Maastricht University

On loss aversion in bimatrix games

Citation for published version (APA):

Driesen, B. W. I., Perea ý Monsuwé, A., & Peters, H. J. M. (2007). On loss aversion in bimatrix games. (METEOR Research Memorandum; No. 033). Maastricht: METEOR, Maastricht University School of Business and Economics.

Document status and date:
Published: 01/01/2007

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

Bram Driesen, Andrés Perea, Hans Peters

On Loss Aversion in Bimatrix Games

RM/07/033

METE**C**R

On Loss Aversion in Bimatrix Games

Bram Driesen[*]   Andrés Perea[*]   Hans Peters[*]

August 2007

Abstract

In this paper we study three different types of loss aversion equilibria in bimatrix games. Loss aversion equilibria are Nash equilibria of games where players are loss averse and where the reference points – points below which they consider payoffs to be losses – are endogenous to the equilibrium calculation. The first type is the fixed point loss aversion equilibrium, introduced in Shalev (2000) under the name of 'myopic loss aversion equilibrium'. There, the players' reference points depend on the beliefs about their opponents' strategies. The second type, the maximin loss aversion equilibrium, differs from the fixed point loss aversion equilibrium in that the reference point is now only based on the *carrier* of the players' beliefs, not on the exact probabilities. In the third, the safety level loss aversion equilibrium, this dependence is completely dispensed with. Finally, we do a comparative statics analysis of all three equilibrium concepts in 2×2 bimatrix games. The results indicate that a player, under some conditions, benefits from his opponent falsely believing he is loss averse.

1   Introduction

Ever since von Neumann and Morgenstern (1944) developed expected utility theory, it has been the dominant approach in individual decision-making

---
[*]Department of Quantitative Economics, Universiteit Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Telephone: +31-43-3883835. Telefax: +31-43-3884874. Email adresses: `B.Driesen@ke.unimaas.nl`, `A.Perea@ke.unimaas.nl`, `H.Peters@ke.unimaas.nl`.

under uncertainty. In spite of the succes expected utility theory has had in the past, its pervasive role in economic theory has also been subject to a lot of criticism. One of the earliest and most famous examples of this is the Allais paradox (Allais, 1953)[1].

Several alternative models for decision-making under uncertainty have been proposed in the literature. One of the most successful non-expected utility theories is named 'prospect theory', and was developed in Kahneman and Tversky (1979). This theory assumes that economic agents make choices between lotteries in two phases: an editing phase and an evaluation phase. In the editing phase, agents observe and interpret the options between which they must choose using several simple heuristics, one of which is the framing of payoffs as gains or as losses, using a reference point. In the evaluation phase, an agent modifies his utility function to a *reference-dependent* utility function, to account for the perception of the payoffs. That is, perceived losses are weighted downwards, a phenomenon frequently referred to as *loss aversion*. The agent then transforms the probabilities with which payoffs are realized using a probability weighting function, and uses these modified probabilities to calculate the expected reference-dependent utility of the lottery. The latter aspect is ignored in this paper, i.e. we will not transform the probabilities to account for how they are perceived by the players.

Although expected utility theory remains the most important model for rational decision-making in mainstream economic theory, the non-expected utility theories – and prospect theory in particular – have proved to be successful challengers of the expected utility paradigm. Many of these theories have an equally solid mathematical basis as expected utility theory, making them acceptable alternatives for economists. More importantly, they tend to incorporate a number of behavioral patterns, documented in the psychology literature[2] that better explain the decisions of economic agents, and as a consequence, are better able to provide a theoretical basis for several empirically observed phenomena that do not fit with the standard theory of rational choice[3].

Although a number of these behavioral aspects have been applied to the

---

[1]For an overview of the literature on violations of expected utility theory, see for instance Shoemaker (1982) or Machina (1987).

[2]See Kaheneman and Tversky (1979), Hershey et al. (1982), and others.

[3]An overview of puzzles and the solutions proposed by prospect theory is provided in Camerer (2002).

specific field of non-cooperative game theory[4], the effects of loss aversion on non-cooperative games have not been extensively studied. There are some papers in which the outcomes of certain well known examples of games are showed to be consistent with experimental or empirical observations, if the players are assumed to be loss averse[5]. However, this type of literature only focuses on specific examples, and furthermore, often starts from the common assumption that players' reference points are given by some exogenous status quo value. This might not fully reflect the idea of reference-dependence as it was originally intended: Kahneman and Tversky (1981) defined the framing of payoffs as *"the decision-maker's conception of the acts, outcomes, and contingencies associated with a particular choice"* (p. 453). This implies that the reference points of players playing a non-cooperative game can not be fixed ex ante, but must be based on their own strategies (the *acts*), their payoffs (the *outcomes*), and the strategies of their opponents (the *contingencies*). Thus, game theory adds another dimension to the issue of framing payoffs, and to loss aversion in general, that is often ignored.

One paper in which reference-dependence is treated consistently with Tversky and Kahneman's definition is Shalev (2000). There, an equilibrium concept is developed in which each player transforms his basic utility payoffs with a reference point such that his expected reference-dependent equilibrium payoff is exactly equal to that reference point. Thus, the players' reference points can be interpreted as their expected payoffs in equilibrium. In line with Tversky and Kahneman's definition, they then depend on the set of equilibrium strategies, and the players' individual basic utility payoff matrices.

We develop two other equilibrium concepts that take into account the players' loss aversion in a way that is consistent with Kahneman and Tversky's definition. Unlike Shalev, we only consider bimatrix games. The first new concept, called 'maximin loss aversion equilibrium', assumes that each player's reference point is equal to his pure maximin value, taking into account only these pure strategies of the opponent that are played with positive probability. This differs from Shalev's equilibrium concept in two significant ways: first, a player's reference point depends on the *carrier* of the opponent's strategy. In addition, it assumes that players are cautious, in the sense

---

[4]Examples are Crawford (1990), Dekel et al. (1991), and Eichberger and Kelsey (1999).

[5]For example, Fershtman (1996) studies an incumbency game, Berejikian (2002) a.o. the game of chicken and the prisoner's dilemma, and Butler (2007) an ultimatum game.

that they base their expectations on 'worst-case' values. Since a player's reference point depends on the carrier of his opponent's strategy, it can exhibit discontinuous jumps when the opponent's carrier changes. Indeed, maximin loss aversion equilibrium may fail to exist. Nonetheless, we show existence if at least one player has at most two pure strategies.

The loss aversion safety level of a player is the value of the matrix game, derived from the basic payoff matrix with that value as reference point. A safety level loss aversion equilibrium is an equilibrium in the bimatrix game obtained by transforming the basic payoffs with these loss aversion safety levels as reference points. This type of equilibrium shares the fixed point idea with Shalev's loss aversion equilibrium and the cautious player property with the maximin loss aversion equilibrium. However, it is based on reference points that no longer depend on the opponent's equilibrium strategy. It is based on what he *could* do rather than what he *will* do.

We conclude the paper with a comparative statics analysis of the three equilibrium concepts in 2×2 bimatrix games. Specifically, we assume that both players are loss neutral, but that only one player has full information regarding the loss attitudes. That is, both players know they are loss neutral themselves, but only one of them has this information about his opponent. We then study for each equilibrium concept what would happen to the payoff of the fully informed player, if his partially informed opponent falsely believes he (the fully informed player) is loss averse. Under some conditions, we find that this situation benefits the fully informed player.

The paper continues as follows. After preliminaries in Section 2, we discuss the 'myopic loss aversion equilibrium' from Shalev (2000) in Section 3. Section 4 discusses the maximin loss aversion equilibrium, and Section 5 the safety level loss aversion equilibrium. In Section 6 we derive the comparative statics results mentioned above. Finally, Section 7 contains some concluding remarks and some directions for future research.

2   Preliminaries

Before introducing the different equilibrium concepts we first define bimatrix games and Nash equilibrium, and indicate how loss aversion of the players can be accounted for.

2.1 Bimatrix Games and Nash Equilibria

Players 1 and 2 have sets of pure strategies $I = \{1, \ldots, m\}$ and $J = \{1, \ldots, n\}$, respectively. If player 1 plays i and player 2 plays j, then player 1 (2) receives a_{ij} (b_{ij}), the number in the payoff matrix A (B) that corresponds with the j-th element of the i-th row.

Let e^k be a vector filled with zeros that has a one in position k. Let $e = (1, 1, \ldots, 1)$. Then the $(k-1)$-dimensional unit simplex Δ^k is defined as

$$\Delta^k := \{\omega \in \mathbb{R}^k \ : \ e^T \omega = 1 \text{ and } \omega_i \geq 0 \text{ for all } i = 1, \ldots, k\}^6.$$

Players' *mixed* strategies are defined as probability distributions over their sets of *pure* strategies, and can thus be represented by elements of a unit simplex in the appropriately sized Euclidean space. The carrier of a player's strategy is the set of pure strategies that a player plays with a strictly positive probability. That is,

$$Car(p) := \{i \in I : p_i > 0\} \text{ and } Car(q) := \{j \in J : q_j > 0\}.$$

A Nash equilibrium in an $m \times n$ bimatrix game (A, B) is a pair (p^*, q^*), such that $p^* A q^* \geq p A q^*$ for all $p \in \Delta^m$ and $p^* B q^* \geq p^* B q$ for all $q \in \Delta^n$.

2.2 Loss Aversion

The way to introduce loss aversion in two-player games (A, B) is to characterize the players by nonnegative *loss aversion coefficients* λ_1 and λ_2, respectively measuring player 1's and player 2's degrees of loss aversion. Since λ_1 and λ_2 are exogenous characteristics of the players, affecting the resulting equilibrium, we henceforth include them in the definition of the bimatrix game. Thus, a bimatrix game is an object of the form $((A, B), (\lambda_1, \lambda_2))$.

In addition to his loss aversion coefficient, player 1 (2) has a number r_1 (r_2) below which he considers the basic utility payoff entries of A (B) to be losses. These points, r_1 and r_2, are the players' respective *reference points*. The idea of loss aversion is captured by transforming the players' basic utility payoffs as follows:

$$a_{ij}^{\lambda_1, r_1} = a_{ij} - \lambda_1 \max\{r_1 - a_{ij}, 0\},$$
$$b_{ij}^{\lambda_2, r_2} = b_{ij} - \lambda_2 \max\{r_2 - b_{ij}, 0\}.$$

---

[6] We henceforth omit the superscript ^{T}, indicating transposition of the vector, as it is always clear from the context what is meant. For the same reason we will no longer mention the length of the vectors.

Observe that this transformation preserves the ordering over deterministic payoffs. That is, a decision-maker prefers basic utility payoff x to y if and only if he prefers $x^{\lambda,r}$ to $y^{\lambda,r}$ for all $\lambda \geq 0$ and $r \in \mathbb{R}$.

We refer to the transformed game $(A^{\lambda_1,r_1}, B^{\lambda_2,r_2})$ as the *loss aversion game*. For each equilibrium concept considered, we require that it is a Nash equilibrium in the loss aversion game $(A^{\lambda_1,r_1}, B^{\lambda_2,r_2})$, where the reference points r_1 and r_2 are endogenous. The three equilibrium concepts differ in the way the reference points are determined.

3 Fixed Point Loss Aversion Equilibria

In Shalev (2000), a concept of loss aversion equilibrium is introduced where the players' reference points are found through a fixed point calculation. First, define

$$\underline{r} := \min \left\{ \min_{(i,j) \in I \times J} a_{ij}, \min_{(i,j) \in I \times J} b_{ij} \right\}$$

and

$$\bar{r} := \max \left\{ \max_{(i,j) \in I \times J} a_{ij}, \max_{(i,j) \in I \times J} b_{ij} \right\}.$$

In words, $\underline{r}$ and $\bar{r}$ are the lowest resp. the highest payoffs in A or B. Then, given a strategy profile (p,q) and some reference point $r_1 \in [\underline{r}, \bar{r}]$, player 1 has an expected payoff of $pA^{\lambda_1,r_1}q$. Observe that

$$pA^{\lambda_1,\underline{r}}q = pAq \geq \min_{(i,j) \in I \times J} a_{ij} \geq \underline{r}$$

and that

$$\bar{r} \geq \max_{(i,j) \in I \times J} a_{ij} \geq pAq \geq pA^{\lambda_1,\bar{r}}q.$$

This and the fact that $pA^{\lambda_1,r_1}q$ is a continuous function of r_1 implies there is an $r_1^* \in [\underline{r}, \bar{r}]$ such that $r_1^* = pA^{\lambda_1,r_1^*}q$. Furthermore, r_1^* is unique because r_1 is strictly increasing on $[\underline{r}, \bar{r}]$, while $pA^{\lambda_1,r_1}q$ is non-increasing on $[\underline{r}, \bar{r}]$. Similarly, there is a unique $r_2^* \in [\underline{r}, \bar{r}]$ such that $r_2^* = pB^{\lambda_2,r_2^*}q$. Clearly, these 'fixed point' reference points can be interpreted as the utilities players expect to realize given the strategy profile (p,q).

Next, Shalev introduces a non-empty, compact- and convex-valued correspondence $\beta : \Delta^m \times \Delta^n \times [\underline{r}, \overline{r}]^2 \to \Delta^m \times \Delta^n \times [\underline{r}, \overline{r}]^2$ where

$$\beta(\hat{p}, \hat{q}, (\hat{r}_1, \hat{r}_2)) := \{(p, q, (r_1, r_2)) \in \Delta^m \times \Delta^n \times [\underline{r}, \overline{r}]^2 :$$
$$r_1 = pA^{\lambda_1, \hat{r}_1}\hat{q} \geq p'A^{\lambda_1, \hat{r}_1}\hat{q} \text{ for all } p' \in \Delta^m, \text{ and}$$
$$r_2 = \hat{p}B^{\lambda_2, \hat{r}_2}q \geq \hat{p}B^{\lambda_2, \hat{r}_2}q' \text{ for all } q' \in \Delta^n\}.$$

Since the (Nash) best reply-correspondence is upper semicontinuous and the players' payoff functions are continuous in their respective reference points, it follows that the correspondence β is also upper semicontinuous. Hence, by Kakutani's fixed point theorem there exists a fixed point $(p^*, q^*, (r_1^*, r_2^*))$. Note that the strategy pair (p^*, q^*) is a Nash equilibrium in the loss aversion game $(A^{\lambda_1, r_1^*}, B^{\lambda_2, r_2^*})$. Since the reference points are determined through a fixed point calculation, we refer to this equilibrium concept as a *fixed point loss aversion equilibrium.*

Although this is an effective way of dealing with loss aversion, it is certainly not the only possible approach. One of the less attractive features of this concept is that reference points are not unique: two loss aversion equilibria generally do not yield the same expected payoffs to the players. Furthermore, a player's reference point depends heavily on his own beliefs about the opponent's strategy. In what follows, we discuss some alternative equilibrium concepts that, if nothing else, at least broaden our view on the issues with endogenous reference point determination.

4   Maximin Loss Aversion Equilibria

In maximin loss aversion equilibrium, each player chooses his reference point in such a way that his maximin payoff w.r.t. the strategies he believes his opponent plays with positive probability, is exactly equal to that reference point. The Nash equilibria in the game that results from using these consistent reference points are maximin loss aversion equilibria.

Maximin loss aversion equilibria are similar to fixed point loss aversion equilibria, because in both concepts players base their reference points on the carriers of their opponents' strategies. In particular, in fixed point loss aversion equilibrium they base them on the probability distribution they believe their opponents have over the pure strategies in their carriers. In maximin loss aversion equilibrium, they also believe there exists some probability distribution over the pure strategies in their opponents' carriers, but

they recognize that only one of these is eventually played. They then base their reference points on the worst possible strategy their opponents could play.

4.1 Definition of Maximin Loss Aversion Equilibria

Formally, player 1 and 2's reference points, given the strategy profile $(p, q) \in \Delta^m \times \Delta^n$, are given by

$$r_1^* := \max_{i \in I} \ \min_{j \in Car(q)} a_{ij}^{\lambda_1, r_1^*} \ \text{ resp. } \ r_2^* := \max_{j \in J} \ \min_{i \in Car(p)} b_{ij}^{\lambda_2, r_2^*}.$$

Given r_1^* and r_2^*, each player has a minimal reference-dependent utility payoff for each of his own pure strategies, given the strategies his opponent plays with positive probability. The reference points are chosen such that for each player the maximum in this set of minimal payoffs, i.e. the maximin payoff, is equal to his reference point.

Observe that we could also do without the consistency argument here. This is so because for all $r_1 \in [\underline{r}, \ \max_{i \in I} \min_{j \in Car(q)} a_{ij}]$, we have that

$$\max_{i \in I} \ \min_{j \in Car(q)} a_{ij}^{\lambda_1, r_1} = \max_{i \in I} \ \min_{j \in Car(q)} a_{ij}.$$

That is, for these values of r_1 we have that the maximin point in the transformed utility payoff matrix, given the opponent's strategy q, is equal to the maximin point in the basic utility payoff matrix, given the opponent's strategy q. Hence, using the reference point $r_1 = \max_{i \in I} \min_{j \in Car q} a_{ij}$ to transform the basic utilities, yields a transformed payoff matrix for which the maximin point, given the strategy q, is exactly equal to that reference point. A similar argument also holds for player 2. Thus, an equivalent definition of the maximin reference points is

$$r_1^* := \max_{i \in I} \ \min_{j \in Car(q)} a_{ij} \ \text{ resp. } \ r_2^* := \max_{j \in J} \ \min_{i \in Car(p)} b_{ij}.$$

Note that a maximin point is unique for each carrier played by the opponent.

Since a player's maximin payoff only depends on the carrier of the strategy played by his opponent, rather than the strategy itself, reference points are more robust against wrong beliefs a player may have about his opponent.

A *maximin loss aversion equilibrium* in a bimatrix game $((A, B), (\lambda_1, \lambda_2))$ is a pair of vectors $(p^*, q^*) \in \Delta^m \times \Delta^n$ such that $p^* A^{\lambda_1, r_1^*} q^* \geq p A^{\lambda_1, r_1^*} q^*$ for

all $p \in \Delta^m$ and $p^* B^{\lambda_2, r_2^*} q^* \geq p^* B^{\lambda_2, r_2^*} q$ for all $q \in \Delta^n$, where r_1^* and r_2^* are the maximin points in A^{λ_1, r_1^*} resp. B^{λ_2, r_2^*} given q^* resp. p^*.

Since a player's reference point only depends on the *carrier* of his opponent's strategy, rather than the strategy itself, this maximin method might provide a valid alternative for fixed point loss aversion. On the other hand, it does not solve the problem of multiple reference points. Furthermore, because reference points no longer depend *continuously* on the strategies played by the opponent, maximin loss aversion equilibria may fail to exist.

4.2 Existence of Maximin Loss Aversion Equilibria

We show by means of a counterexample that maximin loss aversion equilibria may fail to exist. Next, we show existence if one of the players has no more than two pure strategies.

4.2.1 An Example Showing Non-Existence

Consider the following 3×3 bimatrix game:

$$A = \begin{bmatrix} 8 & 1 & 0 \\ 1 & 8 & 0 \\ 4 & 4 & -1 \end{bmatrix} \qquad B = \begin{bmatrix} -4 & 1 & -2 \\ 1 & -4 & -2 \\ -2 & -2 & -1 \end{bmatrix}.$$

Let player 2 be loss neutral, i.e. $\lambda_2 = 0$, and assume $\lambda_1 = 1$. Because of player 2's loss neutrality, r_2 has no influence on the equilibrium. That is, $B = B^{\lambda_2, r_2}$ for all values of r_2. Observe that player 1's best reply against player 2 playing e^1 is e^1, and player 2's best reply against this is e^2. Hence, e^1 can never be an equilibrium strategy for player 2. Similarly, we can exclude e^2 as one of player 2's equilibrium strategies. This implies that player 1's equilibrium reference point is never equal to 8. This leaves two possibilities: $r_1 = 0$ or $r_1 = 4$.

- $r_1 = 0$: In this case, we have

$$A^{\lambda_1, 0} = \begin{bmatrix} 8 & 1 & 0 \\ 1 & 8 & 0 \\ 4 & 4 & -2 \end{bmatrix}.$$

The unique Nash equilibrium in $(A^{\lambda_1, 0}, B)$ is $((.5, .5, 0), (.5, .5, 0))$ implying $r_1 = 4$.

- $r_1 = 4$: In this case, we have

$$A^{\lambda_1,4} = \begin{bmatrix} 8 & -2 & -4 \\ -2 & 8 & -4 \\ 4 & 4 & -6 \end{bmatrix}.$$

The unique Nash equilibrium in $(A^{\lambda_1,4}, B)$ is $\left(\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)\right)$, implying $r_1 = 0$.

Each of player 1's possible reference points implies a loss aversion game in which the carrier of player 2's equilibrium strategy is such that another reference point should be chosen. Hence, there is no maximin loss aversion equilibrium. This implies that at least one of the conditions underlying Kakutani's fixed point theorem is violated for the 'maximin' best-reply correspondence.

Define the players' respective maximin best reply correspondences $\beta_1 : \Delta^n \to \Delta^m$ and $\beta_2 : \Delta^m \to \Delta^n$ as

$$\beta_1(\hat{q}) := \{p \in \Delta^m : pA^{\lambda_1,r_1}\hat{q} \geq p'A^{\lambda_1,r_1}\hat{q} \text{ for all } p' \in \Delta^m \text{ and}$$
$$r_1 = \max_{i \in I} \min_{j \in Car(\hat{q})} a_{ij}\},$$

and

$$\beta_2(\hat{p}) := \{q \in \Delta^n : \hat{p}B^{\lambda_2,r_2}q \geq \hat{p}B^{\lambda_2,r_2}q' \text{ for all } q' \in \Delta^n \text{ and}$$
$$r_2 = \max_{j \in J} \min_{i \in Car(\hat{p})} b_{ij}\}.$$

Then define the overall maximin best-reply correspondence $\beta : \Delta^m \times \Delta^n \to \Delta^m \times \Delta^n$ as $\beta := \beta_1 \times \beta_2$. Note that for all $(\hat{p}, \hat{q})$ there is a single pair of payoff matrices $(A^{\lambda_1,r_1}, B^{\lambda_2,r_2})$, and the set of best replies against $(\hat{p}, \hat{q})$ given these payoff matrices is non-empty, compact and convex. It follows that β is non-empty, and compact- and convex-valued. The fact that a maximin loss aversion equilibrium need not exist then implies that upper semicontinuity of β is violated. More specifically, there are sequences (p^k, q^k) and $(\hat{p}^k, \hat{q}^k)$ in $\Delta^m \times \Delta^n$ converging to (p^*, q^*) resp. $(\hat{p}^*, \hat{q}^*)$ and satisfying $(p^k, q^k) \in \beta(\hat{p}^k, \hat{q}^k)$ for all $k \in \mathbb{N}$ that do not satisfy $(p^*, q^*) \in \beta(\hat{p}^*, \hat{q}^*)$.

In our example, suppose player 2 has a sequence of strategies given by

$$\hat{q}^k = \begin{bmatrix} \frac{1}{3} + \frac{1}{6}\left(1 - \frac{1}{k}\right) \\ \frac{1}{3} + \frac{1}{6}\left(1 - \frac{1}{k}\right) \\ \frac{1}{3} - \frac{1}{3}\left(1 - \frac{1}{k}\right) \end{bmatrix}.$$

Note that this implies $r_1^k = 0$ for all $k \in \mathbb{N}$. Then player 1 obtains $\frac{9}{2} - \frac{3}{2k}$ from his first two strategies, and $4 - \frac{2}{k}$ from his second. Thus, $\beta_1(\hat{q}^k) = \{(\alpha, 1 - \alpha, 0) \text{ with } \alpha \in [0,1]\}$ for all $k \in \mathbb{N}$. The sequence $\hat{q}^k$ converges to $\hat{q}^* = (\frac{1}{2}, \frac{1}{2}, 0)$. Then $r_1^* = 4$ from which it follows that player 1 obtains a payoff of 3 for the first two strategies and 4 for the third one. Hence, $\beta_1(\hat{q}^*) = \{(0, 0, 1)\}$. Since β_1 does not satisfy upper semicontinuity, neither does β.

4.2.2 Existence in $m \times 2$ and $2 \times n$ Games

Although maximin loss aversion equilibria do not exist in general, we do have existence in the case where one of the players has no more than two pure strategies. In the more restrictive setting of 2×2 bimatrix games, we can simply apply Kakutani's fixed point theorem because the maximin best-reply correspondence β is upper semicontinuous. To see this, let $((A, B), (\lambda_1, \lambda_2))$ be a 2×2 bimatrix game, and let $\hat{q}^k$ be a sequence in Δ^2 converging to $\hat{q}^*$, with $\hat{q}_2^k > 0$ for all $k \in \mathbb{N}$, while $\hat{q}_2^* = 0$. Furthermore, let p^k be a sequence in Δ^2, converging to p^*, such that $p^k \in \beta_1(\hat{q}^k)$ for all $k \in \mathbb{N}$. To show that β_1 is upper semicontinuous, it is then sufficient to show that $p^* \in \beta_1(\hat{q}^*)$.

The fact that $p^k \in \beta_1(\hat{q}^k)$ for all k, implies that $p^k A^{\lambda_1, r_1^k} \hat{q}^k \geq p' A^{\lambda_1, r_1^k} \hat{q}^k$ for all $p' \in \Delta^2$ and $k \in \mathbb{N}$, where

$$r_1^k := \max_{i \in I} \min_{j \in Car(\hat{q}^k)} a_{ij}.$$

Let $H := Car(\hat{q}^k)$ for some $k \in \mathbb{N}$. Then $p^* A^{\lambda_1, r_1} \hat{q}^* \geq p' A^{\lambda_1, r_1} \hat{q}^*$ for all $p' \in \Delta^2$, where

$$r_1 := \max_{i \in I} \min_{j \in H} a_{ij}.$$

Observe that player 2 now plays the first column with certainty. That is, $\hat{q}_2^* = 0$. Since transforming a player's payoffs to incorporate his attitude towards losses, leaves his preference ordering over pure alternatives unchanged, this implies that player 1's best reply against $\hat{q}^*$, regardless of his reference point, is to play a strategy that yields him the highest payoff in the first column. That is,

$$p^* \in \arg\max_{p \in \Delta^2} p A^{\lambda_1, \rho} \hat{q}^* = \arg\max_{p \in \Delta^2} p A \hat{q}^*$$

for all $\rho \in [\underline{r}, \bar{r}]$. Hence, $p^* A^{\lambda_1, r_1^*} \hat{q}^* \geq p' A^{\lambda_1, r_1^*} \hat{q}^*$ for all $p' \in \Delta^2$ where

$$r_1^* := \max_{i \in I} \min_{j \in Car(\hat{q}^*)} a_{ij}^{\lambda_1, r_1^*}.$$

Hence, $p^* \in \beta_1(\hat{q}^*)$, implying that player 1's maximin best-reply correspondence β_1 is upper semicontinuous. Using a similar argument we can show that the same holds for β_2. It follows that the joint maximin best-reply correspondence β also satisfies upper semicontinuity, and thus - by Kakutani's fixed point theorem - that maximin loss aversion equilibria exist in 2×2 bimatrix games.

In a more general $m \times 2$ (or $2 \times n$) setting, the maximin best-reply correspondence no longer satisfies upper semicontinuity. To see this, consider the following example:

$$B = \begin{bmatrix} 1 & 4 \\ -5 & 5 \\ -2 & -4 \end{bmatrix}.$$

Furthermore, let

$$\hat{p}^k = \begin{bmatrix} \frac{1}{3} + \frac{1}{6}(1 - \frac{1}{2k}) \\ \frac{1}{3} - \frac{1}{3}(1 - \frac{1}{2k}) \\ \frac{1}{3} + \frac{1}{6}(1 - \frac{1}{2k}) \end{bmatrix},$$

and assume $\lambda_2 > \frac{1}{2}$. Since $Car(\hat{p}^k) = \{1, 2, 3\}$, the reference point of player 2 is $r_2^k = -4$. Hence, the transformed payoff matrix is

$$B^{\lambda_2, r_2^k} = \begin{bmatrix} 1 & 4 \\ -(5 + \lambda_2) & 5 \\ -2 & -4 \end{bmatrix}.$$

Given $\hat{p}^k$, player 2 obtains $-\frac{1}{2} - \frac{9 + 2\lambda_2}{12k}$ from playing column 1, and $\frac{5}{6k}$ from playing column 2, which implies $\beta_2(\hat{p}^k) = \{(0, 1)\}$ for all $k \in \mathbb{N}$. As k converges to infinity, $\hat{p}^k$ converges to $\hat{p}^* = (.5, 0, .5)$. Then player 2's reference point changes to $r_2^* = -2$. The associated loss aversion payoff matrix is

$$B^{\lambda_2, r_2^*} = \begin{bmatrix} 1 & 4 \\ -(5 + 3\lambda_2) & 5 \\ -2 & -(4 + 2\lambda_2) \end{bmatrix}.$$

Given p^*, player 2 obtains a payoff of $-\frac{1}{2}$ from playing the first column, and $-\lambda_2$ from the second. The initial assumption $\lambda_2 > \frac{1}{2}$ implies $\beta_2(\hat{p}^*) = \{(1, 0)\}$. It follows that player 2's maximin best-reply correspondence is not upper semicontinuous. Consequently, existence of maximin loss aversion equilibria in $m \times 2$ (or $2 \times n$) games can no longer be proven by means

of Kakutani's fixed point theorem as in the more restrictive 2×2 setting. However, we can still prove existence.

Proposition 4.1 *For all $\lambda_1, \lambda_2 \geq 0$ and $m \times 2$ matrices A and B, there exists a maximin loss aversion equilibrium in the game $((A, B), (\lambda_1, \lambda_2))$. Similarly, for all $\lambda_1, \lambda_2 \geq 0$ and $2 \times n$ matrices A and B, there exists a maximin loss aversion equilibrium in the game $((A, B), (\lambda_1, \lambda_2))$.*

    Proof. Let A, B be $m \times 2$ matrices, and consider the game $((A, B), (\lambda_1, \lambda_2))$. Define

$$\tilde{r} := \max_{i \in I} \min_{j \in J} a_{ij},$$

and let (p^*, q) be a Nash equilibrium in the game $(A^{\lambda_1, \tilde{r}}, B)$. We can distinguish three (exhaustive) cases:

i. $q = e^t$ where $t \in J$, i.e. player 2's equilibrium strategy is pure. This implies there is an $s \in Car(\hat{p}^*)$ such that $b_{st} = \max_{j \in J} b_{sj}$, because otherwise strategy t would be strictly dominated by the other strategy in J. Since loss aversion preserves the agent's preference ordering over pure alternatives, this implies $b_{st}^{\lambda_2, b_{st}} = \max_{j \in J} b_{sj}^{\lambda_2, b_{st}}$. From the fact that $s \in Car(\hat{p}^*)$, it follows that $a_{st}^{\lambda_1, \tilde{r}} = \max_{i \in I} a_{it}^{\lambda_1, \tilde{r}}$, and thus also $a_{st}^{\lambda_1, a_{st}} = \max_{i \in I} a_{it}^{\lambda_1, a_{st}}$. Hence, (e^s, e^t) is a pure loss aversion equilibrium in $((A, B), (\lambda_1, \lambda_2))$.

ii. $(p^*, q) = (e^s, (\beta, 1 - \beta))$ where $s \in I$ and $\beta \in (0, 1)$. This implies

$$\beta a_{s1}^{\lambda_1, r_1} + (1 - \beta) a_{s2}^{\lambda_1, r_1} \geq \beta a_{i1}^{\lambda_1, r_1} + (1 - \beta) a_{i2}^{\lambda_1, r_1}$$

for all $i \in I$ with $r_1 = \tilde{r}$. Furthermore, it implies $b_{s1} = b_{s2} =: b$ from which it follows that $b_{s1}^{\lambda_2, r_2} = b_{s2}^{\lambda_2, r_2}$ with $r_2 = b$. Hence, $(e^s, (\beta, 1 - \beta))$ is a maximin loss aversion equilibrium in $((A, B), (\lambda_1, \lambda_2))$.

iii. p^* and q satisfy $|Car(p^*)| \geq 2$ and $|Car(q)| = 2$. Then there must exist pure strategies s and s' in $Car(p^*)$ such that

- either $b_{s1} > b_{s2}$ and $b_{s'1} < b_{s'2}$,
- or $b_{s1} < b_{s2}$ and $b_{s'1} > b_{s'2}$,
- or $b_{s1} = b_{s2}$ and $b_{s'1} = b_{s'2}$.

In all three cases, there exists an $\alpha \in (0,1)$ such that both player 2's pure strategies are best replies against player 1's strategy $(\alpha e^s + (1 - \alpha)e^{s'})$, with payoffs given by B^{λ_2, r_2} where $r_2 = \max_{j \in J} \min_{i \in \{s, s'\}} b_{ij}$. But then $((\alpha, 1 - \alpha), q)$ is a maximin loss aversion equilibrium in the game $((A, B), (\lambda_1, \lambda_2))$.

Since all types of Nash equilibria in $(A^{\lambda_1, \tilde{r}_1}, B)$ lead to a maximin loss aversion equilibrium, this establishes existence in $m \times 2$ bimatrix games. The proof of existence in $2 \times n$ bimatrix games is analogous to the $m \times 2$ case. ∎

Note that this proposition also provides a method to calculate maximin loss aversion equilibria. One can modify player 1's payoff matrix under the assumption that player 2 plays both his strategies with a positive probability, and calculate the Nash equilibria in the resulting game. These Nash equilibria can then be transformed into maximin loss aversion equilibria.

5   Safety Level Loss Aversion Equilibria

Here, we introduce the safety level loss aversion equilibrium. Each player takes as a reference point his safety level, i.e. the maximum payoff he would receive if his opponent's only objective were to minimize this expected payoff. This is similar to maximin loss aversion equilibria, but less pessimistic in the sense that players' safety levels are at least as high as their maximin points, given the opponents' carriers. On the other hand, one could argue that the safety level loss aversion equilibrium is more pessimistic than the maximin loss aversion equilibrium, because it is independent of the opponent's strategy in equilibrium, and thus does not exclude the opponent's pure non-equilibrium strategies. Aside from the discussion which concept is more optimistic, the safety level loss aversion equilibrium has some obvious advantages: it not only solves the existence problem, it also provides an answer to the question of multiple reference points.

5.1   Strict Dominance when Players are Loss Averse

A strategy is strictly dominated if there is another strategy that yields a higher payoff for every strategy of the opponent. Thus, a strictly dominated strategy is never in the carrier of a player's equilibrium strategy. Since in the

previous equilibrium concepts, each reference point depended on the opponent's equilibrium strategy, it follows that strictly dominated strategies did not influence the players' equilibrium reference points. Thus, strict dominance did not play a role so far.

In the safety level loss aversion equilibrium concept, however, strictly dominated strategies could play a role. That is, if a player has a strictly dominated strategy, then his opponent could also determine his reference point using the reduced payoff matrix, i.e. the payoff matrix from which the strictly dominated strategy was removed. Elimination of such strategies generally leads to different reference points, and consequently also to different equilibria. To account for the fact that players may be loss averse, we have to extend the standard definition of strict dominance.

Formally, we say that a pure strategy $i \in I$ is a strictly dominated strategy in A if there is a strategy $p \in \Delta^m$ with $p_i = 0$, such that $pAe^j > e^i Ae^j$ for all $j \in J$. A pure strategy $i \in I$ is said to be strictly dominated in (A, λ_1), if it is strictly dominated in $A^{\lambda_1, \rho}$ for all $\rho \in [\underline{r}, \bar{r}]$. Then, to eliminate a pure strategy $i \in I$ from the game, it is no longer sufficient that it is strictly dominated in A. To see this, consider the following example:

$$A = \begin{bmatrix} 5 & 0 \\ 0 & 5 \\ 2 & 2 \end{bmatrix}.$$

Observe that player 1's pure strategy e^3 is strictly dominated by the mixed strategy $(0.5, 0.5, 0)$. Now let $\lambda_1 = 1$ and $r_1 = 2$. Then the transformed payoff matrix is

$$A^{1,2} = \begin{bmatrix} 5 & -2 \\ -2 & 5 \\ 2 & 2 \end{bmatrix}.$$

The conditions for e^3 to be strictly dominated by some $p = (p_1, p_2, 0)$ are

$$5p_1 - 2p_2 > 2, \text{ and}$$
$$-2p_1 + 5p_2 > 2,$$

which together imply $3 > 4$. Thus, e^3 is no longer strictly dominated in $A^{1,2}$, even though it was in A. Since $A = A^{\lambda_1, \underline{r}}$, i being strictly dominated in A is still a *necessary* condition for i to be strictly dominated in (A, λ_1). The following proposition gives a necessary and sufficient condition.

Proposition 5.1 *In the game $((A, B), (\lambda_1, \lambda_2))$ where $\lambda_1 > 0$, a strategy $p \in \Delta^m$ strictly dominates the pure strategy $i \in I$ in (A, λ_1) if and only if p strictly dominates i in $A^{\lambda_1, \rho}$ for all*

$$\rho \in \left[\min_{j \in J} a_{ij}, \max_{(i',j) \in Car(p) \times J} a_{i'j} \right].$$

Proof. Let a pure strategy i be strictly dominated by a strategy p in the payoff matrix A. Then define $\underline{\rho} := \min_{j \in J} a_{ij}$ and $\bar{\rho} := \max_{(i',j) \in Car(p) \times J} a_{i'j}$.

$\Leftarrow$: Let the pure strategy i be strictly dominated in (A, λ_1). Then it is strictly dominated in $A^{\lambda_1, \rho}$ for all $\rho \in [\underline{r}, \bar{r}]$. Since $[\underline{\rho}, \bar{\rho}] \subseteq [\underline{r}, \bar{r}]$, strategy i is strictly dominated in $A^{\lambda_1, \rho}$ for all $\rho \in [\underline{\rho}, \bar{\rho}]$.

$\Rightarrow$: Let the pure strategy i be strictly dominated by p in $A^{\lambda_1, \rho}$ for all $\rho \in [\underline{\rho}, \bar{\rho}]$. Then p strictly dominates i in $A^{\lambda_1, \underline{\rho}}$. That is,

$$p A^{\lambda_1, \underline{\rho}} e^j > e^i A^{\lambda_1, \underline{\rho}} e^j$$

for all $j \in J$. Let $\rho \in [\underline{r}, \underline{\rho}]$. Then $e^i A^{\lambda_1, \underline{\rho}} e^j = e^i A^{\lambda_1, \rho} e^j$. Furthermore, $p A^{\lambda_1, \underline{\rho}} e^j \leq p A^{\lambda_1, \rho} e^j$ since $\underline{\rho} \geq \rho$. Hence, $p A^{\lambda_1, \rho} e^j > e^i A^{\lambda_1, \rho} e^j$ as well. So p strictly dominates i in $A^{\lambda_1, \rho}$ for all $\rho \in [\underline{r}, \underline{\rho}]$.

Similarly, p strictly dominating i in $A^{\lambda_1, \rho}$ for all $\rho \in [\underline{\rho}, \bar{\rho}]$ implies that p strictly dominates i in $A^{\lambda_1, \bar{\rho}}$. That is, $p A^{\lambda_1, \bar{\rho}} e^j > e^i A^{\lambda_1, \bar{\rho}} e^j$ for all $j \in J$. Observe that for all $\rho \in [\bar{\rho}, \bar{r}]$, we have that

$$p A^{\lambda_1, \rho} e^j = (1 + \lambda_1) p A e^j - \lambda_1 \rho.$$

Note that for all $j \in J$ there is an $i' \in Car(p)$ such that $e^i A^{\lambda_1, \rho} e^j \leq e^{i'} A^{\lambda_1, \rho} e^j$. Hence, for all $j \in J$ we have that $e^i A^{\lambda_1, \rho} e^j \leq \rho$ for all $\rho \in [\bar{\rho}, \bar{r}]$, implying

$$e^i A^{\lambda_1, \rho} e^j = (1 + \lambda_1) e^i A e^j - \lambda_1 \rho$$

for all $j \in J$ and $\rho \in [\bar{\rho}, \bar{r}]$. Hence, $(1 + \lambda_1) p A e^j - \lambda_1 \bar{\rho} > (1 + \lambda_1) e^i A e^j - \lambda_1 \bar{\rho}$ for all $j \in J$, and the inequality is preserved if we replace $\bar{\rho}$ by any $\rho \in [\bar{\rho}, \bar{r}]$. But then $p A^{\lambda_1, \rho} e^j > e^i A^{\lambda_1, \rho} e^j$ for all $j \in J$ and $\rho \in [\bar{\rho}, \bar{r}]$. That is, p strictly dominates i in $A^{\lambda_1, \rho}$ for all $\rho \in [\bar{\rho}, \bar{r}]$.

In conclusion, p strictly dominates i in $A^{\lambda_1, \rho}$ for all $\rho \in [\underline{r}, \bar{r}]$, which means that p strictly dominates i in (A, λ_1). ∎

Remark 5.2 It could happen that none of the l strategies $p_1, \ldots, p_l \in \Delta^m$ dominate a pure strategy i in (A, λ_1), while taken together they do. Suppose none of the strategies $p_1, \ldots, p_l$ dominates i in (A, λ_1). Then there are intervals $R_1, \ldots, R_l \subseteq [\underline{r}, \bar{r}]$ such that for all $k = 1, \ldots, l$ we have $p_k A^{\lambda_1, \rho} e^j \leq e^i A^{\lambda_1, \rho} e^j$ for all $\rho \in R_k$ and some $j \in J$. Then, as long as $\bigcap_{k=1}^l R_k = \emptyset$, there is always a strategy p_k strictly better than pure strategy i.

In the following section, when we are talking about a bimatrix game $((A, B), (\lambda_1, \lambda_2))$, we could assume that the payoff matrices A and B are the result of iterated elimination of strategies that are either strictly dominated in (A, λ_1) and (B, λ_2) respectively, or strictly dominated in the weaker sense, explained in the above remark.

5.2 Definition and Existence of Safety Level Loss Aversion Equilibria

The safety level is a concept that dates back to Janos von Neumann's analysis of zero-sum games in von Neumann (1928). Suppose that for each of his strategies, player i assumes that player j plays that (possibly mixed) strategy that minimizes i's expected payoff. Then player i can play that (possibly mixed) strategy that yields him the highest such minimal payoff. Von Neumann showed that when player j adopts the above strategy, then player i obtains the exact same payoff. This payoff is referred to as his *safety level*.

Formally, given a regular bimatrix game with loss neutral players (A, B), the players' safety levels, denoted by $v_1(A)$ resp. $v_2(B)$, are defined by

$$v_1(A) := \max_{p \in \Delta^m} \min_{q \in \Delta^n} pAq, \quad \text{and} \quad v_2(B) := \max_{q \in \Delta^n} \min_{p \in \Delta^m} pBq.$$

Since a player can guarantee his safety level, it would make an intuitively appealing reference point in a bimatrix game where players are loss averse, i.e. $((A, B), (\lambda_1, \lambda_2))$. However, if a player's basic utility payoff matrix has no saddle points, then this implies there is a strictly positive probability that he obtains a payoff lower than his safety level, and thus incurs a loss. Hence, we again require a consistency argument in the sense that a player's payoff matrix should be transformed using a reference point, such that his resulting safety level is equal to that reference point. That is, we are looking for an r_1^* and r_2^* such that $r_1^* := v_1(A^{\lambda_1, r_1^*})$ and $r_2^* := v_2(B^{\lambda_2, r_2^*})$. Such reference points are called *loss aversion safety levels*.

Consider the bimatrix game $((A, B), (\lambda_1, \lambda_2))$. To show that there is a unique r_1^* such that $r_1^* := v_1(A^{\lambda_1, r_1^*})$, it is sufficient to show that $v_1(A^{\lambda_1, r_1})$ is a continuous, non-increasing function of r_1 on the interval $[\underline{r}, \bar{r}]$, and that

$$v_1(A^{\lambda_1, \underline{r}}) \geq \underline{r}, \quad \text{and} \quad v_1(A^{\lambda_1, \bar{r}}) \leq \bar{r}.$$

It is obvious that $v_1(A^{\lambda_1, r_1})$ is continuous in r_1. To show that it is non-increasing in r_1 on the interval $[\underline{r}, \bar{r}]$, let r_1 and s_1 be some reference points in $[\underline{r}, \bar{r}]$ with $r_1 > s_1$, and let

$$p^* \in \arg \max_{p \in \Delta^m} \min_{q \in \Delta^n} \ pA^{\lambda_1, r_1}q, \text{ and}$$

$$q^* \in \arg \min_{q \in \Delta^n} \ p^* A^{\lambda_1, s_1}q.$$

We then have

$$
\begin{aligned}
v_1(A^{\lambda_1, s_1}) \ &= \ \max_{p \in \Delta^m} \min_{q \in \Delta^n} pA^{\lambda_1, s_1}q \\
&\geq \ \min_{q \in \Delta^n} p^* A^{\lambda_1, s_1}q \\
&= \ p^* A^{\lambda_1, s_1}q^*.
\end{aligned}
$$

Note that $pA^{\lambda_1, r_1}q$ is non-increasing in r_1 for any given strategy pair (p, q). Therefore, $p^* A^{\lambda_1, s_1}q^* \geq p^* A^{\lambda_1, r_1}q^*$. Now observe that

$$
\begin{aligned}
p^* A^{\lambda_1, r_1}q^* \ &\geq \ \min_{q \in \Delta^n} p^* A^{\lambda_1, r_1}q \\
&= \ \max_{p \in \Delta^m} \min_{q \in \Delta^n} pA^{\lambda_1, r_1}q \\
&= \ v_1(A^{\lambda_1, r_1}).
\end{aligned}
$$

It follows that a player's unique safety level is a non-increasing continuous function of his reference point. Furthermore,

$$v_1(A^{\lambda_1, \underline{r}}) = v_1(A) \geq \min_{(i,j) \in I \times J} a_{ij} \geq \underline{r},$$

and since $v_1(A^{\lambda_1, r_1})$ is non-increasing in r_1, we also have

$$v_1(A^{\lambda_1, \bar{r}}) \leq v_1(A^{\lambda_1, \underline{r}}) = v_1(A) \leq \max_{(i,j) \in I \times J} a_{ij} \leq \bar{r}.$$

Thus, there must be a unique $r_1^* \in [\underline{r}, \bar{r}]$ such that $r_1^* = v_1(A^{\lambda_1, r_1^*})$. Similarly, if player 2 is loss averse, then there exists a unique $r_2^* \in [\underline{r}, \bar{r}]$ such that

$r_2^* = v_2(B^{\lambda_2, r_2^*})$. In conclusion, given $((A, B), (\lambda_1, \lambda_2))$, each player has a unique loss aversion safety level.

The players transform their payoff matrices using their degrees of loss aversion and their loss aversion safety levels. The safety level loss aversion equilibria are the Nash equilibria in the transformed game. Formally, a *safety level loss aversion equilibrium* is a strategy pair (p^*, q^*), such that $p^* A^{\lambda_1, r_1^*} q^* \geq p A^{\lambda_1, r_1^*} q^*$ for all $p \in \Delta^m$, and $p^* B^{\lambda_2, r_2^*} q^* \geq p^* B^{\lambda_2, r_2^*} q$ for all $q \in \Delta^n$, where $r_1^* := v_1(A^{\lambda_1, r_1^*})$ and $r_2^* := v_2(B^{\lambda_2, r_2^*})$.

Existence of a unique loss aversion safety level and the existence of Nash equilibria together imply general existence of safety level loss aversion equilibria. Here, in contrast with the previous equilibrium concepts, the players' reference points no longer depend on the strategies they believe their opponents will play in equilibrium, but rather on the strategies their opponents would play in a worst-case scenario.

6 Comparative Statics

In this section we consider the effect loss aversion has on the equilibrium payoff of a player in 2×2 bimatrix games. More specifically, we investigate how a player's expected payoff changes when he claims to be loss averse, when in reality, he is neutral against losses, given that the opponent is also loss neutral. The latter assumption implies that we compare the loss aversion equilibrium payoff of a loss neutral player who claims to be loss averse to his standard Nash equilibrium payoff. First, we impose some conditions on the payoffs to make sure there is a mixed Nash equilibrium. Then we obtain sufficient conditions for a player to benefit from his opponent believing he is loss averse, when in fact he is not. Finally, we check for each equilibrium concept whether these sufficient conditions hold.

6.1 Preliminaries

To circumvent the existence problem for maximin equilibria we only consider 2×2 bimatrix games. The basic utilities are represented by the following matrices:

$$A := \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and } B := \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

Since pure Nash equilibria are equivalent to the pure versions of all three types of loss aversion equilibrium considered above, we can restrict ourselves to mixed equilibria. Following Berden and Peters (2006), we exclude the case where player 1 has a weakly dominant strategy, implying $a_{11} \neq a_{21}$ and $a_{12} \neq a_{22}$. W.l.o.g. assume $a_{11} > a_{21}$, $a_{12} < a_{22}$, and $a_{11} \geq a_{22}$. This leaves three exhaustive cases:

  i. $a_{21} \geq a_{22}$;

 ii. $a_{22} \geq a_{21} \geq a_{12}$;

iii. $a_{12} \geq a_{21}$.

In addition, assume $b_{11} < b_{12}$ and $b_{21} > b_{22}$. Hence, player 2 has no weakly dominant strategy.

Let (p^*, q^*) be a Nash equilibrium in this game. The conditions on the players' payoffs imply that (p^*, q^*) is a completely mixed, unique Nash equilibrium. Then

$$p^* = \begin{bmatrix} \gamma \\ 1 - \gamma \end{bmatrix} \text{ and } q^* = \begin{bmatrix} \delta \\ 1 - \delta \end{bmatrix},$$

where

$$\gamma = \frac{b_{22} - b_{21}}{b_{11} - b_{12} - b_{21} + b_{22}} \text{ and } \delta = \frac{a_{22} - a_{12}}{a_{11} - a_{12} - a_{21} + a_{22}}.$$

Assume that player 2 does not have any information about his opponent's attitude towards losses, allowing him to form a wrong belief about it. That is, he may believe player 1 is loss averse, even if he is not. This would also affect player 2's belief about the utility function of the first player, so in order to make him (player 1) indifferent between his two pure strategies, he has to mix differently between his own pure strategies. Specifically, he plays a strategy $\tilde{q} = (\tilde{\delta}, 1 - \tilde{\delta})$ where $\tilde{\delta} \in (0, 1)$. Player 1 knows the utility function of his opponent, so he keeps playing his previous strategy p^*. Thus, if player 2 misperceives λ_1, the mixed loss aversion equilibrium becomes $(p^*, \tilde{q})$. [7]

The equilibrium under player 2's misperception of λ_1 can be explained in two ways. First, player 1 could be naive in the sense that he does not know

---

[7]If player 1's changed degree of loss aversion would change the carrier of his opponent, then it would also change his equilibrium strategy. However, since we are in a 2×2 framework, players keep playing strategies with full carriers for each type of loss aversion equilibrium.

that player 2 does not perceive his degree of loss aversion correctly. Thus, player 1 plays his equilibrium strategy and is surprised by player 2's action. A second explanation would be that player 1 – knowing that player 2 does not have any information about λ_1 – intentionally misrepresents his degree of loss aversion, but is myopic in the sense that he is not able to determine the strategy he has to play in order to optimally exploit player 2's action.

Note that the assumption that player 1 is naive or myopic in the sense described above, also allows for a more general situation where both players are loss averse ex ante, but where player 2 overestimates player 1's degree of loss aversion. However, to avoid cluttered notation, we henceforth stick to the simpler but analogous situation of loss neutral players where player 2 erroneously believes that player 1 is loss averse.

6.2 Main Result

Above, we defined the game and explained what it means that an opponent has a wrong belief about a player's degree of loss aversion. Here, we aim to investigate how an opponent's wrong belief might influence a player's expected payoff. We say that player 2's misperception of λ_1 *benefits* player 1 if $p^* A\tilde{q} \geq p^* Aq^*$, and *hurts* him if $p^* A\tilde{q} \leq p^* Aq^*$.[8] The following theorem presents the comparative statics result.

Theorem 6.1 *In case i. player 1 benefits from player 2 misperceiving λ_1. In cases ii. and iii. player 1 benefits from player 2 misperceiving λ_1 if and only if*

$$\frac{b_{22} - b_{21}}{b_{11} - b_{12} - b_{21} - b_{22}} \geq \frac{a_{22} - a_{21}}{a_{11} - a_{12} - a_{21} - a_{22}}. \tag{1}$$

The proof can be found in the appendix.

Condition (1) can be interpreted as follows. Define $p' := (\gamma', 1 - \gamma')$ with

$$\gamma' := \frac{a_{22} - a_{21}}{a_{11} - a_{21} - a_{12} + a_{22}},$$

and observe that this is exactly the strategy player 1 must play to be indifferent between the actions of his opponent[9]. Furthermore, this strategy

---

[8]Note that player 2's wrong belief about his opponent's loss attitude hurts nor benefits him. That is, $p^* Bq^* = p^* B\tilde{q}$.

[9]Note that p' may not exist, i.e. player 1 can only play p' if $0 \leq \gamma' \leq 1$.

yields the Nash equilibrium payoff $(a_{11}a_{22} - a_{12}a_{21})/(a_{11} - a_{21} - a_{12} + a_{22})$, regardless of the strategy of player 2. Thus, as long as player 2 plays his Nash equilibrium strategy q^*, then player 1 is indifferent between playing p^* and p'. However, if player 2 plays $\tilde{q}$, i.e. erroneously believes player 1 is loss averse, then player 1 gets the Nash equilibrium payoff by playing p', but could get something more or something less by playing p^*.

Given this simple condition on the payoffs of the players, we thus obtain that pretending to be more loss averse makes a player better off. The comparative statics of the fixed point loss aversion equilibrium were investigated before in (Shalev, 2000) with different results. This can be explained as follows. Shalev investigated how a player's reference point moves with his own degree of loss aversion. As this reference point equals this player's equilibrium payoff by definition, this would then yield comparative statics results. However, when a player's degree of loss aversion is changed, also his utility function is changed. Thus, expected utility payoffs are compared, which have been obtained with different utility functions. We investigate how player i's payoff changes when the *belief* player j has about i's degree of loss aversion changes. Since player i's utility function is independent of player j's beliefs about λ_i, this approach circumvents the pitfall of changing utility functions.

7 Concluding Remarks

In this article we have argued that in order to correctly incorporate the concept of loss aversion into non-cooperative game theory, it is necessary to let the reference points of the players depend on the strategies their opponents play. We have examined three different loss aversion equilibrium concepts that satisfied this requirement. Then we established that in 2×2 bimatrix games, a simple condition on the payoffs is sufficient for a player to benefit from his opponent overestimating his (the player's) degree of loss aversion.

A   Proofs for Section 6

It is obvious that a player's reference point is in between the lowest and the highest payoff in his payoff matrix. To simplify the comparative statics analysis of loss aversion equilibria, it is useful to further narrow down this interval for each equilibrium type. In the restrictive 2×2 case, there are three intervals which could contain r_1: the upper, the middle and the lower interval. Let $\pi \in \mathbb{R}^4$ with

$$\pi = \begin{cases} (a_{12}\,,\ a_{22}\,,\ a_{21}\,,\ a_{11}) & \text{in case i.} \\ (a_{12}\,,\ a_{21}\,,\ a_{22}\,,\ a_{11}) & \text{in case ii.} \\ (a_{21}\,,\ a_{12}\,,\ a_{22}\,,\ a_{11}) & \text{in case iii.} \end{cases}$$

The following lemma says that the reference point in fixed point loss aversion equilibrium, lies in the middle interval.

Lemma A.1 *If $(p^*, q^*, (r_1, r_2))$ is a fixed point loss aversion equilibrium in $((A, B), (\lambda_1, \lambda_2))$, then $r_1 \in [\pi_2, \pi_3)$.*

Proof.   Let $\hat{x}$ denote the basic utility payoff x that has been transformed to account for the fact that the decision maker considers it to be a loss. Suppose $r_1 \geq \max\{a_{21}, a_{22}\}$. The expected payoff under the mixed loss aversion equilibrium is equal to a linear combination of $\hat{a}_{21}$ and $\hat{a}_{22}$. Since such a linear combination is strictly smaller than $\max\{\hat{a}_{21}, \hat{a}_{22}\}$, which in turn is smaller than $\max\{a_{21}, a_{22}\}$, this contradicts the assumption $r_1 \geq \max\{a_{21}, a_{22}\}$. Thus, $r_1 < \pi_3$.
To see that the reference point can never be in the lower interval, we must look at each case separately.

Case i.: Here, $a_{21} \geq a_{22}$ and $a_{22} > a_{12}$. Assume $r_1 \in [a_{12}, a_{22})$. The transformed payoff a_{12} is denoted by $\hat{a}_{12}$. Note that $a_{22} - a_{21} \leq 0$ and $\hat{a}_{12} - a_{22} < 0$, and thus $(a_{22} - a_{21})(\hat{a}_{12} - a_{22}) \geq 0$. Working out the lefthand side of the inequality and writing the righthand side as $a_{11}a_{22} - a_{11}a_{22}$ yields

$$a_{22}\hat{a}_{12} - a_{22}^2 - a_{21}\hat{a}_{12} + a_{21}a_{22} \geq a_{11}a_{22} - a_{11}a_{22},$$

from which it follows that

$$a_{11}a_{22} - a_{21}\hat{a}_{12} \geq a_{11}a_{22} - a_{22}\hat{a}_{12} - a_{21}a_{22} + a_{22}^2.$$

Getting a_{22} out of the expression on the righthand side of the inequality yields $a_{22}(a_{11} - \hat{a}_{12} - a_{21} + a_{22})$. The starting assumption of the three cases was that $a_{11} - a_{21} > 0$ and $a_{22} - \hat{a}_{12} > 0$, implying $a_{11} - \hat{a}_{12} - a_{21} + a_{22} > 0$. It follows that

$$\frac{a_{11}a_{22} - a_{21}\hat{a}_{12}}{a_{11} - \hat{a}_{12} - a_{21} + a_{22}} \geq a_{22}.$$

In other words, player 1's expected payoff under loss aversion is larger or equal than a_{22}, contradicting $r_1 < a_{22}$. Thus, $r_1 \geq \pi_2$.

Case ii.: Here, $a_{11} > a_{21}$ and $a_{22} \geq a_{21}$, implying $(a_{11} - a_{21})(a_{22} - a_{21}) \geq 0$. Assume $r_1 \in [a_{12}, a_{21})$, and again, let $\hat{a}_{12}$ represent the transformed a_{12}. Working out the lefthand side of the above inequality, and writing the righthand side as $a_{21}\hat{a}_{12} - a_{21}\hat{a}_{12}$ yields

$$a_{11}a_{22} - a_{21}a_{22} - a_{11}a_{21} + a_{21}^2 \geq a_{21}\hat{a}_{12} - a_{21}\hat{a}_{12}.$$

Rearranging the terms then yields

$$a_{11}a_{22} - a_{21}\hat{a}_{12} \geq a_{11}a_{21} - a_{21}\hat{a}_{12} - a_{21}^2 + a_{21}a_{22}.$$

Getting a_{21} out of the expression on the righthand side results in $a_{21}(a_{11} - \hat{a}_{12} - a_{21} + a_{22})$. Since $a_{11} - \hat{a}_{12} - a_{21} + a_{22} > 0$, we have

$$\frac{a_{11}a_{22} - a_{21}\hat{a}_{12}}{a_{11} - \hat{a}_{12} - a_{21} + a_{22}} \geq a_{21},$$

contradicting $r_1 < a_{21}$. Hence, $r_1 \geq \pi_2$.

Case iii.: Repeating the argument from case ii., with a_{21} replaced by a_{12}, and $\hat{a}_{12}$ by $\hat{a}_{21}$, yields the desired contradiction for case iii.

It follows from the above that $r_1 \in [\pi_2, \pi_3)$. ∎

For the other loss aversion equilibrium concepts, we have a similar result.

Lemma A.2 *If $(p^*, q^*, (r_1, r_2))$ is a maximin loss aversion equilibrium in $((A, B), (\lambda_1, \lambda_2))$, then $r_1 = \pi_2$.*

Proof. Since we have a unique, completely mixed Nash equilibrium, both players play full carriers. Then

Case i.: $\max_{i \in I} \min_{j \in J} a_{ij} = \max\{a_{12}, a_{22}\} = a_{22}$.

Case ii. and iii.: $\max_{i \in I} \min_{j \in J} a_{ij} = \max\{a_{12}, a_{21}\} = a_{21}$.

Thus, $r_1 = \pi_2$. ∎

Lemma A.3 *If $(p^*, q^*, (r_1, r_2))$ is a safety level loss aversion equilibrium in $((A, B), (\lambda_1, \lambda_2))$, then $r_1 \in [\pi_2, \pi_3]$.*

Proof. Recall that for any 2×2 matrix A, we have

$$v_1(A) = \max_{p \in \Delta^2} \min_{q \in \Delta^2} pAq \geq \max_{i \in I} \min_{q \in \Delta^2} e^i Aq = \max_{i \in I} \min_{j \in J} e^i Ae^j.$$

Assume $r_1 < \pi_2$. Then

$$\pi_2 = \max_{i \in I} \min_{j \in J} e^i A^{\lambda_1, r_1} e^j.$$

In safety level loss aversion equilibrium we have $r_1 = v_1(A^{\lambda_1, r_1})$. Then

$$v_1(A^{\lambda_1, r_1}) < \pi_2 = \max_{i \in I} \min_{j \in J} e^i A^{\lambda_1, r_1} e^j,$$

which is a contradiction. Hence, $r_1 \geq \pi_2$.

The safety level, $v_1(A^{\lambda_1, r_1})$, can be interpreted as player 1's Nash equilibrium payoff in the zero-sum game $(A^{\lambda_1, r_1}, -A^{\lambda_1, r_1})$. By a similar reasoning as above, we have that player 2's payoff, $-v_1(A^{\lambda_1, r_1})$, is above $-\pi_3$, implying $v_1(A^{\lambda_1, r_1}) \leq \pi_3$. Hence, $r_1 \in [\pi_2, \pi_3]$. ∎

Recall that $\tilde{q} = (\tilde{\delta}, 1 - \tilde{\delta})$. Next $\tilde{\delta}$ is stated for the three different cases.

Case i.: Here, we have $a_{11} > a_{21} \geq a_{22} > a_{12}$, and by the lemma's A.1 – A.3, $r_1 \in [a_{22}, a_{21}]$. Thus,

$$\begin{aligned}
\tilde{\delta} &= \frac{a_{22} - a_{12} - \lambda_1(r_1 - a_{22}) + \lambda_1(r_1 - a_{12})}{a_{11} - a_{12} - a_{21} + a_{22} + \lambda_1(r_1 - a_{12}) - \lambda_1(r_1 - a_{22})} \\
&= \frac{(1 + \lambda_1)(a_{22} - a_{12})}{a_{11} - a_{21} + (1 + \lambda_1)(a_{22} - a_{12})}.
\end{aligned}$$

Case ii.: Here $a_{11} > a_{22} \geq a_{21} > a_{12}$ with $a_{11} > a_{22}$, $a_{22} > a_{21}$, or both. By the above lemma's, we have $r_1 \in [a_{21}, a_{22}]$. Hence,

$$\tilde{\delta} = \frac{a_{22} - a_{12} + \lambda_1(r_1 - a_{12})}{a_{11} - a_{12} + \lambda_1(r_1 - a_{12}) - a_{21} + \lambda_1(r_1 - a_{21}) + a_{22}}$$

$$= \frac{a_{22} - a_{12} + \lambda_1(r_1 - a_{12})}{a_{11} - a_{12} - a_{21} + a_{22} + \lambda_1(2r_1 - a_{12} - a_{21})}.$$

Case iii.: Here $a_{11} > a_{22} \geq a_{12} > a_{21}$. We thus have $r_1 \in [a_{12}, a_{22}]$, which implies that $\tilde{\delta}$ doesn't change compared to case ii. That is,

$$\tilde{\delta} = \frac{a_{22} - a_{12} + \lambda_1(r_1 - a_{12})}{a_{11} - a_{12} - a_{21} + a_{22} + \lambda_1(2r_1 - a_{12} - a_{21})}.$$

Having specified $\tilde{q}$, player 2's equilibrium strategy given a wrong belief about λ_1, for each case, we can now ask the question how it compares to q^*, player 2's equilibrium strategy given the correct belief about λ_1.

Lemma A.4 *Given the bimatrix game* $((A, B), (\lambda_1, \lambda_2))$, *let* $(p^*, (\delta, 1-\delta))$ *be the unique Nash equilibrium in* (A, B), *and* $(p^*, (\tilde{\delta}, 1-\tilde{\delta}))$ *the Nash equilibrium in* (A^{λ_1, r_1}, B), *where* r_1 *is the equilibrium reference point associated with some loss aversion equilibrium type. Then*

$$\tilde{\delta} \geq \delta$$

in all three cases.

Proof.   Let $x := a_{22} - a_{12}$ and $y := a_{11} - a_{21} - a_{12} + a_{22}$. Note that x and y are strictly positive, and that $\delta = x/y$. Assume $\lambda_1 > 0$. Again, consider the different cases.

Case i.: Here we have

$$\tilde{\delta} = \frac{(1 + \lambda_1)(a_{22} - a_{12})}{a_{11} - a_{21} + (1 + \lambda_1)(a_{22} - a_{12})} = \frac{(1 + \lambda_1)x}{y + \lambda_1 x}.$$

The fact that $a_{11} > a_{21}$ implies $y > x$. Since $\lambda_1 x > 0$, this implies $\lambda_1 xy > \lambda_1 x^2$. Observe that $xy > 0$ and $y(y + \lambda_1 x) > 0$. Hence,

$$\frac{xy + \lambda_1 xy}{y(y + \lambda_1 x)} > \frac{xy + \lambda_1 x^2}{y(y + \lambda_1 x)}.$$

Canceling out y in the lefthand side and $y + \lambda_1 x$ on the righthand side then yields

$$\frac{(1 + \lambda_1)x}{y + \lambda_1 x} > \frac{x}{y},$$

which is equivalent to $\tilde{\delta} > \delta$.

Case ii. and iii.: Here we have

$$\tilde{\delta} = \frac{a_{22} - a_{12} + \lambda_1(r_1 - a_{12})}{a_{11} - a_{12} - a_{21} + a_{22} + \lambda_1(2r_1 - a_{12} - a_{21})} = \frac{x + \lambda_1(r_1 - a_{12})}{y + \lambda_1(2r_1 - a_{12} - a_{21})}.$$

Consider the linear function $f : [a_{21}, a_{22}] \to \mathbb{R}$, given by

$$f(r_1) := \left[\frac{a_{21}}{a_{11} - a_{21}} - \frac{a_{12}}{a_{22} - a_{12}} \right] + \left[\frac{1}{a_{22} - a_{12}} - \frac{1}{a_{11} - a_{21}} \right] r_1.$$

Observe that $f(a_{21}) \geq 0$ and $f(a_{22}) \geq 0$. Since f is a linear function, this implies $f(r_1) \geq 0$ for all $r_1 \in [a_{21}, a_{22}]$. This implies

$$\frac{r_1 - a_{12}}{x} \geq \frac{r_1 - a_{21}}{y - x},$$

which is equivalent to $(r_1 - a_{12})(y - x) \geq (r_1 - a_{21})x$ since both x and $y - x$ are strictly positive. Since $\lambda_1 > 0$, this can be rewritten to $\lambda_1(r_1 - a_{12})y \geq \lambda_1(2r_1 - a_{12} - a_{21})x$. Adding the constant term xy on both sides yields $xy + \lambda_1(r_1 - a_{12})y \geq xy + \lambda_1(2r_1 - a_{12} - a_{21})x$. Since $y > 0$ and $y + \lambda_1(2r_1 - a_{12} - a_{21}) > 0$, it follows that

$$\frac{x + \lambda_1(r_1 - a_{12})}{y + \lambda_1(2r_1 - a_{12} - a_{21})} \geq \frac{x}{y}.$$

Hence, $\tilde{\delta} \geq \delta$.

Thus, $\tilde{\delta} \geq \delta$ in general. ∎

Proof of Theorem 6.1: We have

$$p^* A q^* = \begin{bmatrix} \gamma a_{11} + (1 - \gamma)a_{21} & \gamma a_{12} + (1 - \gamma)a_{22} \end{bmatrix} \begin{bmatrix} \delta \\ 1 - \delta \end{bmatrix},$$

and

$$p^*A\tilde{q} = \begin{bmatrix} \gamma a_{11} + (1-\gamma)a_{21} & \gamma a_{12} + (1-\gamma)a_{22} \end{bmatrix} \begin{bmatrix} \tilde{\delta} \\ 1 - \tilde{\delta} \end{bmatrix}.$$

From Lemma A.4, we have $\tilde{\delta} \geq \delta$. Then $p^*A\tilde{q} \geq p^*Aq^*$ if and only if $\gamma a_{11} + (1-\gamma)a_{21} \geq \gamma a_{12} + (1-\gamma)a_{22}$, which is equivalent to

$$\frac{b_{22} - b_{21}}{b_{11} - b_{12} - b_{21} - b_{22}} \geq \frac{a_{22} - a_{21}}{a_{11} - a_{12} - a_{21} - a_{22}}.$$

Observe that in case i. this condition is trivially satisfied. This concludes the proof. ∎

References

[1] Maurice Allais. Le comportement de l' homme rationnel devant le risque: critique des postulats et axiomes de l' école américaine. *Econometrica*, 21:503–546, 1953.

[2] Caroline Berden and Hans Peters. On the effect of risk aversion in bimatrix games. *Theory and Decision*, 60(4):359–370, 2005.

[3] Jeffrey D. Berejikian. Model building with prospect theory: A cognitive approach to international relations. *Political Psychology*, 23(4):759-786, 2002.

[4] Christopher K. Butler. Prospect theory and coercive bargaining. *Journal of Conflict Resolution*, 51(2):227–250, April 2007.

[5] Colin F. Camerer. Prospect theory in the wild: evidence from the field, in *Choices, Values, and Frames* by Tversky & Kahneman (eds), chapter 16, pp. 288–301. Cambridge University Press, 2002.

[6] Vincent P. Crawford. Equilibrium without independence. *Journal of Economic Theory*, 50:127-154, 1990.

[7] Eddie Dekel, Zvi Safra, and Uzi Segal. Existence and dynamic consistency of Nash equilibrium with non-expected utility preferences. *Journal of Economic Theory*, 55(2):229-246, 1991.

[8] Jürgen Eichberger and David Kelsey. Non-additive beliefs and strategic equilibria. *Games and Economic Behavior*, 30(2):183–215, 1999.

[9] Chaim Fershtman. On the value of incumbency: Managerial reference points and loss aversion. *Journal of Economic Psychology*, 17:245–257, 1996.

[10] John C. Hershey, Howard C. Kunreuther, and Paul J. H. Schoemaker. Sources of bias in assessment procedures for utility functions. *Management Science*, 28(8):936–954, August 1982.

[11] Daniel Kahneman and Amos Tversky. Prospect theory: an analysis of decision under risk. *Econometrica*, 47:263–291, 1979.

[12] Mark J. Machina. Choice under uncertainty: Problems solved and unsolved. *Journal of Economic Perspectives*, 1(1):121–154, 1987.

[13] Jonathan Shalev. Loss aversion equilibrium. *International Journal of Game Theory*, 29(2):269–287, 2000.

[14] Paul J. H. Schoemaker. The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature*, 20(2):529–563, June 1982.

[15] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211:453–458, 1981.

[16] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–1131, 1974.

[17] John von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, December 1928.

[18] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.