Collegio Carlo Alberto

Self-serving biased reference points

Andrea Gallice

No. 223
November 2011

Carlo Alberto Notebooks

www.carloalberto.org/working_papers

Self-serving biased reference points

Andrea Gallice*

University of Torino and Collegio Carlo Alberto

Abstract

This paper formalizes the pervasive phenomenon of the self-serving bias within the framework of reference dependent preferences. This formulation allows the stating of a simple rule to assess the existence of the bias at the aggregate level as well as a procedure that identifies the minimum number of biased agents. We apply the model to two standard situations: a litigation between a plaintiff and a defendant and a bankruptcy problem. In the litigation case, we show how the combination of self-serving bias and reference dependent preferences increases the likelihood that a dispute proceeds to trial. In the bankruptcy case, we show how the existence of individuals with self-serving biased reference points exacerbates the conflict between equity and efficiency of the final allocation.

Keywords: self-serving bias, reference dependent preferences, litigation, trial vs. settlement, bankruptcy problem, optimal allocation.

JEL classification: D03, K41, D63.

*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: European Symposium on Economics and Psychology (Mannheim), LabSi Conference on Political Economy and Public Choice (Siena), 23$^{\text{rd}}$ Annual Congress of the European Economic Association (Milan). The usual disclaimer applies. Contact: andrea.gallice@carloalberto.org

1 Introduction

Self-serving bias 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 is fair and what is not, and to inflate their claims and contributions.[1] As such, self-serving bias (from now on SSB) can have important social and economic implications. For instance, it is considered as one of the main causes of costly impasses in bargaining and negotiations (see Babcock *et al.*, 1995 and Babcock and Loewenstein, 1997) as well as a source of political instability (Heyndels and Ashworth, 2003). 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). Even if the importance of SSB is widely acknowledged in the economic literature, a proper formalization of the concept, as well as the analytical study of its implications, are still somehow scarce and case-specific. In this paper we aim to introduce a more general theoretical framework for modeling and studying the effects of the bias. This framework combines SSB with the notion of reference dependent preferences.

Reference dependent preferences (from now on RDP) 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 intuition goes back to the loss aversion conjecture introduced in the classical article by Kahneman and Tversky (1979) and more recently modeled by Koszegi and Rabin (2006): people define gains and losses with respect to a reference point and losses loom larger than gains.

We postulate that SSB affects agents' reference points in a trivial but systematic

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

way. We claim in fact that, everything else being equal, a self-serving biased agent will have the tendency to set a reference point that is higher then the one his unbiased counterpart would set. This consideration leads to a simple rule for assessing the existence of SSB at the aggregate level: whenever agents' reference points are not mutually compatible (i.e., their sum exceeds the surplus available in the transaction) then one can conclude that at least some of the players are self-serving biased. By recursively applying this rule to progressively smaller sets of agents, we are also able to put a lower bound on the number of biased individuals.

We investigate the implications of the proposed framework in two common situations where SSB is (very) likely to play a role: a litigation between two parties and a bankruptcy problem.[2] In the first case, we model the choices of a plaintiff and a defendant who can either settle their dispute out of court or proceed to a costly trial. This is a major topic as a noticeable percentage of litigations reach the courtroom with obvious costs for the contendants as well as for the efficiency and speed of the legal system.[3] Shavell (1982) and Bebchuk (1984) have been the first to show how wrong beliefs (and in particular optimistic beliefs) about the likelihood of prevailing at trial reduce the space for settlements and make trials more likely. More recently, Farmer and Pecorino (2002) have explicitly analyzed the role of the self-serving bias in litigations. In their formulation, SSB takes the form of a multiplicative bias that inflates (for the plaintiff) or deflates (for the defendant) the objective probability that the judge will decide in favor of the plaintiff. Despite some peculiar situations, the authors find that usually a larger SSB increases the incidence of trials. Bar-Gill (2005) studies instead the evolutionary properties of optimistic beliefs (but argues that the analysis also applies to other cognitive biases such as the SSB) and shows

[2] As such, the paper also contributes to the growing literature about behavioral welfare economics (see Bernheim and Rangel, 2007, for a recent review) that studies the welfare/policy implications of behavioral models vis-à-vis traditional models.

[3] In general, around 10% of legal disputes are litigated in a trial (Bar-Gill, 2005). However, in the case of medical malpractice lawsuits this figure climbs up to 39% (Studdert *et al.*, 2006).

that this bias is evolutionary stable as it makes a party more succesfull in extracting more favorable settlements. Building on these papers, we show how the combination of SSB and RDP confirms the detrimental role that the bias has in the chances of reaching a settlement out of the court but, perhaps surprisingly, does not further amplify it.

The second application that we investigate is a classical bankruptcy problem (see Thomson, 2003, for a recent review about the vast literature on the topic), i.e., the problem of optimally allocating a scarce resource among a finite number of claimants. We show that whenever at least some of the claimants display SSB then the allocation that matches each agent's reference point is unfeasible. The planner is thus forced to disappoint at least some of the claimants and this raises a number of interesting questions. Shall the planner disappoint (a little) all the claimants? Or shall he match the expectations of a few while disappointing (a lot) the remaining ones? If so, who shall the planner favor? We investigate these issues under different social welfare specifications and show how the combination of SSB and RDP exacerbates the trade-off between equity and efficiency of the final allocation.

The paper is organized as follows. Section 2 introduces a model of self-serving biased reference points and presents some general results. Section 3 applies the proposed model to a litigation problem. Section 4 investigates a bankruptcy problem. Section 5 concludes.

2 Reference dependent preferences and self-serving bias

Koszegi and Rabin (2006) introduced the following analytical formulation of reference dependent preferences:

$$u(x, r) = m(x) + \mu(m(x) - m(r))$$

The increasing function $m(\cdot)$ captures the direct effect that the possession or consumption of good x has on total utility $u(\cdot)$. The function $\mu(\cdot)$ is a "universal gain-loss function". Given the reference point r, $\mu(\cdot)$ 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: $\lim_{z \to 0^-} \mu'(z) / \lim_{z \to 0^+} \mu'(z) \equiv \lambda > 1$.

Therefore the function $\mu(\cdot)$ is convex for values of x that are below r (domain of losses) and concave for values of x that are above r (domain of gains). Property P3 also implies that the marginal influence of these perceived gains and losses is decreasing.[4] 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) rather than from the right (gains). Taken together, these last two properties capture the loss aversion phenomenon.

On the other hand, the five properties are silent about how an individual sets his reference point r. This is clearly a problematic issue to tackle given the subjective nature of such a choice. Different individuals can set different reference points according to what they have (as in the traditional status quo formulation of Kahneman

[4] In the two applications that we will later consider (Section 3 and Section 4) we will actually work with linear functions that do not capture diminishing sensitivity but still provide an adequate, and much more tractable, characterization of RDP. In those contexts property P3 will thus be substituted by property P3$'$: For all z, $\mu''(z) = 0$. This alternative property is also mutuated from Koszegi and Rabin (2006, see property A3$'$, page 1140).

and Tversky, 1979), to what they expect (as proposed in Koszegi and Rabin, 2006) or to what they think they deserve, just to name a few possibilities.

No matter the specific features of this introspective process, we argue that the self-serving bias affects in a systematic way the reference point the agent set. Babcock and Loewenstein (1997, p. 110) define SSB as a tendency "to conflate what is fair with what benefits oneself". In line with this definition, we claim that, everything else being equal, a biased agent will set a higher reference point with respect to his hypothetical unbiased counterpart, i.e., $r_{(biased)} > r_{(unbiased)}$. This simple consideration implies that SSB has a negative effect on individual utility. In fact, a biased reference point leads to either smaller perceived gains or larger perceived losses. The following lemma clarifies this point.

Lemma 1 *For any given* x*,* $u(x, r_{(biased)}) < u(x, r_{(unbiased)})$*.*

Proof. Property P1 and the fact that $m(\cdot)$ is an increasing function imply that $\mu(\cdot)$ is decreasing in r. Given the assumption $r_{(biased)} > r_{(unbiased)}$, this implies that, for any given x, $\mu(m(x) - m(r_{(biased)})) < \mu(m(x) - m(r_{(unbiased)}))$. Therefore, $u(x, r_{(biased)}) < u(x, r_{(unbiased)})$. ■

Now consider all those situations in which $n \geq 2$ agents have preferences that can be captured by RDP and that are defined on all the possible allocations of a given surplus of size $S \in \mathbb{R}$. The cases in which the surplus available in the problem under scrutiny is negative (perhaps because of the existence of some kind of transaction costs) are then captured by setting $S < 0$. A classical example is a costly trial between a plaintiff and a defendant in a suit for damages. The case with $S = 0$ mimics a zero-sum game of pure transfers among the agents. The case with $S > 0$ captures instead all the situations in which a positive surplus must be shared among different claimants. Examples include bankruptcy and bargaining problems, principal-agent relations and lobbying.

Let the utility function of agent $i \in \{1, ..., n\}$ be $u_i(x_i, r_i) = m(x_i) + \mu(m(x_i) - m(r_i))$.[5] Now imagine that agents have reference points that are not self-serving biased. Almost tautologically, unbiased reference points should be mutually compatible. This means that the sum of these reference points should be equal to the available surplus, i.e., $\sum_i r_{i(unbiased)} = S$. The framework of unbiased reference points provides a benchmark that can be used to assess the existence of agents that are self-serving biased.

Definition 1 *If $\sum_i r_i > S$ then at least some of the agents are self-serving biased.*

Notice that Definition 1 allows the identification of the existence of SSB only at the aggregate level. For instance, in the case with $S > 0$, we did not define SSB at the individual level with the condition $r_i > \frac{S}{n}$. In fact, it could well be the case that an agent sets $r_i > \frac{S}{n}$ without being biased but simply because he objectively deserves more than others. However, if claims are not compatible (i.e., if $\sum_i r_i > S$) then SSB surely inflates the reference point of some of the players.[6] By recursively applying Definition 1 to progressively smaller sets of agents, it is possible to set a lower bound on the number of biased individuals.

Proposition 1 *Given $n \geq 2$ agents and their reference points r_i where, without loss of generality, $r_1 \leq r_2 \leq ... \leq r_n$, then the number of self-serving biased agents is at least $n - k + 1$ where k is such that $\sum_{i=1}^{k} r_i > S$ and $\sum_{i=1}^{k-1} r_i \leq S$.*

Proof. Consider the set $N = \{1, ..., n\}$ and, without loss of generality, let $r_1 \leq r_2 \leq ... \leq r_n$. If $\sum_{i=1}^{n} r_i > S$, then, by Definition 1, at least one player is biased. Imagine that agent n is the only biased agent. Moreover, imagine that his bias is extreme,

[5]Notice that this specification allows for heterogeneity in agents' reference points but assumes that the functions $m(\cdot)$ and $\mu(\cdot)$ are the same across individuals.

[6]We do not consider the situation of $\sum_i r_i < S$ as this would imply that some agents display a self-defeating bias, an hypothesis whose empirical support is much weaker.

i.e., $r_{n(unbiased)} = 0$. Now consider the set $N \backslash \{n\} = \{1, ..., n-1\}$. If $\sum_{i=1}^{n-1} r_i > S$ then, again by Definition 1, there must be at least another biased agent. Remove agent $n-1$ and apply the same procedure. The process is iterated until one reaches the set $N \backslash \{k, ..., n\} = \{1, ..., k-1\}$ with $\sum_{i=1}^{k} r_i > S$ and $\sum_{i=1}^{k-1} r_i \leq S$. This is the largest possible set that is consistent with the hypothesis of unbiased agents. It follows that $n - k + 1$ is the minimum number of self-serving biased agents within the original set N. ■

Example 1 *Consider two hypothetical situations with* $n = 4$ *and* $S = 1$. *In the first one, let* $r_1 = 0.2$, $r_2 = 0.3$, $r_3 = 0.3$ *and* $r_4 = 0.5$ *such that* $\sum_{i=1}^{4} r_i = 1.3$. *Given that* $\sum_{i=1}^{3} r_i < 1$, *we have that* $k = 4$ *and* $n - k + 1 = 1$. *Therefore, we can only conclude that there is at least one biased claimant. In the alternative scenario, let* $r_1 = 0.4$, $r_2 = 0.7$, $r_3 = 0.8$ *and* $r_4 = 0.9$ *such that* $\sum_{i=1}^{4} r_i = 2.8$. *Given that* $\sum_{i=1}^{2} r_i > 1$ *and* $\sum_{i=1}^{1} r_i < 1$, *we have that* $k = 2$ *and* $n - k + 1 = 3$. *Therefore, there are at least three biased agents.*

3 An application to a litigation problem

In this section we apply the framework of RDP and SSB to a standard litigation problem. We want to investigate which are the implications of the proposed model for what concerns agents' decision to proceed to a costly trial versus a settlement out of court. In particular, we compare the results of the model (subsection 3.3) with two benchmark situations: the case in which litigants have perfect information and rational preferences (subsection 3.1) and the case in which litigants have self-serving biased beliefs but, other than that, still standard rational preferences (subsection 3.2).

The general structure of our analysis follows the one introduced in Shavell (1982) and more recently used by Bar-Gill (2005). As such, we model a litigation as a game between two risk-neutral players: a plaintiff (p) and a defendant (d). The plaintiff moves first and must decide if to sue the defendant. In case of suit, then the two players can either agree on a certain settlement or proceed to court. Let c_p, $c_d > 0$ be the costs that the two players must bear if the case goes to trial and $C = c_p + c_d$ the sum of these legal costs,[7] $q \in (0, 1)$ the true probability of a judgement in favor of the plaintiff, and $W > 0$ the true reimbursement the defendant must pay the plaintiff in case the latter wins the trial. We assume that $qW - c_p > 0$ such that the expected value of the trial for the plaintiff.

3.1 The case with perfect information

If both players know q and W with certainty, the plaintiff always sues the defendant and then the two agents always agree on a settlement where the defendant pays the plaintiff an amount qW. Agents' expected payoffs are in fact given by $E(u_p/suit, settlement) = qW$ and $E(u_d/settlement) = -qW$ that strictly dominate the expected payoffs that would follow a trial, namely $E(u_p/suit, trial) = qW - c_p$ and $E(u_d/trial) = -qW - c_d$. It follows that if information was perfect trials should never be observed.

3.2 The case with self-serving bias

A necessary condition for a litigation to proceed to trial is thus that at least one of the two parties does not possess perfect information and displays some sort of bias in his beliefs about the unknown parameters of the model. In this section, we assume that litigants do not know the true W and that their subjective ex-ante assessment W_i with $i \in \{p, d\}$ can be influenced by the self-serving bias.[8] More

[7] We analyze the situation under the so-called American rule according to which each contendant bears his own legal costs no matter the result of the trial.

[8] Notice that this is a slightly different approach with respect to how Shavell (1982) and Bar-Gill (2005) model optimistic beliefs (see also Langlais, 2008). In fact in these papers agents do not know the true q and optimism influences q_i, the perceived probability of a judgement in favor of the plaintiff. In our model agents do not know W and SSB affects W_i, the perceived reimbursement

precisely, $W_p = W_d = W$ if both agents are unbiased. On the other hand, $W_p > W$ if the plaintiff is biased while $W > W_d$ if the defendant is biased. Expected payoffs are thus given by:

Expected utility of the plaintiff p: $\begin{cases} E(u_p/suit, trial) = qW_p - c_p \\ E(u_p/suit, settlement) = \frac{qW_p - c_p + qW_d + c_d}{2} \end{cases}$

Expected utility of the defendant d: $\begin{cases} E(u_d/trial) = -qW_d - c_d \\ E(u_d/settlement) = -\frac{qW_p - c_p + qW_d + c_d}{2} \end{cases}$

Where, in line with Bar-Gill (2005), we have assumed that in a settlement the defendant pays the plaintiff an amount $\frac{qW_p - c_p + |-qW_d - c_d|}{2}$. This amount corresponds to the mean of the two reservation prices (i.e., the expected payoffs in case of trial).[9] The analysis of the payoffs confirms a well-known result (Shavell, 1982, Bar-Gill, 2005): the litigation proceeds to trial if and only if the condition

$$qW_p - c_p > qW_d + c_d \tag{1}$$

holds. On the other hand, settlement thus occurs whenever $qW_p - c_p \leq qW_d + c_d$.

Condition 1 also allows to determine the maximum level of legal costs c_i with $i \in \{p, d\}$ for which agent i still prefers to proceed to trial. More precisely, agent i proceeds to trial whenever $c_i < \hat{c}_i$ where $\hat{c}_i = qW_p - qW_d - c_j$ with $j \in \{p, d\}$ and $j \neq i$.[10] We will compare these thresholds with those that emerge in a situation where SSB is modeled within RDP.

the defendant must pay if the plaintiff wins the trial. While the analytical implications of the two approaches are basically the same (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 their probability to win).

[9] In other words, we imagine a situation in which the two players announce their reservation prices and then bargain with equal bargaining power such that they settle on the mean value.

[10] This formulation confirms the standard result according to 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 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 $qW_p - qW_d > c_i + c_j$.

3.3 The case with self-serving biased reference points

Now let the two litigants display reference dependent preferences à la Koszegi and Rabin (2006), i.e., $u_i(x_i, r_i) = x_i + \mu(x_i - r_i)$ where x_i is the monetary transfer that agent $i \in \{p, d\}$ receives/pays and r_i is his ex-ante reference point.[11] In such a context reference points are naturally given by players' expectations about the outcome of the trial. More precisely, $r_p = qW_p - c_p > 0$ and $r_d = -qW_d - c_d < 0$ where, as in the previous section, $W_p \geq W \geq W_d$ with strict inequalities if agents are self-serving biased.

Notice moreover that the surplus available in the transaction is negative. In fact, while the amount of money that the defendant pays to the plaintiff is a simple monetary transfer between the two agents, the legal costs that agents incur in case of a trial are dissipated. Therefore, using the notation introduced in Section 2, we have $S < 0$ and in particular $S = -C$ where $C = c_p + c_d$. In line with Definition 1, we thus observe that $r_p + r_d = S$ if both agents are unbiased while $r_p + r_d > S$ whenever at least one of the two agents displays a self-serving biased reference point. We can also express the proposed settlement identified in Subsection 3.2 in terms of the reference points given that $\frac{qW_p - c_p + qW_d + c_d}{2} = \frac{r_p - r_d}{2}$. The condition that leads to trial ($qW_p - c_p > qW_d + c_d$, see condition (1) in the previous subsection) can be reformulated as $r_p > |r_d|$. It follows that settlement occurs if and only if $r_p \leq |r_d|$.

Agents' expected payoffs are given by:[12]

Expected utility of the plaintiff p: $$\begin{cases} E(u_p/suit, trial) = r_p \\ E(u_p/suit, settlement) = \frac{r_p - r_d}{2} + \mu\left(\frac{-r_p - r_d}{2}\right) \end{cases}$$

[11] Notice that with respect to the more general formulation introduced in Section 2 (i.e., $u_i(x_i, r_i) = m(x_i) + \mu(m(x_i) - m(r_i))$), here we are assuming that the function $m(\cdot)$ is such that $m(x) = x$ for any $x \in \mathbb{R}$. This assumption is made for the sake of tractability but it does not undermine the general results.

[12] More precisely $E(u_p/trial) = r_p + \mu(r_p - r_p) = r_p$ because $\mu(0) = 0$ given property P1 of the function μ. And $E(u_p/settlement) = \frac{r_p - r_d}{2} + \mu\left(\frac{r_p - r_d}{2} - r_p\right)$ which becomes $E(u_p/settlement) = \frac{r_p - r_d}{2} + \mu\left(\frac{-r_p - r_d}{2}\right)$. Same simplifications apply to the defendant's payoffs.

Expected utility of the defendant d: $\begin{cases} E(u_d/trial) = r_d \\ E(u_d/settlement) = -\frac{r_p - r_d}{2} + \mu\left(\frac{-r_p - r_d}{2}\right) \end{cases}$

The game with unbiased agents (i.e., $W_p = W_d = W$ and $r_p + r_d = S$) simplifies to the situation analyzed in Section 3.1 such that litigants always agree on the settlement qW.

On the other hand, whenever at least one of the agents is biased, the two reference points are no more compatible ($r_p + r_d > S$) and thus, depending on the size of the legal costs, the option to proceed to trial can be characterized by a higher expected payoff. The following proposition shows a surprising result: while a model of self-serving biased reference points clearly increases the incidence of trials with respect to the case in which agents are unbiased and have perfect information (Subsection 3.1), it does not further increase the likelihood of trials in comparison with the case in which only SSB, but not RDP, is present (Subsection 3.2).

Proposition 2 *If agents display reference dependent preferences and at least one of them has a self-serving biased reference point, then a litigation proceeds to trial whenever legal costs of at least one player are below the threshold* $\tilde{c}_i = \hat{c}_i - \mu\left(\frac{-r_p - r_d}{2}\right)$ *with* $i \in \{p, d\}$ *and where* $\hat{c}_i = qW_p - qW_d - c_j$*, with* $j \neq i$*, is the threshold if agents had no RDP.*

Proof. The condition that makes the plaintiff prefer the trial with respect to the settlement is given by $E(u_p/suit, trial) > E(u_p/suit, settlement)$, i.e., $r_p > \frac{r_p - r_d}{2} + \mu\left(\frac{-r_p - r_d}{2}\right)$. This inequality holds for any $c_p < \tilde{c}_p$ where $\tilde{c}_p = qW_p - qW_d - c_d - \mu\left(\frac{-r_p - r_d}{2}\right)$. Similarly, the defendant prefers the trial if and only if $E(u_d/trial) > E(u_d/settlement)$, i.e., $r_d > -\frac{r_p - r_d}{2} + \mu\left(\frac{-r_p - r_d}{2}\right)$ which holds for any $c_d < \tilde{c}_d$ where $\tilde{c}_d = qW_p - qW_d - c_p - \mu\left(\frac{-r_p - r_d}{2}\right)$. Therefore, for any $i \in \{p, d\}$, $\tilde{c}_i = \hat{c}_i - \mu\left(\frac{-r_p - r_d}{2}\right)$ where $\hat{c}_i$ is the threshold if agents had no RDP. Focusing on the last term, and given property P1 of the $\mu(\cdot)$ function, we have that $\left[-\mu\left(\frac{-r_p - r_d}{2}\right)\right] > 0$ whenever $r_p > |r_d|$

while $\left[-\mu\left(\frac{-r_p-r_d}{2}\right)\right] \leq 0$ whenever $r_p \leq |r_d|$. As such, $\tilde{c}_i > \hat{c}_i$ if $r_p > |r_d|$ and $\tilde{c}_i \leq \hat{c}_i$ if $r_p \leq |r_d|$. ■

Notice therefore that the combination of RDP and SSB does not modify the decision of the litigants if to proceed to trial or settle. In fact, if $r_p > |r_d|$ then $c_i < \hat{c}_i < \tilde{c}_i$ and agent i prefers to go in front of the judge both with (this subsection) and without (Subsection 3.2) RDP. At the opposite, if $r_p \leq |r_d|$ then $c_i \geq \hat{c}_i \geq \tilde{c}_i$ and agent i prefers to settle both with and without RDP. The intuition for this result is that litigants with RDP compare the proposed settlement with their reference point (the expected payoff of the trial). The additional disutility (respectively utility) that agents get from this comparison makes less binding the critical level of legal costs below (resp. above) which the trial (resp. settlement) is preferred. In other words, if legal costs of agent i are below the threshold without RDP ($\hat{c}_i$) then *a fortiori* they are below the threshold with RDP ($\tilde{c}_i$). See the example below. While, at the opposite, if c_i is above $\hat{c}_i$ then *a fortiori* c_i is above $\tilde{c}_i$.

Example 2 *Let $q = 0.5$, $W = 100$, $W_p = 120$, $W_d = 90$, $c_p = 8$ and $c_d = 3$. And if litigants have RDP assume also that $m(x) = x$ and that the function $\mu(\cdot)$ is linear with $\mu(x) = x$ for $x < 0$ and $\mu(x) = \frac{1}{2}x$ if $x \geq 0$. Then $r_p = 52$ and $r_d = -42$ such that the proposed settlement is given by $\frac{52-(-42)}{2} = 47$. Thresholds for legal costs are given by $\hat{c}_p = 12$ and $\hat{c}_d = 7$ if litigants had no RDP, and by $\tilde{c}_p = \hat{c}_p - (-5) = 17$ and $\tilde{c}_d = \hat{c}_d - (-5) = 12$ if litigants had RDP. Given that $c_p < \hat{c}_p < \tilde{c}_p$ and $c_d < \hat{c}_d < \tilde{c}_d$ then both the plaintiff and the defendant choose to procede to trial no matter if they have RDP or not.*

4 An application to a bankruptcy problem

Consider the problem of a social planner who must allocate a homogeneous and perfectly divisible good (whose amount we normalize to $S = 1$) among $n \geq 2$

claimants.[13] Let $N = \{1, ..., n\}$ denote the set of claimants. The notation $x = (x_1, ..., x_n)$ indicates a possible allocation such that x_i is the amount of the good that the planner assigns to claimant $i \in N$. Feasible allocations are the ones for which $x_i \in [0, 1]$ for any i and $\sum_i x_i \leq 1$. The vector $u = (u_1, ..., u_n)$ with $u_i = u_i(x_i)$ collects individual utilities.

The social planner wants to maximize social welfare. His objective function is given by a social welfare function (SWF) $W(u)$ that aggregates individuals' utilities into social utilities. We assume that the social planner is not biased towards any particular claimant and therefore we only consider symmetric SWFs that give equal weight to every agent. More precisely, we consider two classical welfare functions: the utilitarian SWF (Bentham, 1789) defined as $W_{ut}(u) = \sum_i u_i$ and the maxmin SWF (Rawls, 1971) defined as $W_{mm}(u) = \min\{u_1, ..., u_n\}$. We will indicate an optimal allocation with the vector $\hat{x}_w = (\hat{x}_1^w, ..., \hat{x}_n^w)$ where $\hat{x}_w = \arg\max W_w(u)$ and $w \in \{ut, mm\}$.

4.1 The case with rational preferences

Traditional neoclassical analysis postulates agents have preferences that lead to continuous, increasing, and concave utility functions. If claimants are endowed with preferences of this kind, the utilitarian SWF selects $\hat{x}_{ut}$ such that $u_i'(\hat{x}_i^{ut}) = u_j'(\hat{x}_j^{ut})$ for any $i, j \in N$. In fact, the function $W_{ut}(u)$ is concave (it is the sum of n concave functions) and it is thus maximized by the allocation that equalizes agents' marginal utility. If, on the other hand, the social planner adopts the maxmin SWF, the optimal allocation is the one that equalizes individuals' actual utility, i.e., $\hat{x}_{mm}$ is such that $u_i(\hat{x}_i^{mm}) = u_j\left(\hat{x}_j^{mm}\right)$ for any $i, j \in N$. Alternatively, another common

[13] Countless are the possible examples for such a situation: a parent who wants to divide a chocolate bar among her children, a boss who must share a monetary bonus among his subordinates, a judge called to decide how to divide the belongings of a divorcing couple, an organization that must allocate humanitarian aid to different villages hit by a natural disaster.

formulation of rational utility functions is the linear one.[14] In this case, $\hat{x}_{ut}$ is such that $\hat{x}_i^{ut} = 1$ for the i (assumed to be unique) with $u_i' > u_j'$ for any $j \neq i$ and $\hat{x}_{mm}$ is such that $u_i\left(\hat{x}_i^{mm}\right) = u_j\left(\hat{x}_j^{mm}\right)$ for any $i, j \in N$.

4.2 The case with self-serving biased reference points

With respect to the rational formulation, RDP seem better suited to model the preferences of claimants involved in an allocation problem. This is, in fact, a typical situation in which a claimant's utility, although mainly depending on the amount of resource that the agent gets, is usually also affected by comparisons between the actual allocation and the expected one (i.e., the reference point). And, as already discussed, reference points are in turn likely to be affected by the self-serving bias.[15]

Let claimants have preferences à la Koszegi and Rabin (2006), i.e., $u_i(x_i, r_i) = m(x_i) + \mu(m(x_i) - m(r_i))$. As in Section 3, here we assume again that $m(x_i) = x_i$ such that $u_i(x_i, r_i) = x_i + \mu(x_i - r_i)$. In the context of a bankruptcy problem this assumption proves to be very useful as the linear form of $m(\cdot)$ implies that the properties of the function $\mu(\cdot)$ directly translate into equivalent properties of the utility function $u_i(\cdot)$.[16] Finally, let $\sum_i r_i > 1$ such that, in line with Definition 1, at least some of the claimants have self-serving biased reference points. The planner knows the vector $r = (r_1, ..., r_n)$ but he does not know the size of individual biases so that he cannot correct for them.

In such a situation, utilitarian SWF is given by $W_{ut} = 1 + \sum_i \mu(x_i - r_i)$. Notice that the function W_{ut} is not guaranteed to be concave. In fact, the allocation $x = (r_1, ..., r_n)$ is unfeasible and the planner is forced to disappoint at least some of the claimants, i.e., he must allocate $x_i < r_i$ to some $i \in N$. This implies that

[14] This formulation can be considered as an approximation of a concave function for the cases in which the admissible range of x_i is small enough to make the marginal decreases in utility negligible. Because of this, linear utility functions are often implicitly assumed in many low stakes experimental studies about strategic interactions (bargaining games, ultimatum games, dictator games).

[15] In such a context, reference points can be seen as claims on shares of the total amount S.

[16] See Proposition 2 in Koszegi and Rabin (2006) for a formal statement and proof of this result.

some of the $\mu(x_i - r_i)$ functions are convex. Nevertheless, it is easy to prove that the optimal utilitarian allocation cannot be such that $x_i < r_i$ for all i.

Proposition 3 *The optimal utilitarian allocation $\hat{x}_{ut} = (\hat{x}_1^{ut}, ..., \hat{x}_n^{ut})$ is such that $\hat{x}_i^{ut} \geq r_i$ for some $i \in N$.*

Proof. By contradiction. Assume $\hat{x}_{ut}$ is such that $\hat{x}_i^{ut} < r_i$ for all i. Property P3 of the function $\mu(\cdot)$ states that $\mu_i''(\hat{x}_i^{ut}) > 0$ such that, given the linear form of $m(\cdot)$, the functions u_i are convex at $\hat{x}_i^{ut}$. Therefore the function W_{ut} is also convex. This implies that $\hat{x}_{ut}$ cannot be a maximum because it fails the second order necessary condition. ■

In terms of utilitarian welfare, any allocation such that $x_i < r_i$ for all i (like for instance the famous proportional rule that assigns $x_i = r_i / \left(r_i + \sum_{j \neq i} r_j \right))$ is thus inefficient. In particular, these allocations are dominated by any allocation that matches the reference points of some agents and leaves the others as residual claimants. In other words, it is more efficient to satisfy some agents and disappoint a lot the remaining ones rather than to disappoint a little all of them. The question is then how to decide who are the agents to disappoint and by how much. The following proposition addresses this point.

Proposition 4 *Assume that the constraint $x_i \leq r_i$ for any $i \in N$ must hold and, without loss of generality, order the claimants such that $r_1 \leq r_2 \leq ... \leq r_n$. Then the allocation $\hat{x}_{ut} = (\hat{x}_1^{ut}, ..., \hat{x}_n^{ut})$ with $\hat{x}_i^{ut} = \min \left\{ r_i, \max \left\{ 1 - \sum_{j<i} r_j, 0 \right\} \right\}$ is optimal.*

Proof. The planner's problem is given by $\max W_{ut} = 1 + \sum_i \mu(x_i - r_i)$. This is equivalent to $\min \sum_i \mu(x_i - r_i)$ given that $x_i \leq r_i$ must hold and therefore, by property P1, the functions $\mu(\cdot)$ are non positive. Moreover, property P3 ensures that $\mu(\cdot)$ exhibits diminishing marginal sensitivity such that $\mu(a) + \mu(b) < \mu(0) + \mu(a+b)$

for any $a, b < 0$. This implies that the planner must allocate $x_i = r_i$ to as many agents as possible (i.e., starting from those with the lowest r_i) while disappointing as much as possible the claimants that can be disappointed the most. The allocation rule $\hat{x}_i^{ut} = \min\left\{r_i, \max\left\{1 - \sum_{j<i} r_j, 0\right\}\right\}$ fulfills this goal. ■

The optimal allocation $\hat{x}_i^{ut}$ identified by Proposition 4 is unique whenever $r_{n-1} < \sum_{i=1}^{n} r_i - 1 \leq r_n$. If this condition does not hold then there could be multiple optimal allocations. Still, $\hat{x}_i^{ut}$ always belongs to the set of optimal solutions.

Example 3 *Consider two hypothetical situations with $n = 4$. In the first one let $r_1 = 0.1$, $r_2 = 0.3$, $r_3 = 0.5$ and $r_4 = 0.7$. Given that $r_3 < \sum_i r_i - 1 \leq r_4$, Proposition 4 identifies the unique optimal solution $\hat{x}_{ut} = (0.1, 0.3, 0.5, 0.1)$. In the alternative scenario let $r_1 = 0.1$, $r_2 = 0.2$, $r_3 = 0.5$ and $r_4 = 0.6$. Given that the condition $r_3 < \sum_i r_i - 1 \leq r_4$ does not hold, there are multiple optimal allocations. Proposition 4 identifies $\hat{x}_{ut} = (0.1, 0.2, 0.5, 0.2)$. But also the allocation $\hat{x}'_{ut} = (0.1, 0.2, 0.1, 0.6)$ achieves the maximal welfare $W_{ut} = 1 + \mu(-0.4)$.*[17]

Proposition 4 provides the solution to the problem when the condition $x_i \leq r_i$ must hold for all i. Relaxing this constraint, can it be welfare enhancing to allocate $x_i > r_i$ to some of the agents? The answer to this question clearly depends on the specific shape of claimants' utility functions. In particular, starting from the allocation $\hat{x}_{ut}$ identified by Proposition 4, the answer can be positive if and only if the decrease in welfare associated with further disappointing the residual claimant $\tilde{\imath}$ (in both scenarios of Example 3 this would be agent 4) by allocating him $\hat{x}_{\tilde{\imath}}^{ut} - \epsilon$ with $\hat{x}_{\tilde{\imath}}^{ut} = 1 - \sum_{j=1}^{\tilde{\imath}-1} r_j \geq \epsilon > 0$ is more than compensated by the increase stemming from redistributing ϵ among the claimants $i = \{1, ..., \tilde{\imath} - 1\}$. Formally, if and only if the condition $(\tilde{\imath} - 1)\mu\left(\frac{\epsilon}{\tilde{\imath}-1}\right) > -\left[\mu(\hat{x}_{\tilde{\imath}}^{ut} - \epsilon - r_{\tilde{\imath}}) - \mu(\hat{x}_{\tilde{\imath}}^{ut} - r_{\tilde{\imath}})\right]$ holds. If this is

[17] Notice anyway that a social planner with lexicographic preferences defined over utilitarian welfare and equality would strictly prefer the allocation $\hat{x}_{ut}$ over $\hat{x}'_{ut}$.

the case, then these unconstrained optimal allocations are identified by first- and second-order conditions or they emerge as a corner solution.

In any case, utilitarian welfare is surely smaller than 1. Utilitarian welfare would be larger ($W_{ut} = 1$) if the social planner could match all claims, i.e., if agents were unbiased and $\sum_i r_i = 1$. In other words, self-serving bias is welfare detrimental not only at the individual level (see Lemma 1) but also at the aggregate level.

Consider now what happens if the social planner adopts the maxmin SWF. This function selects the allocation at which utility functions intersect. Given the shape of agents' utility functions and the hypothesis of SSB, such a condition, if feasible, will usually arise in the interval where $x_i < r_i$ for all i. This implies that in general the optimal maxmin allocation is inefficient from a utilitarian point of view. The following case is particularly striking:

Proposition 5 *If* $\sum_i r_i > 1$ *and* $r_i = r$ *for all* i *then* $\hat{x}_{mm} = \left(\frac{1}{n}, ..., \frac{1}{n}\right) = \arg\min W_{ut}$.

Proof. If $r_i = r$ for all i then claimants are perfectly symmetric and the only feasible and Pareto efficient allocation that equalizes their utility is the egalitarian one. It follows that $\hat{x}_{mm} = \left(\frac{1}{n}, ..., \frac{1}{n}\right)$. Symmetry also implies that this is the unique allocation for which the FOCs of $\max W_{ut}$ are satisfied and $\sum_i x_i = 1$ holds. But given that $x_i < r_i$ for all i the functions u_i are convex and so is W_{ut}. Therefore, $\hat{x}_{mm}$ coincides with the minimum of the utilitarian SWF. ■

When claimants are perfectly symmetric, the maxmin SWF supports the egalitarian allocation. Indeed, possibly also because of its ethical appeal (in line with Aristotle's celebrated prescription that "equals should be treated equally"), this is certainly the most common solution implemented in reality. Still, Proposition 5 shows that such a choice implies a high efficiency cost. In fact, the egalitarian allocation happens to be the worst possible outcome from a utilitarian point of view.

5 Conclusion

We proposed a framework that allows to explicitly model the self-serving bias. In particular, we introduced self-serving bias within the family of reference dependent preferences by claiming that the bias systematically inflates agents' reference points. This consideration provides a simple rule to assess the existence of the bias at the aggregate level as well as a procedure to set a lower bound on the number of biased agents. We applied the model to two standard situations: a litigation between a plaintiff and a defendant and a bankruptcy problem. In the first case, the proposed model essentially confirms the detrimental role that the self-serving bias has on the probability of solving a dispute through a settlement rather then through a trial. In the second situation, the model of self-serving biased reference points amplifies instead the trade-off between the efficiency and the equity of the final allocation.

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

References

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

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

[3] 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.

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

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

[6] Bentham, J., 1789. *Introduction to the principles of morals and legislation*. London: Athlone.

[7] Bernheim, B. D., Rangel, A., 2007. Behavioral public economics: welfare and policy analysis with non-standard decision makers. In: Diamond, P., Vartiainen, H. (Eds.): Behavioral economics and its applications. Princeton, NY: Princeton University Press.

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

[9] Heyndels, B., Ashworth, J., 2003. Self-serving bias in tax perceptions: federalism as a source of political instability. *Kyklos* 56 (1), 47-68.

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

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

[12] Langlais, E., 2008. Asymmetric information, self-serving bias and the pretrial bargaining impasse. EconomiX working paper 2008-30, Universitè Paris X-Nanterre.

[13] Rawls, J., 1971. *A theory of justice*. Cambridge, MA: Harvard University Press.

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

[15] 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.

[16] 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.

[17] Studdert, D., Mello, M., Gawande, A., Gandhi, T., Kachalia, A., Yoon, C., Puopolo, A.L., Brennan, T., 2006. Claims, errors, and compensation payments in medical malpractice litigations. *The New England Journal of Medicine* 354 (19), 2024-33.

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

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