turn affects the strength of the endowment and attachment effect. Thus, there are in some sense externalities between the outcomes of different types.

Indeed, because the agents' expected utilities endogenously depend on the mechanism through the reference point, point-wise maximization of the objective function is not possible. To see this, rewrite equation (RM') to

$$\int_a^b \int_a^b \Big[J_B(\theta_B)(1-\Lambda_B) - J_S(\theta_S)(1+\Lambda_S) \Big] y^f(\theta_B,\theta_S) dF_B(\theta_B) dF_S(\theta_S)$$
$$+ \Lambda_B \int_a^b y_B(\theta_B)^2 dF_B(\theta_B) + \Lambda_S \int_a^b y_S(\theta_S)^2 dF_S(\theta_S).$$

In the first line, we have made use of the linearity in the y_B and y_S terms to "move out" the integral from within these terms. In the second line, however, we cannot do this as y_B and y_S enter quadratically. Hence, a pointwise maximization over the ex-post allocation rule y^f is not feasible at this stage. To get rid of the quadratic terms, we first show that the optimal allocation rule takes a particular form.

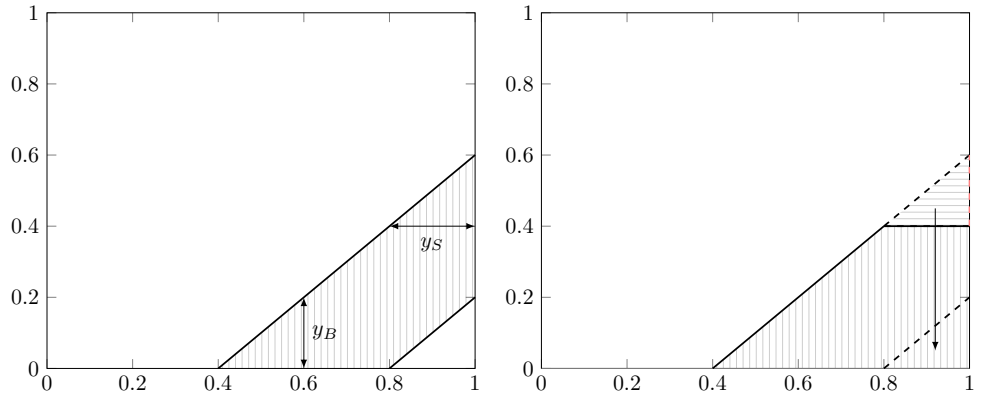
Lemma 1 *The solution to the problem* (RM') *can be written as*

$$y^f(\theta_B,\theta_S) = \begin{cases} 1 & 0 \leq \theta_S \leq \theta_S^*(\theta_B) \\ 0 & \textit{otherwise,} \end{cases} \tag{11}$$

for some function $\theta_S^* : [a,b] \to [a,b]$.

The above lemma has a straightforward interpretation. Fix a buyer type θ_B and suppose that it is optimal to induce trade for some seller type $\theta_S^*(\theta_B)$. Then, it is optimal to also induce trade for all seller types θ_S that are lower, i.e., for all $\theta_S \leq \theta_S^*(v_B)$. This reflects the intuition that the designer would like to induce trade with low seller types, as they will be willing to give up the good at a low price, yielding a high profit margin.

The proof of the lemma proceeds as follows. First, suppose some (ex-post) allocation rule $\hat{y}$ is optimal. This allocation rule gives rise to expected trade probabilities for buyer and seller, where $\hat{y}_B$ denotes the buyer's trade probability. From there, construct a new



(a) Original allocation rule (b) Constructed allocation rule

Figure 1: The above figures illustrate the construction of allocation rules in the proof of Lemma 1. Buyer and seller types are depicted on the x and y axes, respectively and trade takes place in the vertically-shaded areas. Panel (a) shows an allocation rule which satisfies CPEIC but is not of the form in Lemma 1. The arrows denoted by y_B and y_S indicate how to deduce the expected trade probabilities $y_B(\theta_B)$ and $y_S(\theta_S)$. To show that this allocation rule is not optimal, we compare it with the allocation rule in Panel (b). To obtain this allocation rule "shift" all trade probability from the horizontally-shaded area in Panel (b) down, so that the trade probability for any buyer type does not change and so that trade takes place with the lowest seller types.

allocation rule y^f which holds the trade probability of the buyer constant but shifts all trade probability to the lowest seller types, as illustrated in Figure 1. Note that this allocation rule is of the form in Lemma 1. Further, the buyer's transfers are the same under both allocation rules, as the trade probability is the same. To complete the proof, we need to show that the transfers to the sellers are lower in this new allocation rule. To do so, we prove a technical lemma in the appendix (Lemma 2).

Having proved that the optimal allocation rule is as equation (11), we can make use of this structure to obtain

$$y_B(\theta_B)^2 = \left(\int_a^b y^f(\theta_B,\theta_S)dF_S(\theta_S)\right)^2 = \left(\int_a^{\theta_S^*(\theta_B)} 1dF_S(\theta_S)\right)^2 = F_S(\theta_S^*(\theta_B))^2$$
$$= F_S(\theta_S)^2\big|_a^{\theta_S^*(\theta_B)} = 2\int_a^{\theta_S^*(\theta_B)} F_S(\theta_S)dF_S(\theta_S)dt = 2\int_a^b y^f(\theta_B,\theta_S)F_S(\theta_S)dF_S(\theta_S),$$

where we integrate by parts at the beginning of the second line. We analogously obtain $y_S(\theta_S)^2 = 2\int_a^b y^f(\theta_B,\theta_S)(1-F_B(\theta_B))dF_B(\theta_B)$. We can use these equalities to get rid of the terms $y_B(\theta_B)^2$ and $y_S(\theta_S)^2$ in equation (RM') which becomes

$$\int_a^b\int_a^b \left(\tilde{J}_B(\theta_B) - \tilde{J}_S(\theta_S)\right) K(\theta_B,\theta_S)y^f(\theta_B,\theta_S)dF_S(\theta_S)dF_B(\theta_B), \tag{12}$$

where

$$\tilde{J}_i(\theta_i) := \frac{J_i(\theta_i)}{1 - \Lambda_{-i} + 2\Lambda_{-i}F_i(\theta_i)} \tag{13}$$

are modified virtual types and $K(\theta_B, \theta_S) := (1-\Lambda_S+2\Lambda_S F_B(\theta_B))(1-\Lambda_B+2\Lambda_B F_S(\theta_S)) \geq 0$. We now have a concave maximization problem. If the modified virtual types $\tilde{J}_i$ are increasing, the monotonicity constraints on y_i are satisfied. In this case, an allocation rule y^f is optimal if and only if a trade takes place whenever $\tilde{J}_B(\theta_B) - \tilde{J}_S(\theta_S) \geq 0$ (see, e.g., Theorems 1 and 2 in Luenberger, 1969, p. 217 and p. 221). However, in the presence of loss aversion, increasing virtual types (Assumption 2) does not necessarily imply increasing modified virtual types so we are not quite done yet. We can, however, iron these modified virtual types as in Myerson (1981). To do so, define the ironed (modified) virtual types by[33]

$$I_B(\theta_B) := \frac{d}{dt}\left[\text{conv}\left(\int_0^t \frac{J_B(F_B^{-1}(x))}{1 - \Lambda_S + 2\Lambda_S x}\, dx\right)\right]\Bigg|_{t=F_B(\theta_B)}, \tag{14}$$

$$I_S(\theta_S) := \frac{d}{dt}\left[\text{conv}\left(\int_0^t \frac{J_S(F_S^{-1}(x))}{1 - \Lambda_B + 2\Lambda_B x}\, dx\right)\right]\Bigg|_{t=F_S(\theta_S)}. \tag{15}$$

These ironed virtual types are increasing and thus the allocation rule in which trade takes place whenever $I_B(\theta_B) \geq I_S(\theta_S)$ satisfies the monotonicity constraints and, as we show in the proof of the following proposition, maximizes the designer's revenue.

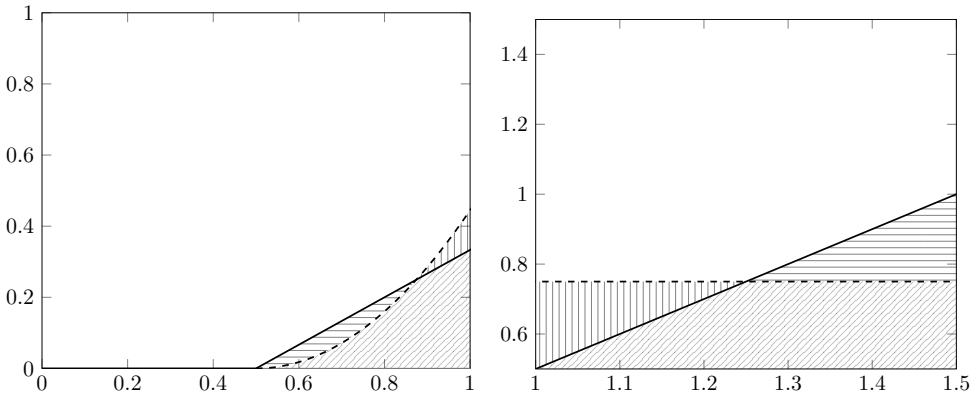
Proposition 6 *The revenue-maximizing allocation rule is given by*

$$y^{RM}(\theta_B, \theta_S) = \begin{cases} 1 & \text{if } I_B(\theta_B) - I_S(\theta_S) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

This result deserves some discussion, as it has several noteworthy features. First, as in Myerson and Satterthwaite (1983), the designer is optimally inducing trade when the buyer's (ironed) virtual type is bigger than the seller's. The difference is that we consider modified virtual types, which correct for agents' loss aversion, although the result from MS is contained as a special case.

Second, depending on the distribution of types, the trade frequency can increase or decrease compared to the case with no loss aversion, as illustrated in Figure 2. This heterogeneity across types comes as a surprise, as loss aversion affects any buyer (seller) type in the same way through a reduction in the expected utility due to the ex-post variation in payoffs and by reducing (increasing) their information rent. So, where does it come from, and what determines for which types the designer increases or decreases

[33]This definition follows Mierendorff (2016). See footnote 36 on page 210 therein for details.



(a) Partial insurance (b) Full insurance

Figure 2: The above figures depict optimal allocation rules with the buyer and seller types on the x and y axes, respectively. Panel (a): Types are drawn from the unit interval with buyer types distributed according to $F_B(\theta_B) = \theta_B$ and seller according to $F_S(\theta_S) = \sqrt{\theta_S}$. Loss-aversion parameters read $(\Lambda_B, \Lambda_S) = (1, 0)$. Panel (b): Buyer types are drawn uniformly from $[1, 1.5]$ and seller types uniformly from $[0.5, 1.5]$. Loss-aversion parameters read $(\Lambda_B, \Lambda_S) = (0, 1/2)$. In both panels the solid line corresponds to the case without loss aversion; the dashed line corresponds to the case with loss aversion. For profiles (θ_B, θ_S), loss aversion leads to a reduction of trade in the horizontally-shaded area and to an increase in trade in the vertically-shaded area; in the diagonally-shaded area trade takes place with and without loss aversion.

the trade probability? The answer lies in loss-averse agents' dislike for ex-post variation in payoffs. To get an intuition for this, consider some pair (θ'_B, θ'_S) for which the designer is indifferent between inducing trade or not, i.e., $I_B(\theta'_B) = I_S(\theta'_S)$.[34] It follows from the shape of the optimal allocation rule (see Lemma 1) that for such a θ'_S the trade, probability reads $y_S(\theta'_S) = \int_{\theta'_B}^{b} dF_B(t) = 1 - F_B(\theta'_B)$. Notice further, that for any type the ex-post variation in the payoff is maximized when the trade probability is $1/2$, as then either trade or no trade is equally likely. Moreover, this ex-post variation is reduced by moving in either direction, i.e., moving towards trade or towards no trade. Now, observe that in the objective function (12) we can rewrite

$$\begin{aligned}&\left(\tilde{J}_B(\theta_B) - \tilde{J}_S(\theta_S)\right) K(\theta_B, \theta_S)\\ &= J_B(\theta_B)\left[1 - \Lambda_B + 2\Lambda_B F_S(\theta_S)\right] - J_S(\theta_S)\left[1 - \Lambda_S + 2\Lambda_S F_B(\theta_B)\right].\end{aligned}$$

To understand how a change in Λ_S affects the optimal trade probability for our type θ'_S take the derivative of the above equation with respect to Λ_S, yielding $J_S(\theta'_S)(1-2F_B(\theta'_B))$. This is positive if and only if $F_B(\theta'_B) \leq 1/2$. Therefore, for any pair (θ'_S, θ'_B) for which $y_S(\theta'_S) = 1 - F_B(\theta'_B) \geq 1/2$, the designer will optimally increase the trade probability and

[34] For the sake of exposition we suppose that the modified and ironed virtual types coincide so that there are no non-monotonicity issues in the modified types as we describe the intuition in these paragraphs.

otherwise decrease it. In summary, the designer reduces ex-post variation for all types, by either increasing or decreasing the trade probability, which increases the utility of loss-averse agents. This reduction in ex-post variation is achieved in a way that is in line with the designer's material interests, as it increases the trade probability for "profitable types" (high buyer and low seller types) and decreases it with "not-so-profitable" types, thus maximizing the designer's revenue. While the optimal allocation rule does in general not provide agents with full insurance in the trade dimension, it may do so as illustrated in Figure 2b, where some seller types trade with probability 1 and some seller types trade with probability 0, while buyers have a constant trade probability of 0.25 (but are not loss averse in that example).

Finally, let us discuss the need for ironing. If we are willing to impose stricter bounds on loss aversion than required by the assumption of no dominance of gain-loss utility (Assumption 1), no ironing is necessary, as then the assumption of increasing virtual types (Assumption 2) is enough to ensure that the modified virtual types are increasing. For some type distributions, we do not need to impose further assumptions on the parameters of loss aversion.[35] To obtain a general solution, however, ironing is necessary.

To close this section, we focus our attention on the case when types are uniformly distributed on $[a, a+1]$ and impose additional bounds on the degree of loss aversion. This eliminates the need for ironing and allows us to derive closed-form solutions of the optimal mechanism, allowing for interesting comparative statics.

Corollary 2 *Consider the case of uniformly distributed types on the interval* $[a, a+1]$ *for* $a \geq 0$ *and suppose that* $\Lambda_B \leq 1/(a+1), \Lambda_S \leq \min\{1, 1/a\}$. *Then, the optimal allocation rule reads*

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

where

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

Moreover, for $(\Lambda_B, \Lambda_S) \neq (0, 0)$, *increasing the parameter* a *reduces the optimal trade frequency, eventually eliminating all trade.*

We can interpret an increase of a as an increase in the stakes. Thus, for higher stakes, less trade takes place for any positive degree of loss aversion with trade vanishing completely, as we hit the bound. This is in sharp contrast to the case without loss

[35] For instance, when seller types are distributed on some interval $[0, b]$ with $f_S'(\theta_S) \leq 0$, we can allow for any $\Lambda_B \leq 1$.

aversion, where the optimal mechanism is independent of the size of the stakes. Intuitively, the potential material gains from trade remain the same even when the stakes are high because only the difference between valuation matters. However, as the stakes increase, the potential losses increase. Since the designer needs to compensate the agents for these losses with appropriate transfers to maintain participation, the losses eventually eat up all the potential material gains.[36] Hence, at some point the best the designer can do is to induce no trade at all. Contrary to conventional wisdom, the behavioral effects of loss aversion are not mitigated when the stakes are large. Rather, loss aversion has the biggest impact precisely when the stakes are large.

4.2 Maximizing the Gains from Trade

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

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

In order to formulate the maximization problem, we impose a budget balance condition in addition to CPEIC and IR. Namely, we do not want the designer to inject money into the economy on average. This is in line with the preceding section, where we looked at

[36]The combination of increasing stakes while keeping constant the potential gains from trade is key here. If types are instead uniformly distributed on $[0, b]$ and we increase b the effect does not materialize. Intuitively, while the stakes increase, so does the size of the cake that can be distributed to the agents.

ex-ante revenue maximization. We say that a mechanism is ex-ante budget balanced if

$$\int_a^b \int_a^b \left(t_S^f(\theta_S, \theta_B) - t_B^f(\theta_S, \theta_B) \right) \, dF_S(\theta_S) \, dF_B(\theta_B) = 0. \tag{AB}$$

We consider two maximization problems given by

$$\max_{(y^f, t_B^f, t_S^f) \in \mathcal{F}} \int_a^b U_S(\theta_S, s_B^t | \theta_S) \, dF_S(\theta_S) + \int_a^b U_B(\theta_B, s_S^t | \theta_B) \, dF_B(\theta_B),$$
$$\text{subject to CPEIC, IR and AB.} \tag{TG}$$

and

$$\max_{(y^f, t_B^f, t_S^f) \in \mathcal{F}} \int_a^b (-\theta_S y_S(\theta_S) + \bar{t}_S(\theta_S)) \, dF_S(\theta_S) + \int_a^b (\theta_B y_B(\theta_B) - \bar{t}_B(\theta_B)) \, dF_B(\theta_B),$$
$$\text{subject to CPEIC, IR and AB.} \tag{MG}$$

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

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

The proof is analogous to the revenue maximization problem and thus omitted. From here we proceed as we did for the derivation of the revenue-maximizing mechanism, albeit carrying the Lagrange multiplier γ from the budget constraint with us. For the problem of maximizing total gains from trade, everything goes through as it did with the revenue-maximization problem. For the case of maximizing material gains from trade only, however, the ironing approach fails, as the modified virtual types of buyer and seller cannot be separated from each other. Imposing bounds on loss aversion, we can nevertheless derive the optimal allocation rule.[37]

Proposition 8 *The optimal allocation rule for the problem of maximizing total gains*

[37] See the proof of the result for the precise definitions of the expressions in the result. They are analogous to the ones in the revenue-maximization problem above.

from trade reads

$$y^{TG}(\theta_B, \theta_S) = \begin{cases} 1 & \text{if } I_B^{TG}(\theta_B, \gamma) - I_S^{TG}(\theta_S, \gamma) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

Assuming the parameters Λ_S and Λ_B are sufficiently small, the optimal allocation rule for the problem of maximizing material gains from trade reads

$$y^{MG}(\theta_B, \theta_S) = \begin{cases} 1 & \text{if } \tilde{J}_B^{MG}(\theta_B, \theta_S, \gamma) - \tilde{J}_S^{MG}(\theta_B, \theta_S, \gamma) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

Naturally, the two allocation rules coincide when $\Lambda_B = \Lambda_S = 0$, in which case we find ourselves in the setting as in MS. In general, however, the optimal allocation rules (and thus transfers) for the two problems will be distinct and it will matter what stance the designer takes regarding gain-loss utility. Yet, in some instances, it does not matter whether the designer treats gain-loss utility as a mistake or not, as the following corollary shows.[38]

Corollary 3 *Consider the case of uniformly distributed types on the unit interval. If $\Lambda_S = \Lambda_B = \Lambda$, then the optimal mechanisms for the problems (TG) and (MG) coincide.*

5 Conclusion

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

We first formally identify the endowment and attachment effects and study their impact on the agents' information rents. Using these insights, we can show that it remains impossible to implement ex-post materially efficient trade, but that buyer loss aversion can mitigate the severity of this impossibility. Turning to the design of optimal mechanisms, we find that the designer should reduce ex-post variation in payoffs. More specifically, when maximizing revenue or gains from trade, agents receive full insurance in the money

[38] Without the restriction to uniformly distributed types, the problem is not sufficiently tractable to make further comparisons between the two problems (TG) and (MG). However, the equivalence proved in Corollary 3 does not extend in general to uniformly distributed types, as one can numerically show that for $\Lambda_B \neq \Lambda_S$ the optimal mechanisms are distinct.

dimension in the form of interim-deterministic transfers and partial insurance in the ownership dimension in the form of increased or decreased trade probabilities relative to the case with no loss aversion. Whether the trade probability increases depends on how attractive a trade between some types is from a material perspective.

One may wonder whether other models than the one by Kőszegi and Rabin can also explain the attachment and endowment effect and thus constitute alternatives to the present analysis. One obvious alternative is a model of loss aversion with a fixed reference point, such as classical prospect theory by Kahneman and Tversky.[39] Indeed, with an appropriately chosen, fixed reference point, we can obtain the attachment and endowment effect. However, the question of what constitutes an appropriate reference point remains open.

Maintaining the assumption of an endogenously determined reference point, one could instead consider a preferred personal equilibrium (PPE) as introduced in Kőszegi and Rabin (2006). The key distinction from the CPE is that, when considering deviations, the agents' reference points do not change. This seemingly small change substantially impacts the above analysis, despite agents' expected utility in equilibrium remaining unchanged. While the revelation principle still holds under PPE, the classical characterization of incentive-compatible mechanisms does not obtain. Indeed, fixing the reference point as deviations are considered is reminiscent of the type-dependent reference point in Carbajal and Ely (2016), in which the standard characterization of incentive-compatible mechanisms fails, too. Given the important role of the characterization of incentive-compatible mechanisms throughout, we are left to speculate on how the results would change under PPE. First, we expect interim-deterministic transfers to remain optimal. Just as with CPE, any ex-post variation in transfers reduces agents' expected payoffs and could be recouped by the designer using interim-deterministic transfers. Second, in terms of the optimal allocation rules, we expect qualitatively similar results with PPE as with CPE. Agents' dislike of ex-post variation of payoffs under PPE should also lead to a reduction in optimal trade frequencies. As CPE yields more risk aversion than PPE (Proposition 8, Kőszegi and Rabin, 2006), the changes in the trade frequency could be less pronounced with PPE than with CPE.

Another take would be to allow for the endogenous reference point to be given by the certainty equivalent of a lottery rather than the full lottery as in Bell (1985) and Loomes and Sugden (1986). Kőszegi and Rabin (2007) note that these models of disappointment aversion are very similar to the KR's framework with CPE. However, Masatlioglu and

[39]Salant and Siegel (2016) study the efficient allocation of a divisible asset for different types of reallocation costs. For concave reallocation cost, the initial allocation can be interpreted as the reference point and, deviations from the reference point lead to losses (but no gains) that are symmetric across agents. They show that ex-post efficiency may not be attained, suggesting a robustness of the impossibility in the presence of a fixed reference point, which was also documented in an older version of the present paper.

Raymond (2016) find that the intersection of preferences induced by expectations-based loss aversion with CPE and any of these disappointment-aversion models is only standard expected utility, and thus while seemingly similar, the models are quite different. Nevertheless, Benkert (2022) shows that the optimal mechanisms for the two types of models are equivalent across a range of mechanism design settings. In particular, the optimal mechanisms derived in the present paper are also optimal if we instead work with a model of disappointment aversion as in Bell (1985) and Loomes and Sugden (1986).[40] This finding is of practical relevance, as the designer of some economic institution may have evidence that individuals are loss averse, but be unsure about the precise formation process of the reference point, be it fixed, as a full lottery over outcomes or, as the certainty equivalent of the lottery. There appears to be some robustness, which suggests that lacking this information may not be too much of a problem, as long as loss-averse individuals are provided with insurance as derived above.

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

[40] The impossibility result would not be affected either.

A Proofs

Proof of Proposition 1

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 $\Gamma = (M_S, M_B, g)$ that implements f in CPE. If $s^* = (s_S^*, s_B^*)$ 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 $s^* = (s_S^*, s_B^*)$ 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

$$g(s_S^*(\theta_S), s_B^*(\theta_B)) = f(\theta_S, \theta_B),$$

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.

Proof of Proposition 2

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

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

and analogously

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

Thus,

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

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

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

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. ■

Proof of Proposition 3

As noted in the main text we can write $\tilde{v}_B(\theta_B) = y_B(\theta_B)(1 - \Lambda_B(1 - y_B(\theta_B)))$ and $\tilde{v}_S(\theta_S) = y_S(\theta_S)(1 + \Lambda_S(1 - y_S(\theta_S)))$, allowing us to rewrite the information rents in equations (6) and (7)

$$IR_B(\theta_B) = \int_{a_B}^{\theta_B} \tilde{v}_B(t)\, dt = \int_{a_B}^{\theta_B} y_B(t)(1 - \Lambda_B(1 - y_B(t)))\, dt$$

and

$$IR_S(\theta_S) = \int_{\theta_S}^{b_S} \tilde{v}_S(t)\, dt = \int_{\theta_S}^{b_S} y_S(t)(1 + \Lambda_S(1 - y_S(t)))\, dt.$$

Taking derivatives we obtain

$$\frac{\partial IR_B(\theta_B)}{\partial \Lambda_B} = -\int_{a_B}^{\theta_B} y_B(t)(1-y_B(t))\,dt \leq 0,$$

and

$$\frac{\partial IR_S(\theta_S)}{\partial \Lambda_S} = \int_{\theta_S}^{b_S} y_S(t)(1-y_S(t))\,dt \geq 0,$$

completing the proof.

Proof of Proposition 4

We know from equation (8) that $\tilde{v}_B(\theta_B) = y_B(\theta_B)(1+\Lambda_B(y_B(\theta_B)-1))$ and $\tilde{v}_S(\theta_S) = y_S(\theta_S)(1-\Lambda_S(y_S(\theta_S)-1))$, where

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

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

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

Note that the monotonicity constraints are satisfied due to Assumption 1, i.e., $\Lambda_B, \Lambda_S \leq 1$. Further, from the discussion in the main text, we know that we can set the loss aversion in the money dimension to zero, as it only makes the problem harder. This allows us to express the sum of the agents' ex-ante expected utilities as

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

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

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

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

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

Myerson and Satterthwaite (1983) show in their proof of Theorem 1 (p. 269) that

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

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

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

Careful inspection of the limits of the integrals shows that

$$
U_B(a_B) + U_S(b_S) = - \int_{\max\{a_B, a_S\}}^{\min\{b_S, b_B\}} (1 - F_B(x)) F_S(x) \, dx
$$

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

violating individual rationality.

Proof of Proposition 5

As noted in the text, the objective function reads

$$
\begin{aligned}
&\int_a^b \left(\eta_B^2 w_B(\theta_B) + \theta_B \tilde{v}_B(\theta_B) - \int_a^{\theta_B} \tilde{v}_B(t)\,dt\right)\, dF_B(\theta_B) \\
&+ \int_a^b \left(\eta_S^2 w_S(\theta_S) - \theta_S \tilde{v}_S(\theta_S) - \int_{\theta_S}^b \tilde{v}_S(t)\,dt\right)\, dF_S(\theta_S),
\end{aligned}
$$

where we observe that w_B and w_S enter positively. Next, note that

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

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 perform a change of variables for the resulting integral. This shows that the two integrals are the same and allows us to sum them up. Thus, since $\lambda_S^2 > 1$ we find $w_S(\theta_S) \leq 0$. The argument for w_B is analogous.

Given that w_B and w_S enter the designer's objective function positively, the designer optimally sets $w_i(\theta_i) = 0$. Further, a transfer achieves $w_i(\theta_i) = 0$ if and only if the transfer is independent of almost all buyer types. Thus, interim deterministic transfers are the

only transfers that achieve $w_i(\theta_i) = 0$.

Proof of Lemma 1

We begin by proving the following technical lemma.

Lemma 2 *Let $f, g : [a, b] \to \mathbb{R}$ be two integrable functions with the following properties:*

(1) We have

$$\int_a^b f(x)dx \geq \int_a^b g(x)dx$$

(2) There exists a $x_0 \in [a, b]$ such that

a) $f(x) \geq g(x)$ for a.e. $x \leq x_0$

b) $f(x) \leq g(x)$ for a.e. $x \geq x_0$

Further, define $\varphi : [a, b] \to \mathbb{R}$. Then, if φ is monotonically decreasing we have

$$\int_a^b \varphi(x)f(x)dx \geq \int_a^b \varphi(x)g(x)dx,$$

and if φ is monotonically increasing we have

$$\int_a^b \varphi(x)f(x)dx \leq \int_a^b \varphi(x)g(x)dx.$$

Proof. We prove the statement for the case when φ is monotonically decreasing. For the case of an increasing φ simply reverse the appropriate inequalities. We begin by rewriting property (1) to

$$\int_a^b f(x)dx \geq \int_a^b g(x)dx \Leftrightarrow \int_a^{x_0} (f(x) - g(x))dx \geq \int_{x_0}^b (g(x) - f(x))dx$$

and note that both integrands are weakly positive due to property (2). Then, once more by (2), we have for a.e. $x \leq x_0$

$$\varphi(x)(f(x) - g(x)) \geq \varphi(x_0)(f(x) - g(x))$$

for a.e. $x \leq x_0$, which we can integrate to obtain

$$\int_a^{x_0} \varphi(x)(f(x) - g(x))dx \geq \int_a^{x_0} \varphi(x_0)(f(x) - g(x))dx. \tag{19}$$

Proceeding analogously, we obtain the inequality

$$\int_{x_0}^{b} \varphi(x_0)(g(x) - f(x))dx \geq \int_{x_0}^{b} \varphi(x)(g(x) - f(x))dx. \tag{20}$$

Further, we also have by

$$\int_{a}^{x_0} \varphi(x_0)(f(x) - g(x))dx = \int_{x_0}^{b} \varphi(x_0)(g(x) - f(x))dx \tag{21}$$

by property (1). Combining the inequalities in equations (19) to (21) we obtain

$$\int_{a}^{x_0} \varphi(x)(f(x) - g(x))dx \geq \int_{x_0}^{b} \varphi(x)(g(x) - f(x))dx,$$

which we can rearrange to

$$\int_{a}^{b} \varphi(x)f(x)dx \geq \int_{a}^{b} \varphi(x)g(x)dx,$$

completing the proof. ■

With this in hand, we can prove Lemma 1. We begin by performing a change of variables, which will simplify the analysis. Let $v_i = F_i(\theta_i)$ and define $\varphi_i(v_i) = F_i^{-1}(v_i)$. Further, define

$$q(v_B, v_S) = y(\varphi_B(v_B), \varphi_S(v_S)), \quad q_i(v_i) = \int_0^1 q(v_i, v_{-i})dv_{-i}$$

as well as

$$M_B(v_B) = J_B(\varphi_B(v_B)) = \varphi_B(v_B) - (1 - v_B)\varphi_B'(v_B),$$
$$M_S(v_S) = J_S(\varphi_S(v_S)) = \varphi_S(v_S) + v_S\varphi_S'(v_S).$$

The problem in (RM') then becomes

$$\int_0^1 M_B(v_B)q_B(v_B)(1 - \Lambda_B(1 - q_B(v_B)))dv_B - \int_0^1 M_S(v_S)q_S(v_S)(1 + \Lambda_S(1 - q_S(v_S))dv_S \tag{22}$$

subject to the monotonicity constraints. Let $\hat{q} : [0,1] \times [0,1] \to \{0,1\}$ be any candidate for optimality. Associate to $\hat{q}$ the function

$$q(v_B, v_S) = \begin{cases} 1 & 0 \leq v_S \leq \hat{q}_B(v_B) \\ 0 & o.w., \end{cases} \tag{23}$$

where $\hat{q}_B(v_B) = \int_0^1 \hat{q}(v_B, v_S)dv_S$. First, note that $q_B = \hat{q}_B$ by construction. Thus, the first integral in equation (22) is not affected by a change from $\hat{q}$ to q. However, we will show that the second integral, which enters negatively, will become smaller. To do so, we will show

$$\int_0^1 M_S(v_S)q_S(v_S)(1+\Lambda_S)dv_S \leq \int_0^1 M_S(v_S)\hat{q}_S(v_S)(1+\Lambda_S)dv_S \tag{24}$$

and

$$\int_0^1 \Lambda_S M_S(v_S)(q_S^2(v_S) - \hat{q}_S^2(v_S))dv_S \geq 0. \tag{25}$$

To prove (24), fix $v_B \in [0,1]$ and apply Lemma 2 by setting $f(v_S) = q(v_B, v_S)$, $g(v_S) = \hat{q}(v_B, v_S)$ and $x_0 = \hat{q}_B(v_B)$. Note that the properties (1) and (2) in the lemma are satisfied by the construction of q from $\hat{q}$. Further, $M_S(v_S)(1+\Lambda_S)$ is a monotonically increasing function so that Lemma 2 yields

$$\int_0^1 M_S(v_S)(1+\Lambda_S)q(v_B, v_S)dv_S \geq \int_0^1 M_S(v_S)(1+\Lambda_S)\hat{q}(v_B, v_S)dv_S.$$

Let us now integrate this with respect to v_B and apply Fubini's theorem to reverse the order of integration to obtain

$$\begin{aligned}
&\int_0^1 \int_0^1 M_S(v_S)(1+\Lambda_S)q(v_B, v_S)dv_S dv_B \geq \int_0^1 \int_0^1 M_S(v_S)(1+\Lambda_S)\hat{q}(v_B, v_S)dv_S dv_B \\
&\Leftrightarrow \int_0^1 \int_0^1 M_S(v_S)(1+\Lambda_S)q(v_B, v_S)dv_B dv_S \geq \int_0^1 \int_0^1 M_S(v_S)(1+\Lambda_S)\hat{q}(v_B, v_S)dv_B dv_S \\
&\Leftrightarrow \int_0^1 M_S(v_S)q_S(v_S)(1+\Lambda_S)dv_S \leq \int_0^1 M_S(v_S)\hat{q}_S(v_S)(1+\Lambda_S)dv_S
\end{aligned}$$

as claimed in equation (24).

To prove (25), we begin by noting that we can rewrite this inequality to

$$\begin{aligned}
&\int_0^1 M_S(v_S)[q_S^2(v_S) - \hat{q}_S^2(v_S)]dv_S \geq 0 \\
&\Leftrightarrow \int_0^1 [q_S(v_S) + \hat{q}_S(v_S)][M_S(v_S)q_S(v_S) - M_S(v_S)\hat{q}_S(v_S)]dv_S \geq 0.
\end{aligned}$$

We will once more apply Lemma 2. Fix $v_B \in [0,1]$ and set $f(v_S) = M_S(v_S)q(v_B, v_S)$,

$g(v_S) = M_S(v_S)\hat{q}(v_B, v_S)$. Note that by (24)

$$\int_0^1 f(v_S)dv_S \geq \int_0^1 g(v_S)dv_S$$

so that property (1) is satisfied. Further, if $v_S \leq \hat{q}_B(v_B)$, then $M_S(v_S)q(v_S, v_B) = M_S(v_S) \geq M_S(v_S)\hat{q}(v_B, v_S)$. Similarly, if $v_S \geq \hat{q}_B(v_B)$, then $M_S(v_S)q(v_S, v_B) = 0 \leq M_S(v_S)\hat{q}(v_B, v_S)$. Together, this shows that property (2) is satisfied. Further, define $\phi(v_S) = q_S(v_S) + \hat{q}_S(v_S)$ and note that it is a decreasing function, as the candidate solution $\hat{q}_S$ is decreasing by assumption and the associated q_S by the construction in (23).[41] Therefore, it follows from Lemma 2 and by once more integrating with respect to v_B and applying Fubini's theorem that

$$\begin{aligned}
&\int_0^1 [q_S(v_S) + \hat{q}_S(v_S)][M_S(v_S)q_S(v_S) - M_S(v_S)\hat{q}_S(v_S)]dv_S \geq 0 \\
\Leftrightarrow &\int_0^1 [q_S(v_S) + \hat{q}_S(v_S)]M_S(v_S)(q_S(v_S) - \hat{q}_S(v_S))dv_S \geq 0 \\
\Leftrightarrow &\int_0^1 \Lambda_S M_S(v_S)(q_S^2(v_S) - \hat{q}_S^2(v_S))dv_S \geq 0,
\end{aligned}$$

as claimed in equation (25).

Putting equations (24) and (25) together, we obtain that

$$\begin{aligned}
&\int_0^1 M_S(v_S)\hat{q}_S(v_S)(1 - \Lambda_S(\hat{q}_S(v_S) - 1))\, dv_S - \int_0^1 M_S(v_S)q_S(v_S)(1 - \Lambda_S(q_S(v_S) - 1))\, dv_S \\
= &\int_0^1 M_S(v_S)(1 + \Lambda_S)\hat{q}_S(v_S) - M_S(v_S)\Lambda_S\hat{q}_S^2(v_S)\, dv_S \\
&- \int_0^1 M_S(v_S)(1 + \Lambda_S)q_S(v_S) - M_S(v_S)\Lambda_S q_S^2(v_S)\, dv_S \\
= &\int_0^1 M_S(v_S)(1 + \Lambda_S)[\hat{q}_S(v_S) - q_S(v_S)] + M_S(v_S)\Lambda_S[q_S^2(v_S) - \hat{q}_S^2(v_S)]\, dv_S \geq 0,
\end{aligned}$$

showing that the second integral in equation (22) indeed becomes smaller when moving from $\hat{q}$ to q. Hence, for any $\hat{q}$ not of the form (11) we can construct a function in this class which does better, completing the proof.

Proof of Proposition 6

As noted in the main text, we can leverage Lemma 1 to rewrite the objective function to

$$\int_a^b \int_a^b \left(\tilde{J}_B(\theta_B) - \tilde{J}_S(\theta_S)\right) K(\theta_B, \theta_S)y^f(\theta_B, \theta_S)dF_S(\theta_S)dF_B(\theta_B), \tag{26}$$

[41] To see this, note that $q_S(v_S) = 1 - q_B^{-1}(v_S)$. Thus, since q_B is increasing, q_S is decreasing.

where

$$\tilde{J}_i(\theta_i) := \frac{J_i(\theta_i)}{1 - \Lambda_{-i} + 2\Lambda_{-i}F_i(\theta_i)} \tag{27}$$

are modified virtual types and $K(\theta_B, \theta_S) := (1-\Lambda_S+2\Lambda_S F_B(\theta_B))(1-\Lambda_B+2\Lambda_B F_S(\theta_S)) \geq 0$. Having defined the virtual types in equations (14) and (15), we will show (closely following Myerson, 1981) that the allocation rule in (16)

$$y^f(\theta_B, \theta_S) = \begin{cases} 1 & \text{if } I_B(\theta_B) - I_S(\theta_S) \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

maximizes the designer's revenue. We begin by defining the functions $g_i(\theta_i) = f_i(\theta_i)(1 - \Lambda_j + 2\Lambda_j F_i(\theta_i))$, allowing us to rewrite the objective to

$$\int_a^b \int_a^b \left(\tilde{J}_B(\theta_B) - \tilde{J}_S(\theta_S)\right) y^f(\theta_B, \theta_S) g_S(\theta_S) g_B(\theta_B) d\theta_S d\theta_B$$

Note that the functions g_i are PDFs, so that in particular $G_i(a) = 0$ and $G_i(b) = 1$. With this in hand and abusing notation a bit so that $y_B(\theta_B) = \int_a^b y(\theta_B, \theta_S) dG_S(\theta_S)$ we have

$$\begin{aligned}
&\int_a^b \int_a^b [\tilde{J}_B(\theta_B) - I_B(\theta_B)] y(\theta_B, \theta_S) dG_S(\theta_S) dG_B(\theta_B) \\
&= \int_a^b [\tilde{J}_B(\theta_B) - I_B(\theta_B)] y_B(\theta_B) dG_B(\theta_B) \\
&= \left[y_B(\theta_B) \int_a^{\theta_B} [\tilde{J}_B(\theta_B) - I_B(\theta_B)] g_B(t) dt \right]_a^b - \int_a^b \left(\int_a^{\theta_B} [\tilde{J}_B(\theta_B) - I_B(\theta_B)] g_B(t) dt \right) y_B'(\theta_B) d\theta_B \\
&= \left[y_B(\theta_B) G_B(\theta_B) \int_a^{\theta_B} [\tilde{J}_B(\theta_B) - I_B(\theta_B)] dt \right]_a^b - \int_a^b \left(G_B(\theta_B) \int_a^{\theta_B} [\tilde{J}_B(\theta_B) - I_B(\theta_B)] dt \right) y_B'(\theta_B) d\theta_B \\
&= - \int_a^b \left(G_B(\theta_B) \int_a^{\theta_B} [\tilde{J}_B(\theta_B) - I_B(\theta_B)] dt \right) y_B'(\theta_B) d\theta_B
\end{aligned} \tag{28}$$

where the last equality follows because the generalized virtual types are the convex hull of the modified virtual types, which are continuous so that the endpoint terms are zero. The analogous expression can be derived for the seller. Now, we can obtain

$$\begin{aligned}
&\int_a^b \int_a^b \left(\tilde{J}_B(\theta_B) - \tilde{J}_S(\theta_S)\right) K(\theta_B, \theta_S) y^f(\theta_B, \theta_S) dF_S(\theta_S) dF_B(\theta_B) \\
&= \int_a^b \int_a^b \left(\tilde{J}_B(\theta_B) - \tilde{J}_S(\theta_S)\right) y^f(\theta_B, \theta_S) dG_S(\theta_S) dG_B(\theta_B) \\
&= \int_a^b \int_a^b \left(I_B(\theta_B) - I_S(\theta_S)\right) y^f(\theta_B, \theta_S) dG_S(\theta_S) dG_B(\theta_B)
\end{aligned} \tag{29}$$

$$+\int_a^b\int_a^b\left(\tilde{J}_B(\theta_B)-I_B(\theta_b)\right)y^f(\theta_B,\theta_S)dG_S(\theta_S)dG_B(\theta_B) \tag{30}$$

$$+\int_a^b\int_a^b\left(I_S(\theta_S)-\tilde{J}_S(\theta_S)\right)y^f(\theta_B,\theta_S)dG_S(\theta_S)dG_B(\theta_B) \tag{31}$$

Now consider the allocation rule in question, that is,

$$y^f(\theta_B,\theta_S)=\begin{cases}1 & \text{if } I_B(\theta_B)\geq I_S(\theta_S)\\ 0 & \text{otherwise.}\end{cases}$$

First, the expression in equation (29) is maximized with this rule. Second, note that the expression in (30) is equal to equation (28). Consider any allocation rule which is CPEIC so that y_B and y_S are increasing and decreasing. Then, the expression in (28) is weakly negative since $I_B\leq\tilde{J}_B$. Consider now the candidate allocation rule and observe that whenever $\tilde{J}_B\neq I_B$ we must have $\tilde{J}_B(\theta_B)>I_B(\theta_B)$ since I_B is the convex hull of $\tilde{J}_B$ and that I_B must thus be flat on those intervals, implying that y_B must be flat on those intervals, too, as the trade probability does not increase when I_B is flat. But then, this implies

$$\begin{aligned}&\int_a^b\int_a^b\left(\tilde{J}_B(\theta_B)-I_B(\theta_b)\right)y^f(\theta_B,\theta_S)dG_S(\theta_S)dG_B(\theta_B)\\&=-\int_a^b\left(G_B(\theta_B)\int_a^{\theta_B}[\tilde{J}_B(\theta_B)-I_B(\theta_B)]dt\right)y_B'(\theta_B)d\theta_B\\&=0.\end{aligned}$$

An analogous argument implies that the expression in equation (31) is equal to zero. Putting this together, it follows that our candidate allocation rule is indeed the one which maximizes the objective function and that it satisfies the monotonicity constraints. Thus, applying Theorems 1 and 2 in Luenberger, 1969, p. 217 and p. 221, equation (16) is the revenue-maximizing mechanism.

Proof of Corollary 2

We impose that types are uniformly distributed types on the interval $[a,a+1]$ for $a\geq 0$. Further, we for now assume that no ironing is necessary due to the bounds on loss aversion and will verify that the modified virtual types are indeed monotonic. Then, the optimal allocation rule can be written as

$$\frac{2\theta_B-a-1}{1-\Lambda_S+2\Lambda_S(\theta_B-a)}\geq\frac{2\theta_S-a}{1-\Lambda_B+2\Lambda_B(\theta_S-a)}$$

$$\theta_S \leq \delta(\theta_B) := \frac{(2\theta_B - 1 - a)(1 - \Lambda_B(2a+1) + a\Lambda_S) + a - \Lambda_S a^2}{2(1 - \Lambda_B(2\theta_B - a - 1) + \Lambda_S(2\theta_B - 1 - 2a))}.$$

One can now verify that this allocation rule is appropriately monotonic for $\Lambda_B \leq 1/(a+1), \Lambda_S \leq \min\{1, 1/a\}$. Taking derivatives one can show that the allocation rule induces less trade for given Λ_B, Λ_S as a increases, eventually eliminating trade.

Proof of Proposition 8

The derivations of the mechanisms maximizing the total and the material gains from trade proceed analogously. We here present the derivations for the case of maximizing the total gains from trade. Making use of Proposition 7 and the budget constraint (AB), we eliminate the transfers from the problem and can rewrite the objective function to

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

Further, the budget constraint AB and the CPEIC can be jointly written as

$$\begin{aligned}
&\int_a^b J_B(\theta_B) y_B(\theta_B) \left(1 - \Lambda_B \left(1 - y_B(\theta_B)\right)\right)\, dF_B(\theta_B) \\
=&\int_a^b J_S(\theta_S) y_S(\theta_S) \left(1 + \Lambda_S \left(1 - y_S(\theta_S)\right)\right)\, dF_S(\theta_S),
\end{aligned}$$

as well as the monotonicity constraints. We can set up the Lagrangian as

$$\begin{aligned}
\mathcal{L}(y^f, \gamma) = &\int_a^b (\theta_B + \gamma J_B(\theta_B)) y_B(\theta_B) \left(1 - \Lambda_B \left(1 - y_B(\theta_B)\right)\right)\, dF_B(\theta_B) \\
&- \int_0^1 (\theta_S + \gamma J_S(\theta_S)) y_S(\theta_S) \left(1 + \Lambda_S \left(1 - y_S(\theta_S)\right)\right)\, dF_S(\theta_S).
\end{aligned}$$

Note that we must have $\gamma \geq 0$, because relaxing the budget constraint (i.e., allowing the designer to run a deficit) can only increase the objective. Hence, by Assumption 2, $\theta_B + \gamma J_B(\theta_B)$ and $\theta_S + \gamma J_S(\theta_S)$ are strictly increasing in θ_B and θ_S, respectively. Therefore, the arguments from the proof of the revenue-maximizing mechanism carry through and we obtain

$$\mathcal{L}(y^f,\gamma)=\int_a^b\int_a^b\left(\tilde{J}_B^{TG}(\theta_B,\gamma)-\tilde{J}_S^{TG}(\theta_S,\gamma)\right)K(\theta_B,\theta_S)y^f(\theta_B,\theta_S)dF_S(\theta_S)dF_B(\theta_B),,$$

where

$$\tilde{J}_i^{TG}(\theta_i,\gamma):=\frac{\theta_i+\gamma J_i(\theta_i)}{1-\Lambda_{-i}+2\Lambda_{-i}F_i(\theta_i)}.$$

From here we can apply the analogous ironing technique and follow the same steps as in the proof of the revenue-maximizing mechanism to obtain the result.

The Lagrangian for the case of material gains from trade is obtained analogously and reads

$$\begin{aligned}\mathcal{L}(y^f,\gamma)=&\int_a^b\int_a^b\underbrace{\left(\theta_B+\gamma J_B(\theta_B)\left[1-\Lambda_B+2\Lambda_B F_S(\theta_S)\right]\right)}_{\tilde{J}_B^{MG}(\theta_B,\theta_S,\gamma)}y^f(\theta_B,\theta_S)dF_S(\theta_S)dF_B(\theta_B)\\&-\int_a^b\int_a^b\underbrace{\left(\theta_S+\gamma J_S(\theta_S)\left[1-\Lambda_S+2\Lambda_S(F_B(\theta_B)\right]\right)}_{\tilde{J}_S^{MG}(\theta_B,\theta_S,\gamma)}y^f(\theta_B,\theta_S)dF_S(\theta_S)dF_B(\theta_B),.\end{aligned}$$

Inspecting this expression, one can see that the ironing technique does not work in this case as the virtual types cannot be separated. Thus, to derive the optimal mechanism we need to impose bounds on Λ_i which ensures the monotonicity. These points can be obtained by taking the derivative with respect to θ_i from, $\tilde{J}_B^{MG}(\theta_B,\theta_S,\gamma)-\tilde{J}_S^{MG}(\theta_B,\theta_S,\gamma)$. Then for θ_B this derivative needs to be positive and for θ_S it needs to be negative, allowing to solve for Λ_i to obtain the bound.

Proof of Corollary 3

To obtain the expression in Corollary 3 plug in the assumptions on the distributions of types and the parameters of loss aversion to rewrite the trade condition in equation (17) (note that we do not need to iron types, as the modified virtual types are monotonic given the assumptions on the type distributions) to

$$\theta_S\leq\frac{(\theta_B+\gamma(2\theta_B-1))(1-\Lambda)}{1+2g-\Lambda}.$$

We can plug this into the budget constraint and solve for the Lagrange multiplier, yielding $\gamma = (\Lambda + \sqrt{1 + \Lambda + \Lambda^2})/2$ which yields the optimal allocation rule

$$y^{TG}(\theta_B, \theta_S) = \begin{cases} 1 & \theta_S \leq \frac{(1-\Lambda)\left(2\theta_B\left(\sqrt{\Lambda^2+\Lambda+1}+\Lambda+1\right)-\sqrt{\Lambda^2+\Lambda+1}-\Lambda\right)}{2\left(\sqrt{\Lambda^2+\Lambda+1}+1\right)} \\ 0 & o.w. \end{cases}$$

Proceeding analogously for the case of material gains from trade, we obtain the same allocation rule.

References

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

Adar, D. (2021): "The underlying motivations of NFT trading," https://uxdesign.cc/the-underlying-motivations-of-nft-trading-504e4036dda7, last accessed 21.07.2023.

Apesteguia, J. and M. Ballester (2015): "A Measure of Rationality and Welfare," *Journal of Political Economy*, 123, 1278–1310.

Backus, M., T. Blake, B. Larsen, and S. Tadelis (2020): "Sequential Bargaining in the Field: Evidence from Millions of Online Bargaining Interactions," *The Quarterly Journal of Economics*.

Banerji, A. and N. Gupta (2014): "Detection, Identification, and Estimation of Loss Aversion: Evidence from an Auction Experiment," *American Economic Journal: Microeconomics*, 6, 91–133.

Bartling, B., L. Brandes, and D. Schunk (2015): "Expectations as Reference Points: Field Evidence from Professional Soccer," *Management Science*, 61, 2646–2661.

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

Benkert, J.-M. (2022): "On the equivalence of optimal mechanisms with loss and disappointment aversion," *Economics Letters*, 214.

Benkert, J.-M. and N. Netzer (2018): "Informational Requirements of Nudging," *Journal of Political Economy*, 126, 2323–2355.

Bernheim, B. and A. Rangel (2009): "Beyond Revealed Preference: Choice-Theoretic Foundations For Behavioral Welfare Economics," *Quarterly Journal of Economics*, 124, 51–104.

Bierbrauer, F. and N. Netzer (2016): "Mechanism Design and Intentions," *Journal of Economic Theory*, 163, 557–603.

Carbajal, J. C. and J. C. Ely (2016): "A Model of Price Discrimination under Loss Aversion and State-Contingent Reference Points," *Theoretical Economics*, 11, 455–485.

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

Coase, R. (1960): "The Problem of Social Cost," *Journal of Law and Economics*, 3, 1–44.

Cramton, P., R. Gibbons, and P. Klemperer (1987): "Dissolving a partnership Efficienty," *Econometrica*, 55, 615–632.

Crawford, V. P. (2021): "Efficient Mechanisms for Level-k Bilateral Trading," *Games and Economic Behavior*, 127, 80–101.

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

Dato, S., A. Grunewald, D. Müller, and P. Strack (2017): "Expectation-based loss aversion and strategic interaction," *Games and Economic Behavior*, 104, 681–705.

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

Dreyfuss, B., O. Heffetz, and M. Rabin (2022): "Expectations-Based Loss Aversion May Help Explain Seemingly Dominated Choices in Strategy-Proof Mechanisms," *American Economic Journal: Microeconomics*.

Driesen, B., A. Perea, and H. Peters (2012): "Alternating offers Bargaining with loss aversion," *Mathematical Social Sciences*, 64, 103–118.

Duraj, J. (2015): "Mechanism Design with News Utility," Personal Communcation.

——— (2018): "Mechanism Design with News Utility," Mimeo.

Eisenhuth, R. (2019): "Reference-Dependent Mechanism Design," *Economic Theory Bulletin*, 7, 77–103.

Eisenhuth, R. and M. Grunewald (2018): "Auctions with Loss Averse Bidders," *International Journal of Economic Theory*, 16, 129–152.

Ellis, A. and D. J. Freeman (2021): "Revealing Choice Bracketing," Tech. rep.

Ericson, K. M. M. and A. Fuster (2011): "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments," *Quarterly Journal of Economics*, 126, 1879–1907.

——— (2014): "The Endowment Effect," *Annual Review of Economics*, 6, 555–579.

Fehr, E. and L. Goette (2007): "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment," *American Economic Review*, 97, 298–317.

Fieseler, K., T. Kittsteiner, and B. Moldovanu (2003): "Partnerships, lemons, and efficient trade," *Journal of Economic Theory*, 113, 223–234.

Garratt, R. and M. Pycia (2023): "Efficient Bilateral Trade," Mimeo, UCSB and UZH.

Gershkov, A., B. Moldovanu, P. Strack, and M. Zhang (2021): "Optimal Auctions: Non-expected Utility and Constant Risk Aversion," *The Review of Economic Studies*.

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

Gneezy, U., L. Goette, C. Sprenger, and F. Zimmermann (2017): "The Limits of Expectations-Based Reference Dependence," *Journal of the European Economic Association*, 15, 861–876.

Heffetz, O. (2021): "Are reference points merely lagged beliefs over probabilities?" *Journal of Economic Behavior and Organization*, 181, 252–269.

Heffetz, O. and J. A. List (2014): "Is the Endowment Effect an Expectations Effect?" *Journal of the European Economic Association*, 12, 1396–1422.

Heidhues, P. and B. Kőszegi (2014): "Regular Prices and Sales," *Theoretical Economics*, 9, 217–251.

Herweg, F., D. Müller, and P. Weinschenk (2010): "Binary Payment Schemes: Moral Hazard and Loss Aversion," *American Economic Review*, 100, 2451–2477.

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–291.

Karle, H., G. Kirchsteiger, and M. Peitz (2015): "Loss Aversion and Consumption Choice: Theory and Experimental Evidence," *American Economic Journal: Microeconomics*, 7, 101–120.

Karle, H. and M. Möller (2020): "Selling in Advance to Loss Averse Consumers," *International Economic Review*, 61, 441–468.

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

Karle, H. and H. Schumacher (2017): "Advertising and Attachment: Exploiting Loss Aversion through Pre-Purchase Information," *RAND Journal of Economics*, 48, 875–1135.

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

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

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

——— (2009): "Reference-Dependent Consumption Plans," *American Economic Review*, 99, 909–936.

Kneeland, T. (2022): "Mechanism Design with Level-k Types: Theory and an Application to Bilateral Trade," *Journal of Economic Theory*, 201.

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

Larsen, B. J. (2021): "The Efficiency of Real-World Bargaining: Evidence from Wholesale Used-Auto Auctions," *Review of Economic Studies*, 88, 851–882.

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

Manzini, P. and M. Mariotti (2014): "Welfare economics and bounded rationality: the case for model-based approaches," *Journal of Economic Methodology*, 21, 343–360.

Masatlioglu, Y. and C. Raymond (2016): "A Behavioral Analysis of Stochastic Reference Dependence," *The American Economic Review*, 106, 2760–2782.

Maskin, E. and J. Riley (1984): "Optimal Auctions with Risk Averse Buyers," *Econometrica*, 52, 1473 – 1518.

Meisner, V. and J. von Wangenheim (2021): "School choice and loss aversion," Tech. rep.

Mierendorff, K. (2016): "Optimal dynamic mechanism design with deadlines," *Journal of Economic Theory*, 161, 190–222.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58.

——— (1989): *Mechanism Design*, Springer.

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

Pope, D. G. and M. E. Schweitzer (2011): "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review*, 101, 129–157.

Post, T., M. J. van den Assem, G. Baltussen, and R. H. Thaler (2008): "Dear or No Deal? Decision Making under Risk in a Large-Payoff Game Show," *American Economic Review*, 98, 38–71.

Rosato, A. (2016): "Selling Substitute Goods to Loss-Averse Consumers: Limited Availability, Bargains and Rip-offs," *Rand Journal of Economics*.

——— (2017): "Sequential Negotiations with Loss-Averse Buyers," *European Economic Review*, 91, 290–304.

——— (2023): "Loss aversion in sequential auctions," *Theoretical Economics*, 18, 561–596.

Rubinstein, A. and Y. Salant (2012): "Eliciting Welfare Preferences from Behavioural Data Sets," *Review of Economic Studies*, 79, 375–387.

Salant, Y. and R. Siegel (2016): "Reallocation Costs and Efficiency," *American Economic Journal:Microeconomics*, 8, 203–227.

Shalev, J. (2002): "Loss Aversion and Bargaining," *Theory and Decisions*, 52, 201–232.

Thaler, R. H. (1980): "Toward a positive theory of consumer choice," *Journal of Economic Behavior and Organization*, 1, 39–60.

——— (1999): "Mental Accounting Matters," *Journal of Behavioral Decision Making*, 12, 183–206.

Wolitzky, A. (2016): "Mechanism Design with Maxmin Agents: Theory and an Application to Bilateral Trade," *Theoretical Economics*, 11, 971–1004.