Self-Serving Biased Reference Points

Andrea Gallice*

University of Torino and Collegio Carlo Alberto

Abstract

This paper formalizes the notion of self-serving bias within a framework of reference-dependent preferences. We argue that the bias affects agents' expectations in a systematic way and, through this channel, also influences their reference points. We derive some general results both at the individual and the aggregate level and then apply the model to two common situations: a bankruptcy problem and a litigation between two parties. In the first case, we provide a ranking of various allocative rules on the basis of the level of welfare they generate. In the second case, we show that self-serving biased reference points do not affect the incidence of trials but can increase the likelihood of appeals.

Keywords: self-serving bias, reference dependent preferences, expectations, bankruptcy problem, litigation.

JEL classification: D03, K41, D63.

1 Introduction

Self-serving bias (SSB) is a pervasive phenomenon that influences individual behavior in a variety of ways: people tend to overestimate their own merits and abilities, to favorably acquire and interpret information, to give biased judgments about what

*Contact: Department of Economics and Statistics, University of Torino, Corso Unione Sovietica 218bis, 10134, Torino, Italy; Collegio Carlo Alberto, Via Real Collegio 30, 10024, Moncalieri, Italy. Email: andrea.gallice@carloalberto.org. Telephone: +390116705287. Fax: +390116705082.

is fair and what is not, and to inflate their claims and contributions.[1] As such, SSB can have important social and economic implications. For instance, it is considered one of the main causes of costly impasses in bargaining and negotiation (Babcock *et al.*, 1995; Babcock and Loewenstein, 1997; Farmer *et al.*, 2004), and it can be a source of political instability (Passarelli and Tabellini, 2013). Moreover, it has been argued that SSB increases the propensity to strike (Babcock *et al.*, 1996), the incidence of trials (Farmer and Pecorino, 2002), and the intensity of marital conflicts (Schütz, 1999).

While the importance of SSB is widely acknowledged in the literature, a proper formalization of the concept, and an analytical study of its implications, continue to be scarce and case-specific. We aim to introduce a general theoretical framework for modeling and studying the effects of the bias. This framework combines SSB with the notion of reference-dependent preferences (RDPs). RDPs explicitly acknowledge the fact that an agent's evaluation of a given outcome can be influenced by comparing it with a certain reference point. This approach is based on the loss aversion conjecture, introduced in the classic article by Kahneman and Tversky (1979): that people define gains and losses with respect to a reference point and losses loom larger than gains.

In the first part of this paper, we consider a model of RDP à la Koszegi and Rabin (2006), and postulate that SSB affects the way agents set their reference points in a simple but systematic way. We argue that the bias influences agents' expectations and, through this channel, ultimately determines their reference points. In our model, a self-serving biased agent has biased expectations, i.e., expectations that foresee a more favorable outcome than a rational assessment of the situation would warrant. As such, the agent unconsciously sets an inflated reference point. We investigate some implications of the model, both at the individual and the aggregate level. For any possible outcome, a self-serving biased agent is worse off with respect to his rational alter-ego as biased reference points lead to either larger perceived

[1] Research in psychology and sociology provides many convincing examples of the existence of such a bias. For instance, Svenson (1981) reports that the overwhelming majority of subjects (93%) feel they drive better than average, while Ross and Sicoly (1979) show how, for married couples, the sum of the their self-assessed personal contributions to household tasks usually exceeds 100%.

losses or smaller perceived gains. Moreover, we show that one can infer the existence of biased individuals whenever the sum of the reference points of all agents involved in a transaction exceeds the available surplus, regardless of whether the latter is exogenously given or endogenously determined by agents' actions. By recursively applying this result to progressively smaller sets of agents, it is possible to identify a lower bound on the number of biased individuals.

In the second part of the paper, we apply our model to two typical situations where the hypothesis that agents may have self-serving biased reference points seems particularly appropriate: a bankruptcy problem and a litigation between a plaintiff and a defendant. In a bankruptcy problem a principal must allocate a scarce resource among a finite number of claimants. We study how the principal's problem gets modified when claimants have biased reference points and we provide a ranking of a number of standard allocative rules in terms of the welfare they generate. More precisely, we show that when all the claimants are biased in the same proportional way, the so-called constrained equal losses rule dominates the proportional rule, which in turn dominates the constrained equal awards rule. In the case of a litigation between a plaintiff and a defendant, we show instead that the possibility that litigants have self-serving biased reference points does not affect the incidence of trials but may increase the probability of appeals. The losing party in fact suffers twice: not only does he bear the monetary consequences of the verdict, he also experiences a further loss in that the outcome falls short of his expectations. In our framework, even the winning party may actually feel disappointed by the judge's decision and thus regret not having settled outside the court.

2 Reference dependent preferences and self-serving bias

The paper introduces the notion of SSB within a framework of RDPs. As a model of the latter, we follow the fruitful approach introduced by Koszegi and Rabin (2006), which we briefly review in Section 2.1. We then discuss in Section 2.2 the issue of reference point formation. Finally, in Section 2.3, we define the notion of self-serving biased reference points.

2.1 A model of reference-dependent preferences

Koszegi and Rabin (2006) (KR06 in what follows) introduce a formal model that captures the notion of RDPs: an individual's perception of a given outcome is shaped not only by the outcome per se but also by how this outcome compares with some reference point.

More precisely, the KR06 model postulates a utility function $u(x \mid r)$ (to be shortly defined), where $x \in \mathbb{R}^K$ is a K-dimensional outcome and $r \in \mathbb{R}^K$ is the reference point. The model allows for stochastic outcomes (such as lotteries defined over deterministic outcomes) as well as for stochastic reference points. If x is distributed according to F and r is distributed according to G, then the agent's expected utility is given by:

$$U(F|G) = \int \int u(x \mid r) dG(r) dF(x)$$

where the function $u(x \mid r)$ takes the following form:

$$u(x \mid r) = \sum_k m_k(x_k) + \sum_k \mu(m_k(x_k) - m_k(r_k))$$

The strictly increasing function $m_k(\cdot)$ captures the direct effect that the possession or consumption of good x_k has on $u(x \mid r)$. The function $\mu(\cdot)$ is a "universal gain-loss function". Given the reference point r_k, $\mu(\cdot)$ reflects the additional effects that perceived gains or losses have on $u(\cdot)$. More precisely, and in line with the original prospect theory formulation of Kahneman and Tversky (1979), $\mu(\cdot)$ is assumed to satisfy the following properties:

P1: $\mu(z)$ is continuous for all z, strictly increasing and such that $\mu(0) = 0$.

P2: $\mu(z)$ is twice differentiable for $z \neq 0$.

P3: $\mu''(z) > 0$ if $z < 0$ and $\mu''(z) < 0$ if $z > 0$.

P4: if $y > z > 0$ then $\mu(y) + \mu(-y) < \mu(z) + \mu(-z)$.

P5: $\lim_{z\to 0^-} \mu'(z)/\lim_{z\to 0^+} \mu'(z) \equiv \lambda > 1$.

The $\mu(\cdot)$ function thus displays a kink when $z = 0$, i.e., when the actual outcome x_k matches the reference point r_k. Property P3 then indicates that $\mu(\cdot)$ is convex for values of x_k that are below r_k (domain of losses) and concave for values of x_k that are above r_k (domain of gains). The same property also implies that the marginal influence of these perceived gains and losses is decreasing. Property P4 means that, for large absolute values of z, the function $\mu(\cdot)$ is more sensitive to losses than to gains. P5 implies the same result for small values of z: $\mu(\cdot)$ is steeper approaching the reference point from the left (losses) than from the right (gains). Taken together, these last two properties capture the loss aversion phenomenon.

2.2 Reference points as (rational) expectations

A key aspect of any model of RDPs is specifying how an agent comes to define his reference point. Kahneman and Tversky (1979) proposed the so-called status quo formulation, which states that individuals set their reference points in line with what they are used to. An alternative possibility is that agents define their reference points according to what they expect rather than to what they have. Consider, for instance, the situation of a worker who expects a wage increase of \$500 but then actually gets an increase of just \$200; this outcome is likely to feel more like a loss with respect to expectations rather than a gain with respect to the status quo.

Supported by recent empirical evidence (Abeler *et al.*, 2011; Ericson and Fuster, 2011), the general consensus now acknowledges the role of expectations as the main determinant of reference points. However, theoretical models that try to embed such a feature face an additional challenge. In order to get sensible results and falsifiable predictions, these models must analytically define agents' reference points. Therefore, they need to identify the precise nature of an agent's expectations. The natural approach, then, is to rely on the notion of rational expectations. For instance, the KR06 model postulates that the reference point is defined by the rational expectations the agent held in the recent past about the outcome of the problem at issue.

As such, KR06 endogenize the reference point: in their notion of personal equilibrium, an agent expects to implement those actions (and thus reach those outcomes) that he indeed finds optimal to pursue in the specific "state of the world" that will occur. While KR06 considers nonstrategic situations with only one individual, Shalev (2000) investigates the issue of reference points formation when the final outcome is determined by the interaction of different agents. He also defines reference points through rational expectations. In particular, he introduces the concept of loss-aversion equilibrium, i.e., an equilibrium in which an agent's reference point coincides with the actual outcome of a modified game where the original payoff is adjusted to capture the loss that the agent expects to experience.

In this paper, we also define reference points as expectations. However, in our framework agents' expectations (and thus their reference points) will not necessarily be rational, precisely because they may be unconsciously influenced by SSB.

2.3 Self-serving biased reference points

We start by formalizing the concept of rational reference point. We will then use this notion as a benchmark to define and identify the alternative concept of self-serving biased reference point.

Consider a set of agents $N = \{1, ..., n\}$ whose preferences are defined over the possible allocations of a resource of size $S \neq 0$. Let

$$X = \left\{ (x_1, ..., x_n) \mid x_i \cdot S \geq 0 \text{ for all } i \in N \text{ and } \sum_{i=1}^{n} x_i = S \right\}$$

denote the set of feasible and non-wasteful allocations, where x_i is the amount received by agent $i \in N$.[2] The actual allocation is determined by a draw from a probability distribution θ which is defined on X, i.e., $\theta(x_1, ..., x_n)$ is the probability that allocation $(x_1, ..., x_n)$ emerges and $\int_X \theta(x_1, ..., x_n)\, dx = 1$. We define an agent's rational reference point as his expected allocation given θ.

[2] In the definition of X, the feasibility condition $x_i \cdot S \geq 0$ for all $i \in N$ requires that the amount x_i that generic agent i gets must be non-negative whenever S is positive, while on the contrary x_i must be non-positive whenever S is negative.

Definition 1 *A rational agent has a reference point* $r_i^{rat} = \int_X x_i \theta(x_1, ..., x_n)\, dx$.

Definition 1 is very general and subsumes some special cases that we will more carefully investigate in the course of the analysis. For instance, if θ is degenerate, r_i^{rat} coincides with the deterministic allocation that will occur (see Example 1 in Section 2.3). Alternatively, when the probability distribution θ is non-degenerate, this can be exogenously given, in which case r_i^{rat} is given by the expected value of x_i (see the bankruptcy problem discussed in Section 3), or it can rather emerge endogenously if agents strategically interact and can influence the final outcome through their actions (see the litigation problem in Section 4). When this is the case, r_i^{rat} is endogenously determined and coincides with the (stochastic or deterministic) allocation that arises on the equilibrium path.

Having defined the concept of rational reference point, we then argue that SSB affects agents' reference points in a simple but systematic way. In line with Babcock and Loewenstein's (1997, p. 110) definition of SSB as a tendency "to conflate what is fair with what benefits oneself", we claim that, everything else being equal, a self-serving biased agent has a higher reference point with respect to the one that his hypothetical unbiased alter ego would set through rational expectations.

Definition 2 *A self-serving biased agent has a reference point* $r_i^{ssb} > r_i^{rat}$.

Clearly, the definition does not analytically pin down a unique value for a self-serving biased reference point. Indeed, any reference point $r_i > r_i^{rat}$ qualifies as a biased reference point.[3] Obviously, the larger the difference $\Delta r_i = \left(r_i^{ssb} - r_i^{rat}\right)$, the larger the agent's bias. Notice also that, by the same logic, one could define an agent who sets a reference point $r_i < r_i^{rat}$ as an individual who is influenced by a sort of "self-defeating bias". However, the empirical evidence for the existence of such a bias is much weaker and a formal analysis of its consequences lies beyond the scope of this paper. Therefore, in what follows we assume $r_i \geq r_i^{rat}$ for every

[3] For instance, our definition of biased reference point is consistent with the approach that Hart and Moore (2008) pursue in their study about the pros and cons of flexible contracts. The authors assume in fact that parties set as their reference point the *best* outcome that the contract permits, which is then obviously larger than the average outcome.

$i \in N$. In other words, each individual has either a rational or a self-serving biased reference point.

Definition 2 also implies that, in the context of RDPs, SSB negatively affects an agent's utility and this negative effect is increasing in the size of the agent's bias. The bias in fact inflates the reference point and thus leads to either smaller perceived gains or larger perceived losses. Proposition 1 formalizes this result (all proofs appear in the appendix).

Proposition 1 *Let agent i have RDPs such that $u_i(x_i \mid r_i) = m(x_i) + \mu(m(x_i) - m(r_i))$. Then, $u_i(x_i \mid r_i^{ssb}) < u_i(x_i \mid r_i^{rat})$ for any possible x_i. Moreover, the agent's utility is strictly decreasing in $\Delta r_i = \left(r_i^{ssb} - r_i^{rat}\right)$.*

The following example shows how SSB can lead an agent to have an inflated reference point even in a deterministic situation. It also highlights the negative consequences the bias has on an individual's well-being.

Example 1 *A principal must allocate a monetary bonus of size $S = 1$ to two workers on the basis of the effort the two exerted in performing a certain task. Let $u_i(x_i \mid r_i) = m(x_i) + \mu(m(x_i) - m(r_i))$ be the utility function of worker $i \in \{a, b\}$, where $x_i \in [0, 1]$ is the bonus the worker receives. The principal perfectly observes agents' effort $e_a, e_b \geq 0$ and implements the allocation $x^* = \{x_a^*, x_b^*\}$ where $x_i^* = f(e_i)$ with $f(e_i) : \mathbb{R}^+ \to [0, 1]$ and $f'(e_i) > 0$. The function $f(e_i)$ is common knowledge among all agents. Worker a is rational. He correctly anticipates what he will get and sets his reference point accordingly: $r_a^{rat} = x_a^*$. He thus enjoys utility $u_a(x_a^* \mid r_a^{rat}) = m(x_a^*)$ as $\mu(0) = 0$ (property P1 of the μ function). Worker b is self-serving biased. He overestimates the effort he exerted (i.e., he considers $\tilde{e}_b > e_b$) and thus expects to receive an amount $\tilde{x}_b > x_b^*$ given that $f(\tilde{e}_b) > f(e_b)$.[4] Worker b thus unconsciously sets his reference point such that $r_b^{ssb} = \tilde{x}_b$ with $r_b^{ssb} > r_b^{rat}$. He enjoys utility $u_b(x_b^* \mid r_b^{ssb}) = m(x_b^*) + \mu(m(x_b^*) - m(r_b^{ssb}))$. In line with Proposition*

[4] In other words, and consistently with how we defined SSB in the first few lines of the introduction, the agent overestimates his own merits (i.e., $\tilde{e}_b > e_b$) and inflates his claims and contributions (i.e., $\tilde{x}_b > x_b^*$).

1, $u_b(x_b^* \mid r_b^{ssb}) < u_b(x_b^* \mid r_b^{rat})$ *given that* $m(x_b^*) - m(r_b^{ssb}) < 0$*, and the worker experiences a loss.*

Example 1 also illustrates a more general result: whenever all agents are rational then their reference points are mutually compatible, i.e., their sum matches the size of the resource to be shared. This simple consideration leads to the statement of the following proposition:

Proposition 2 *Consider a profile of reference points* $r = (r_1, ..., r_n)$ *and let* $S \neq 0$ *be the size of the available surplus. Then, if* $\sum_i r_i = S$*, all agents have rational reference points. If instead* $\sum_i r_i > S$*, at least some of the agents have self-serving biased reference points.*

We now propose a procedure that, by recursively applying Proposition 2 to progressively smaller sets of agents, is able to identify a lower bound on the number of biased individuals.

Proposition 3 *Given the profile of reference points* $r = (r_1, ..., r_n)$ *such that* $\sum_i r_i > S$ *and where, without loss of generality,* $r_1 \leq ... \leq r_n$*, the number of self-serving biased agents is at least* $n-m+1$*, where* m *is such that* $\sum_{i=1}^{m} r_i > S$ *and* $\sum_{i=1}^{m-1} r_i \leq S$*.*

Example 2 *Consider two hypothetical situations with* $n = 4$ *and* $S = 1$*. In the first situation, let* $r = (0.2, 0.3, 0.3, 0.5)$ *such that* $\sum_{i=1}^{4} r_i = 1.3$*. Given that* $\sum_{i=1}^{3} r_i < 1$*, it follows that* $m = 4$ *and* $n - m + 1 = 1$*. Therefore, one can only conclude that there is at least one biased agent. In the alternative scenario, let* $r = (0.4, 0.7, 0.8, 0.8)$ *such that* $\sum_{i=1}^{4} r_i = 2.7$*. Given that* $\sum_{i=1}^{2} r_i > 1$ *and* $\sum_{i=1}^{1} r_i < 1$*, it follows that* $m = 2$ *and* $n - m + 1 = 3$*. Therefore, at least three agents have a biased reference point.*

In what follows we apply our proposed framework of self-serving biased reference points to two common situations: a bankruptcy problem (Section 3) and a litigation between two parties (Section 4). Within these contexts, the model generates further implications and lead to some novel predictions.

3 Application to a bankruptcy problem

In a bankruptcy problem a principal must allocate a finite resource of size S among a number of agents whose claims sum up to more than S. A typical example is a bankrupt firm that must be liquidated. The proposed framework of self-serving biased reference points seems particularly appropriate in the context of bankruptcy problems. These are, in fact, typical situations in which the two conditions that underlie the model are likely to hold. First, a claimant's utility is likely to be affected not only by the actual allocation but also by how this compares with his expectations (i.e., his reference point). Second, a claimant's expectations are likely to be affected by SSB such that the agent may have inflated reference points. For instance, the creditor of a bankrupt firm may think that his claims deserve a higher priority compared to those of other claimants and may thus expect a larger reimbursement.

We model the problem as follows: a principal must allocate a homogeneous and perfectly divisible good whose size we normalize to $S = 1$ among $n \geq 2$ claimants. Let $c = (c_1, ..., c_n)$ with $c_i \in \mathbb{R}_+$ and such that $\sum_i c_i > 1$ be a vector that collects individual claims. The vector $x = (x_1, ..., x_n)$ with $\sum_i x_i = 1$ denotes instead a possible allocation. Claimants have RDPs, i.e., $u_i(x_i \mid r_i) = m(x_i) + \mu(m(x_i) - m(r_i))$ where, as before, r_i is the agent's reference point. In what follows we actually set $m(x_i) = x_i$ such that $u_i(x_i \mid r_i) = x_i + \mu(x_i - r_i)$.[5]

We are then interested in studying some welfare properties of three standard allocative rules that are commonly advocated in the literature (see Thomson, 2003 and 2013 for exhaustive reviews). These rules are:

- The *proportional* rule (*prop*), which allocates amounts proportional to claims:

$$x_i^{prop} = \lambda^{prop} c_i \text{ with } \sum_i \lambda^{prop} c_i = 1$$

[5] This assumption simplifies the analysis as the linear form of $m(\cdot)$ implies that the utility function $u_i(x_i \mid r_i)$ satisfies the same properties that characterize the $\mu(\cdot)$ function (see Proposition 2 in KR06). However, to measure a claimant's standard utility in terms of the amount of resource that he obtains is consistent with the approach that is usually adopted in bankruptcy problems (see Thomson, 2013).

- The *constrained equal awards* rule (*cea*), which assigns equal amounts to all claimants subject to no one receiving more than his claim:

$$x_i^{cea} = \min\{c_i, \lambda^{cea}\} \text{ with } \sum_i \min\{c_i, \lambda^{cea}\} = 1$$

- The *constrained equal losses* rule (*cel*), which assigns equal amount of losses to all claimants subject to no one receiving a negative amount:

$$x_i^{cel} = \max\left\{0, c_i - \lambda^{cel}\right\} \text{ with } \sum_i \max\left\{0, c_i - \lambda^{cel}\right\} = 1$$

All three rules select allocations that satisfy some basic desirable properties (Thomson, 2013), such as *non-negativity* (no agent is asked to pay: $x_i^\tau \geq 0$ for any $i \in N$ and any $\tau \in \{prop, cea, cel\}$), *claims boundedness* (no agent receives more than his claim: $x_i^\tau \leq c_i$ for any i and any τ), and *balance* (the principal allocates all the resource: $\sum_i x_i^\tau = 1$ for any τ).[6]

As a measure of welfare, we use the utilitarian social welfare function whose generic form is given by $W_{ut}(x) = \sum_i u_i(x_i)$. Therefore, and given the balance property, $W_{ut}(x^\tau) = 1 + \sum_i \mu(x_i^\tau - r_i)$ for any $\tau \in \{prop, cea, cel\}$.

3.1 The case with rational claimants

Rational claimants have rational reference points. Let $\theta(x^\tau) \in [0,1]$ denote the probability that the principal will implement allocation x^τ with $\tau \in \{prop, cea, cel\}$ (or, equivalently, claimants' beliefs that the principal is of type τ). In line with Definition 1, a claimant's rational reference point is then given by $r_i^{rat} = \sum_\tau x_i^\tau \theta(x^\tau)$ with $\sum_i r_i^{rat} = 1$.[7] Given that the rational reference point of agent i is a weighted average of all the possible realizations of x_i and since $x_i^\tau \leq c_i$ for any τ (claim boundedness property), we can conclude that $r_i^{rat} \leq c_i$ for any i. In other words, a

[6] As an example of how the three rules work in practice, consider a bankruptcy problem with $n = 3$ and let $c = (0.3, 0.5, 0.8)$. The rules would then select the following allocations: $x^{prop} = (0.1875, 0.3125, 0.5)$, $x^{cea} = (0.3, 0.35, 0.35)$, and $x^{cel} = (0.1, 0.3, 0.6)$.

[7] Following on the example that was introduced in the previous footnote, assume $\theta(x^\tau) = \frac{1}{3}$ for any $\tau \in \{prop, cea, cel\}$. The vector of rational reference points is then given by $r^{rat} = (0.196, 0.321, 0.483)$. Clearly, $\sum_i r_i^{rat} = 1$.

rational claimant realizes that he will possibly get less than his claim. As such, he sets his expectations, and thus his reference point, accordingly.

When all the agents are rational two other interesting relations hold. First, the actual gains or losses that an agent will perceive conditional on the specific allocation that the principal will implement cancel out across rules. More formally, $\sum_{\tau}\left(x_i^{\tau}-r_i^{rat}\right)=0$ for any $i\in N$. This implies that in expectations an agent does not experience any gain or loss. Second, within any rule, individual gains and losses also cancel out across agents, i.e., $\sum_i\left(x_i^{\tau}-r_i^{rat}\right)=0$ for any τ. However, this last condition does not necessarily imply the condition $\sum_i\mu(x_i^{\tau}-r_i^{rat})=0$ given that the μ function weights gains and losses differently. It follows, that in general $W_{ut}(x^{\tau})\neq 1$.[8] Since losses loom larger than gains, one may actually expect $W_{ut}(x^{\tau})<1$ most of the times.[9]

3.2 The case with self-serving biased claimants

We now study the bankruptcy problem when some of the claimants have a self-serving biased reference point. The first result that we show is quite straightforward: the bias is welfare detrimental not only at the individual level (see Proposition 1) but also at the aggregate one.

Proposition 4 $W_{ut}(x^{\tau}\mid r')<W_{ut}(x^{\tau}\mid r^{rat})$ *for any rule* $\tau\in\{prop,cea,cel\}$ *and any vector* $r'\geq r^{rat}$ *with* $r'\neq r^{rat}$.

It is then interesting to study how the different rules perform in terms of the actual level of welfare that they generate. We tackle this issue in a restricted domain

[8]Indeed, $W_{ut}(x^{\tau})=1$ only in some specific situations such as when $\theta(x^{\tau})=1$ for some $\tau\in\{prop,cea,cel\}$ (i.e., no uncertainty about the principal's type) or when $c_i=c_j$ for all $i,j\in N$ (i.e., claimants are symmetric such that $x^{\tau}=\left(\frac{1}{n},...,\frac{1}{n}\right)$ for any τ and thus $r^{rat}=x^{\tau}$).

[9]For instance, it is always the case that $W_{ut}(x^{\tau})<1$ when $n=2$ and $c_i\neq c_j$ as the perceived loss of agent i weights more than the perceived gain of agent j. On the contrary, the condition $W_{ut}(x^{\tau})>1$ verifies only in specific situations. A necessary condition is that the number of agents that experience a gain is larger than the number of those that experience a loss. The intuition is that, due to diminishing sensitivity of the μ function (Property P3), the aggregate benefit experienced by a large number of claimants that receive little more than their reference point may overcome the negative effect experienced by a small number of agents that suffer large losses.

where it is possible to define a clear ranking. In particular, we focus on the case in which claimants know with certainty the principal's type, i.e., $\theta(x^\tau) = 1$ for some $\tau \in \{prop, cea, cel\}$. This implies that, from the point of view of the claimants, the outcome of the allocation procedure is deterministic such that $r_i^{rat} = x_i^\tau$. We then further assume that all claimants are biased in the same proportional way. More formally, $r_i^{ssb} = (1+\beta) r_i^{rat}$ for all $i \in N$ where the parameter $\beta > 0$ captures the proportion by which an agent's reference point is inflated with respect to the rational benchmark.

Within such a frawework, utilitarian welfare can be expressed as $W_{ut}(x^\tau) = 1 + \sum_i \mu(x_i^\tau - r_i^{ssb}) = 1 + \sum_i \mu(-\beta x_i^\tau)$. The distribution of perceived losses across agents (i.e., the distribution of $-\beta x_i^\tau$) is thus proportional to the actual distribution that rule τ implements (i.e., the distribution of x_i^τ). Because of the diminishing sensitivity to losses displayed by the μ function, such a relationship leads to a clear ranking of the three allocative rules. The first step toward this goal is to evaluate how evenly the different rules allocate the contested resource across claimants. As a measure of inequality we use the standard Lorenz dominance criterion (see Sen, 1973), which we briefly review.

Consider two allocations $x^\tau = (x_1^\tau, ..., x_n^\tau)$ and $x^\kappa = (x_1^\kappa, ..., x_n^\kappa)$ with $\tau, \kappa \in \{prop, cea, cel\}$ and then define the new vectors $\hat{x}^\tau = (\hat{x}_1^\tau, ..., \hat{x}_n^\tau)$ and $\hat{x}^\kappa = (\hat{x}_1^\kappa, ..., \hat{x}_n^\kappa)$ which display the coordinates of the original vectors in increasing order, i.e., $\hat{x}_i^\tau \leq \hat{x}_{i+1}^\tau$ and $\hat{x}_i^\kappa \leq \hat{x}_{i+1}^\kappa$ for any $i \in \{1, ..., n-1\}$. We say that x^τ Lorenz dominates x^κ, and we denote such a relation by writing $x^\tau \succ_L x^\kappa$, if:

$$\hat{x}_1^\tau \geq \hat{x}_1^\kappa, \quad \hat{x}_1^\tau + \hat{x}_2^\tau \geq \hat{x}_1^\kappa + \hat{x}_2^\kappa, \quad ... \quad, \text{ and } \hat{x}_1^\tau + ... + \hat{x}_{n-1}^\tau \geq \hat{x}_1^\kappa + ... + \hat{x}_{n-1}^\kappa$$

with at least one strict inequality. We instead write $x^\tau \succeq_L x^\kappa$ if none of the above mentioned inequalities holds strictly. We can then say that rule τ Lorenz dominates rule κ, a relation that we denote as $\tau \succ_L \kappa$, whenever $x^\tau \succeq_L x^\kappa$ for any possible vector of claims and $x^\tau \succ_L x^\kappa$ for at least one vector of claims. In other words, the relation $\tau \succ_L \kappa$ indicates that rule τ always implements an allocation that is (weakly) less skewed with respect to the allocation that rule κ implements. Bosmans and Lauwers (2011) show that the constrained equal awards rule Lorenz dominates the proportional rule, which in turn Lorenz dominates the constrained equal losses rule. More formally, $cea \succ_L prop \succ_L cel$. Such a result plays a key role in the proof of the following proposition:

Proposition 5 *When all claimants have a self-serving biased reference point* $r_i^{ssb} = (1+\beta)\, r_i^{rat}$ *with* $\beta > 0$ *and* $\theta\,(x^\tau) = 1$ *for some* $\tau \in \{prop, cea, cel\}$*, the following ranking emerges:* $W_{ut}(x^{cel}) \geq W_{ut}(x^{prop}) \geq W_{ut}(x^{cea})$.

This result highlights a peculiar characteristic of bankrupcty problems when claimants have RDPs and biased reference points. From a purely utilitarian point of view, the allocations that generate the highest level of welfare are those that implement the most uneven distributions. Because of the diminishing sensitivity of the gain-loss function $\mu\,(\cdot)$ (Property P3), it is in fact more efficient to disappoint a lot just a few claimants rather than disappoint a little all of them.

4 Application to a litigation problem

As a second application of our model of self-serving biased reference points, we consider a litigation between a plaintiff and a defendant. Litigations are another typical situation in which agents' perception of the final outcome is likely to be affected by how this compares to the expected outcome, which is in turn likely to be influenced by SSB. We investigate whether and how the possibility that litigants have biased reference points modifies their decision to proceed to a costly trial versus a settlement out of court, as well as the probability of appealing against the judge's verdict. We first consider a dispute in which litigants may be self-serving biased but still display standard utility (i.e., their well-being is solely determined by the monetary consequences generated by the final outcome). We then consider the case in which agents have RDPs.

4.1 The case with self-serving biased litigants

We follow the standard structure introduced in Shavell (1982) and model a litigation as a game between two risk-neutral players: a plaintiff (p) and a defendant (d). The game develops as follows:

- p moves first and decides whether to sue d or not (in which case the game ends).

- If p brings suit, litigants can negotiate an agreement and thus settle out of court.

- If litigants fail to settle, they go to trial and the judge decides the outcome.

If the litigation enters the courtroom, let $q \in [0, 1]$ represent the probability that the judge decides in favor of the plaintiff and W_i with $i \in \{p, d\}$ be agent i's assessment of the actual reimbursement W the defendant will be ordered to pay the plaintiff. We allow for the possibility that SSB influences litigants' views and thus assume $W_p \geq W \geq W_d$. More precisely, $W_p > W$ if the plaintiff is biased (he overestimates the damage he suffered) and $W > W_d$ if the defendant is biased (he underestimates his responsibility in causing the accident). The condition $W_i = W$ defines instead an unbiased agent.[10] We indicate with $c_i > 0$ agent i's legal costs such that $C = c_p + c_d$ represents the total costs of going to trial. We assume each side bears its own costs no matter the outcome of the trial (the so-called American rule).

Litigants' expected monetary outcomes in case of a trial are thus given by:

$$x_{p|D} = qW_p - c_p \quad , \quad x_{d|D} = -qW_d - c_d$$

where we use the notation $x_{i|D}$ to indicate the payoff that litigant $i \in \{p, d\}$ obtains in the "disagreement outcome" (D) as agents fail to settle. Settlement can actually

[10] Notice that this is a slightly different approach with respect to how Shavell (1982) and Bar-Gill (2005) model "optimism". In these papers, agents do not know the true q and optimism influences q_i, which is the perceived probability of a judgement in favor of the plaintiff. In our model, agents do not know W and SSB affects W_i, which is the perceived reimbursement the defendant must pay if the plaintiff wins the trial. While the analytical implications of the two approaches are similar (what matters are expected payoffs), we think that our characterization better matches the definition of SSB as a bias that inflates "how much an agent thinks he deserves" (and similarly their approach better describes optimism, i.e., a bias that leads agents to overestimate the probability of winning).

occur whenever there exists an expected surplus that agents can share, i.e., whenever the amount of money the defendant expects to pay is larger than the net reimbursement the plaintiff expects to get. If this is the case, we assume that the settlement outcome is determined through Nash bargaining such that litigants equally share the surplus of size $S = \left|x_{d|D}\right| - x_{p|D} = q(W_d - W_p) + C$. We can thus express agent i's expected payoff in the "agreement outcome" (A) that emerge when litigants settle out of court as $x_{i|A} = x_{i|D} + \frac{S}{2}$.[11] More explicitly:

$$x_{p|A} = q\left(\frac{W_p + W_d}{2}\right) + \frac{c_d - c_p}{2} \quad , \quad x_{d|A} = -q\left(\frac{W_p + W_d}{2}\right) + \frac{c_p - c_d}{2}$$

Notice that whenever both agents are unbiased, the plaintiff always sues the defendant (we assume that $qW - c_p > 0$) and the two agents always agree on the settlement. In fact, whenever $W_p = W = W_d$, the agreement payoffs simplify to $x_{p|A} = qW + \frac{c_d - c_p}{2}$ and $x_{d|A} = -qW + \frac{c_p - c_d}{2}$ such that the condition $x_{i|A} \geq x_{i|D}$ for any $i \in \{p, d\}$ holds.

Therefore, a necessary condition for a litigation to proceed to trial is that at least one agent is self-serving biased. More precisely, trial occurs if and only if $x_{i|D} > x_{i|A}$ for at least one $i \in \{p, d\}$, i.e., if and only if the condition $qW_p - c_p > qW_d + c_d$ holds.[12] This condition allows to determine the maximum level of legal costs that makes litigant i prefer to go to trial rather than to settle: the agent prefers to enter the courtroom as far as $c_i < q(W_p - W_d) - c_j$ with $i, j \in \{p, d\}$ and $j \neq i$.[13]

[11] More precisely, and given the surplus $S = q(W_d - W_p) + C$, Nash bargaining selects the allocation $s^* = (s_p^*, s_d^*)$, with $s_p^* + s_d^* = S$, that maximizes the so-called Nash product (NP), i.e., the product of the differences between agents' agreement and disagreement payoffs. Formally, s^* maximizes the function $NP = \left[x_{p|A} - x_{p|D}\right] * \left[x_{d|A} - x_{d|D}\right]$, where $x_{i|A} = x_{i|D} + s_i$ for both $i \in \{p, d\}$. Therefore, $NP = [s_p * s_d] = [s_p * (S - s_p)]$ given that $s_p + s_d = S$. First and second order conditions are given by $\frac{\partial NP}{\partial s_p} = S - 2s_p = 0$ and $\frac{\partial^2 NP}{\partial s_p^2} = -2$ such that $s_p^* = s_d^* = \frac{S}{2}$.

[12] As an example, consider a litigation problem with $W_p = 14$, $W = 10$, $W_d = 8$, $c_p = 4$, $c_d = 2$, and $q = 0.5$. The condition $qW_p - c_p > qW_d + c_d$ therefore does not hold such that the litigation will not reach the courtroom. To see this notice that $x_{p|D} = 3$, $x_{d|D} = -6$, and $S = \left|x_{d|D}\right| - x_{p|D} = 3$ with $\frac{S}{2} = 1.5$. Therefore, $x_{p|A} = 4.5$, $x_{d|A} = -4.5$, and litigants agree to settle given that $x_{i|A} \geq x_{i|D}$ for both $i \in \{p, d\}$.

[13] This formulation confirms the standard result, in which "*under the American system, there will be a trial if and only if the plaintiff's estimate of the expected judgment exceeds the defendant's*

4.2 The case with self-serving biased reference points

Assume litigants now display RDPs. More precisely, let their utility function be given by $u_{i|\eta} = x_{i|\eta} + \mu(x_{i|\eta} - r_{i|\eta})$, where $x_{i|\eta}$ is the expected monetary amount that agent $i \in \{p, d\}$ receives/pays in outcome $\eta \in \{D, A\}$, as identified in the previous section. Agent i's reference point is given by $r_{i|\eta}$ and, coherently with the approach that we pursued throughout the paper, we postulate that agents' reference points are determined by their expectations. However, the peculiarity here is that a litigant's reference point forms endogenously: agents' espectations in fact do differ across outcomes and which outcome $\eta \in \{D, A\}$ will actually occur is in turn determined by agents' actions. We thus have to distinguish between two different reference points depending on the outcome that verifies along the equilibrium path:

- If the litigation enters the courtroom ($\eta = D$, the disagreement outcome), litigants' reference point is given by the expected value of the trial. Therefore:

$$r_{p|D} = x_{p|D} = qW_p - c_p \quad , \quad r_{d|D} = x_{d|D} = -qW_d - c_d$$

It is immediate to verify that a rational plaintiff ($W_p = W$) has a rational reference point ($r^{rat}_{p|D} = qW - c_p$) while a self-serving biased plaintiff ($W_p > W$) has an inflated reference point ($r^{ssb}_{p|D} = qW_p - c_p$ such that $r^{ssb}_{p|D} > r^{rat}_{p|D}$). Similar considerations hold for the defendant: $r^{rat}_{d|D} = -qW - c_d$ and $r^{ssb}_{d|D} = -qW_d - c_d$ with $r^{ssb}_{d|D} > r^{rat}_{d|D}$ given that $W_d < W$.[14]

- If agents settle ouside the court ($\eta = A$, the agreement outcome), litigants' reference point is instead given by the settlement amount:

estimate by at least the sum of their legal costs" (Shavell, 1982, page 63). In fact, trial occurs if and only if $c_i < \hat{c}_i$, i.e., $c_i < qW_p - qW_d - c_j$ for some $i \in \{p, d\}$ and $j \neq i$. It follows that trial occurs if and only if $qW_p - qW_d > c_i + c_j$.

[14]Notice also that, in line with our general model (see Proposition 2), $\sum_i r_{i|D} = -C$ whenever $r_{i|D} = r^{rat}_{i|D}$ for both $i \in \{p, d\}$. The trial generates in fact a negative surplus of size $S = -C$ since the reimbursement that the defendant possibly pays to the plaintiff is a simple monetary transfer between the two agents while legal costs are dissipated. The condition $\sum_i r_{i|D} > -C$ characterizes instead all those situations in which at least one litigant has a self-serving biased reference point.

$$r_{p|A} = x_{p|A} = q\left(\frac{W_p + W_d}{2}\right) + \frac{c_d - c_p}{2} \quad , \quad r_{d|A} = x_{d|A} = -q\left(\frac{W_p + W_d}{2}\right) + \frac{c_p - c_d}{2}$$

where $r_{i|A}$ is a now a deterministic value since litigants reach the settlement outcome through negotiations and thus know with certainty what they will get.[15]

Despite of the fact that litigants may have different reference points depending on the outcome that will emerge in equilibrium, the following proposition shows what may at first seem a surprising result: the possibility that litigants have self-serving biased reference points does not affect the incidence of trials.

Proposition 6 *The fact that litigants may have RDPs does not influence their decision to proceed to trial, regardless of whether their reference points are rational or self-serving biased.*

The intuition for such a result is the following. Litigants decide whether to settle or proceed to court on the basis of the expected utility these two actions lead to. However, agents "expect to get what they expect", no matter if they are rational or biased. This implies that they do not anticipate any perceived gain or loss. As such, and because of property P1 of the μ function, RDPs collapse to standard preferences that are solely defined on the monetary consequences of the litigation. Litigants' incentives thus coincide with those that characterize a situation in which agents have standard preferences and so do their equilibrium actions. Therefore, the incidence of trials remains unaffected.

However, the fact that agents may have RDPs and self-serving biased reference points modifies litigants' perception of the actual outcome of the trial once this comes to an end. For example, if the judge decides in favor of the plaintiff (we indicate this specific realization of the disagreement outcome by writing $\eta = D_p$), the actual utility $u_{i|D_p}$ that litigants experience is given by:

[15] Notice that in this case $\sum_i r_{i|A} = 0$ regardless of whether agents have rational or biased reference points. A settlement is a purely monetary transfer between litigants and it thus generates a null surplus. As such, Proposition 2 does not apply.

$$u_{p|D_p} = W - c_p + \mu(W - qW_p) \quad , \quad u_{d|D_p} = -W - c_d + \mu(-W + qW_d)$$

If on the contrary the judge decides in favor of the defendant ($\eta = D_d$), the litigants' actual utility $u_{i|D_d}$ is given by:

$$u_{p|D_d} = -c_p + \mu(-qW_p) \quad , \quad u_{d|D_d} = -c_d + \mu(qW_d)$$

In both cases the losing party enjoys lower utility with respect to the case in which he had standard preferences (i.e., utility functions as above but without the $\mu(\cdot)$ part). In fact, in addition to the negative monetary consequences he has to suffer, a losing agent with RDPs and biased reference point perceives a further loss as the actual payoff falls short of the expected one. Assuming that the likelihood with which a losing party appeals against the judge's decision is an increasing function of the agent's degree of disappointment, we can thus state the following result:

Proposition 7 *The likelihood with which a losing party appeals against the verdict of a trial is increasing in the size of his self-serving bias.*

Finally, notice that while a victorious defendant is certainly happy about the final outcome ($\mu(qW_d) > 0$, because the agent feels as if he saved an amount qW_d), a victorious plaintiff can still feel disappointed by the judge's verdict. In fact, whenever $qW_p > W$, the plaintiff experiences a loss since the reimbursement he gets from the defendant is lower than the expected one. While it is obvious that a losing party would ex-post have preferred to agree on a settlement outside the court, this last consideration indicates that even a winning plaintiff might ex-post regret his decision to proceed to trial rather than to settle.

5 Conclusion

We presented a general framework that can be used to explicitly model the self-serving bias and study some of its consequences from an analytical point of view. In particular, we introduced the bias within the family of reference dependent preferences by arguing that the bias systematically inflates agents' expectations, and thus their reference points. We then applied the model to two common situations where the existence of self-serving biased reference points is likely to play an important role: a bankruptcy problem and a litigation between two parties. In the bankruptcy case, we provided a ranking of a number of standard allocative rules on the basis of the welfare that they generate. In the litigation case, our model essentially confirmed the detrimental role that self-serving bias has on the probability of solving a dispute outside the court. However, we showed that the possibility that litigants may have self-serving biased reference points increases the probability of appeals.

Despite some obvious limitations, we feel that the proposed formulation provides a simple but fruitful way to formally analyze some of the consequences of the self-serving bias, captures the main ingredients of many real-life problems and, generally contributes to the recent literature regarding the public policy implications of research in behavioral economics.

Acknowledgements

I thank Pascal Courty, Botond Koszegi and Klaus Schmidt for helpful suggestions. I also acknowledge useful comments received at seminars at the universities of Florence, Munich and Turin as well as at the following conferences: Annual Congress of the European Economic Association (Milan), European Symposium on Economics and Psychology (Mannheim), LabSi Conference on Political Economy and Public Choice (Siena), SIDE 8^{th} Annual Conference (Rome).

Appendix

Proof of Proposition 1

The $m(\cdot)$ function is strictly increasing. The $\mu(\cdot)$ function is also strictly increasing. Therefore, $\mu(\cdot)$ is strictly decreasing in r_i. By Definition 2, $r_i^{ssb} > r_i^{rat}$. It follows that, for any given x_i, $\mu(m(x_i) - m(r_i^{ssb})) < \mu(m(x_i) - m(r_i^{rat}))$ which implies $u_i(x_i \mid r_i^{ssb}) < u_i(x_i \mid r_i^{rat})$. More in general, the fact that $\mu(\cdot)$ is strictly decreasing in r_i implies that $\frac{\partial u_i(x_i|r_i^{ssb})}{\partial \Delta r_i} < 0$ for any x_i.

Proof of Proposition 2

By Definition 1, $r_i^{rat} = \int_X x_i \theta\,(x_1, ..., x_n)\, dx$. Therefore,

$$\begin{aligned}\sum\nolimits_i r_i^{rat} = \sum\nolimits_i \left(\int_X x_i \theta\,(x_1, ..., x_n)\, dx\right) = \\ = \int_X \left(\sum\nolimits_i x_i\right) \theta\,(x_1, ..., x_n)\, dx = \\ = \int_X S \cdot \theta\,(x_1, ..., x_n)\, dx = S\end{aligned}$$

By Definition 2, a self-serving biased agent has instead a reference point $r_i^{ssb} > r_i^{rat}$. It follows that $\sum_i r_i > S$ whenever there is at least one agent $i \in N$ such that $r_i = r_i^{ssb}$.

Proof of Proposition 3

Consider the set $N = \{1, ..., n\}$ and let $r = (r_1, ..., r_n)$ be such that $r_1 \leq r_2 \leq ... \leq r_n$. If $\sum_{i=1}^{n} r_i > S$, then, by Proposition 2, at least one player is biased. Assume that agent n is the unique biased agent. Moreover, assume his bias is extreme, i.e., $r_n^{rat} = 0$ such that $\Delta r_n = r_n$. Now consider the set $N \setminus \{n\} = \{1, ..., n-1\}$. If $\sum_{i=1}^{n-1} r_i > S$, then, again by Proposition 2, there must be at least one other biased agent. Remove agent $n-1$ and apply the same procedure. The process is iterated until one reaches the set $N \setminus \{m, ..., n\} = \{1, ..., m-1\}$ with $\sum_{i=1}^{m} r_i > S$ and $\sum_{i=1}^{m-1} r_i \leq S$. This is the largest possible set that is consistent with the hypothesis of unbiased agents. It follows that $n - m + 1$ is the minimum number of self-serving biased agents within the original set N.

Proof of Proposition 4

Let $W_{ut}\left(x^{\tau} \mid r^{rat}\right) = 1 + \sum_i \mu(x_i^{\tau} - r_i^{rat})$ be the level of welfare generated by rule $\tau \in \{prop, cea, cel\}$ when all the claimants have rational reference points. Now consider the vector $r' \geq r^{rat}$ such that there exists at least one claimant whose reference point is biased, i.e., $r_i' = r_i^{ssb} > r_i^{rat}$ for some $i \in N$. Let $W_{ut}(x^{\tau} \mid r') = 1 + \sum_i \mu(x_i^{\tau} - r_i')$ be the associated level of welfare. Notice that $\sum_i \left(x_i^{\tau} - r_i^{rat}\right) = 0$ while $\sum_i (x_i^{\tau} - r_i') < 0$ given that $\sum_i r_i' > 1$ (see Proposition 2). Therefore, it must be the case that $(x_i^{\tau} - r_i') < \left(x_i^{\tau} - r_i^{rat}\right)$ for some $i \in N$, which in turn implies $\mu\left(x_i^{\tau} - r_i'\right) < \mu\left(x_i^{\tau} - r_i^{rat}\right)$ for some $i \in N$ given that $\mu(\cdot)$ is strictly increasing (Property P1). It follows that $W_{ut}(x^{\tau} \mid r') < W_{ut}(x^{\tau} \mid r^{rat})$.

Proof of Proposition 5

Given that $r_i^{ssb} = (1+\beta)\, r_i^{rat}$ for all $i \in N$ and $\theta\left(x^{\tau}\right) = 1$ for some $\tau \in \{prop, cea, cel\}$, utilitarian welfare is given by $W_{ut}\left(x^{\tau}\right) = 1 + \sum_i \mu(x_i^{\tau} - r_i^{ssb}) = 1 + \sum_i \mu(-\beta x_i^{\tau})$ where $\mu(-\beta x_i^{\tau}) < 0$ such that $\sum_i \mu(-\beta x_i^{\tau}) < 0$. $W_{ut}\left(x^{\tau}\right)$ is strictly increasing in $\sum_i \mu(-\beta x_i^{\tau})$. Because of the strict convexity of the μ function in the domain of losses (Property P3), we have that $\mu(a)+\mu(b) < \mu(a-\epsilon)+\mu(b+\epsilon) < 0$ for any $a \leq b < 0$ and $\epsilon \in (0, b)$. Therefore, the rule that achieves the highest $\sum_i \mu(-\beta x_i^{\tau})$ is the rule that implements the more skewed distribution of $-\beta x_i^{\tau}$. Given that $cea \succ_L prop \succ_L cel$ (Bosmans and Lauwers, 2011) and $\beta > 0$, it follows that $-\beta x^{cea} \succ_L -\beta x^{prop} \succ_L -\beta x^{cel}$. Therefore, $\sum_i \mu(-\beta x_i^{cea}) \leq \sum_i \mu(-\beta x_i^{prop}) \leq \sum_i \mu(-\beta x_i^{cel}) < 0$ such that $W_{ut}(x^{cel}) \geq W_{ut}(x^{prop}) \geq W_{ut}(x^{cea})$.

Proof of Proposition 6

Let litigant $i \in \{p, d\}$ have RDPs such that, for any possible outcome $\eta \in \{D, A\}$ of the litigation, $u_{i|\eta} = x_{i|\eta} + \mu(x_{i|\eta} - r_{i|\eta})$. However, $r_{i|\eta} = x_{i|\eta}$ because the litigant expects to get $x_{i|\eta}$ and thus sets his reference point accordingly. Therefore, $u_{i|\eta} = x_{i|\eta} + \mu(0) = x_{i|\eta}$ for any $i \in \{p, d\}$ and any $\eta \in \{D, A\}$ as $\mu(0) = 0$ by property P1 of the $\mu(\cdot)$ function. The problem is therefore analogous to the one studied in Section 4.1 and thus leads to the same results. In particular, the litigation proceeds to trial if and only if the condition $qW_p - c_p > qW_d + c_d$ holds or, equivalently, if and only if litigants' legal costs are such that $c_i < q(W_p - W_d) - c_j$ with $i, j \in \{p, d\}$ and $j \neq i$.

Proof of Proposition 7

Denote with $z_i = \mu(x_{i|D_j} - r_{i|D})$ the perceived gain ($z_i > 0$) or loss ($z_i < 0$) that agent $i \in \{p, d\}$ experiences when he compares the *actual* outcome of the trial ($x_{i|D_j}$ where D_j with $j \in \{p, d\}$ indicates the outcome in which the judge decides in favor of litigant j) with his ex-ante reference point ($r_{i|D}$). Let $\pi_i(z_i) \in [0, 1]$ define the probability that litigant i will appeal against the judge's verdict and assume that $\pi_i(z_i) = 0$ if $z_i > 0$ while $\frac{\partial \pi_i(\cdot)}{\partial z_i} < 0$ if $z_i < 0$. Finally, and as already introduced, let $\Delta r_i = \left(r_i^{ssb} - r_i^{rat}\right)$ be the size of litigant i's self-serving bias. A losing plaintiff perceives a loss $z_p = \mu(-qW_p) < 0$. The size of his bias is given by $\Delta r_p = q(W_p - W)$. Therefore, $\frac{\partial z_p}{\partial \Delta r_p} < 0$ since as the bias increases (W_p goes up), a losing plaintiff experiences a larger loss. Similarly, a losing defendant perceives a loss $z_d = \mu(-W + qW_d) < 0$. The size of his bias is given by $\Delta r_d = q(W - W_d)$. Therefore, $\frac{\partial z_d}{\partial \Delta r_d} < 0$ since as the bias increases (W_d goes down), a losing defendant experiences a larger loss. Therefore, $\frac{\partial z_i}{\partial \Delta r_i} < 0$ for both $i \in \{p, d\}$. Given that by assumption $\frac{\partial \pi_i(\cdot)}{\partial z_i} < 0$ for $z_i < 0$, it follows that $\frac{\partial \pi_i(\cdot)}{\partial \Delta r_i} > 0$.

References

[1] Abeler, J., Falk, A., Goette, L., Huffman, D., 2011. Reference points and effort provision. *American Economic Review*, 101 (2), 470-492.

[2] Babcock, L., Loewenstein, G., 1997. Explaining bargaining impasse: the role of self-serving biases. *Journal of Economic Perspectives* 11 (1), 109-126.

[3] Babcock, L., Loewenstein, G., Issacharoff, S., Camerer, C., 1995. Biased judgments of fairness in bargaining. *American Economic Review* 85 (5), 1337-1343.

[4] Babcock, L., Wang, X., Loewenstein, G., 1996. Choosing the wrong pond: social comparisons in negotiations that reflect a self-serving bias. *Quarterly Journal of Economics* 111 (1), 1-19.

[5] Bar-Gill, O., 2005. The evolution and persistence of optimism in litigation. *Journal of Law, Economics and Organization*, 22 (2), pp. 490-507.

[6] Bebchuk, L.A., 1984. Litigation and settlement under imperfect information. *RAND Journal of Economics*, 15 (3), 404-415.

[7] Bosmans, K., Lauwers, L., 2011. Lorenz comparisons of nine rules for the adjudication of conflicting claims. *International Journal of Game Theory*, 40 (4), pp. 791-807.

[8] Ericson, K. Fuster, A., 2011. Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. *Quarterly Journal of Economics*, 126 (4), 1879-1907.

[9] Farmer, A., Pecorino, P., 2002. Pretrial bargaining with self-serving bias and asymmetric information. *Journal of Economic Behavior and Organization*, 48 (2), 163-176.

[10] Farmer, A., Pecorino, P., Stango, V., 2004. The causes of bargaining failure: evidence from major league baseball. *Journal of Law and Economics*, 47 (2), 543-568.

[11] Hart, O., Moore, J., 2008. Contracts as reference points. *Quarterly Journal of Economics* 123 (1), 1-48.

[12] Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47 (2), 263-291.

[13] Koszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. *Quarterly Journal of Economics* 121 (4), 1133-1165.

[14] Passarelli, F., Tabellini, G., 2013. Emotions and political unrest. IGIER Working Papers, 474, Bocconi University.

[15] Ross, M., Sicoly, F., 1981. Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology* 37 (3), 322-336.

[16] Schütz, A., 1999. It was your fault! Self-serving biases in autobiographical accounts of conflicts in married couples. *Journal of Social and Personal Relationships* 16 (2), 193-208.

[17] Sen, A., 1973. On economic inequality. Clarendon Press, Oxford.

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

[19] Shavell, S., 1982. Suit, Settlement, and Trial: A Theoretical Analysis under Alternative Methods for the Allocation of Legal Costs. *Journal of Legal Studies* 11 (1), 55-81.

[20] Svenson, O., 1981. Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica* 47 (2), 143-148.

[21] Thomson, W., 2003. Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: a survey. *Mathematical Social Sciences* 45 (3), 249-297.

[22] Thomson, W., 2013. Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: an update. RCER working papers, N. 578, University of Rochester.

[23] Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5 (4), 297-323.