DEPFID

DEPARTMENT OF ECONOMIC POLICY, FINANCE AND DEVELOPMENT

UNIVERSITY OF SIENA

DEPFID WORKING PAPERS

Andrea Gallice

Self-Serving Biased Reference Points

9 / 2009

DIPARTIMENTO DI POLITICA ECONOMICA, FINANZA E SVILUPPO

UNIVERSITÀ DI SIENA

DEPFID Working Papers – 9 / November 2009

Self-Serving Biased Reference Points

Andrea Gallice

Abstract

The paper formalizes the pervasive phenomenon of the self-serving bias within the framework of reference dependent preferences. This formulation allows to state 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. As an application, we study the problem of the optimal allocation of a scarce resource among a finite number of claimants. We analyze the performance of different welfare criteria and show how the existence of self-serving biased individuals exacerbates the conflict between equity and efficiency of the final allocation.

KEYWORDS: Self-Serving Bias, Reference Dependent Preferences, Optimal Allocation.

JEL CLASSIFICATION: D00, D60.

ADDRESS FOR CORRESPONDENCE: andrea.gallice@eui.eu

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: European Symposium on Economics and Psychology (Mannheim), LabSi Conference on Political Economy and Public Choice (Siena), 23rd Annual Congress of the European Economic Association (Milan). The usual disclaimer applies.

Dipartimento di Politica Economica, Finanza e Sviluppo – Università di Siena

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, 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, 2000) 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 missing. In this paper we take a step in this direction by formally introducing SSB within the framework of reference dependent preferences.

Reference dependent preferences (from now on RDP) explicitly acknowledge the fact that an agent's evaluation of a given outcome is usually influenced by comparing it with an ex-ante reference point. This intuition goes back to the loss aversion conjecture introduced in the classical article by Kahneman and Tversky (1979): people define gains and losses with respect to a reference point and losses loom

[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%.

larger than gains.

In this paper we postulate that SSB affects agents' reference points in a trivial but systematic 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 to be shared) 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 use the proposed framework to study a well-known welfare problem, namely the optimal allocation of a scarce resource among a finite number of (self-serving biased) claimants.[2] The fact that claimants are biased implies that the allocation that matches each agent's reference point is unfeasible. The planner is thus forced to disappoint at least some of the claimants. Shall the planner disappoint (a little) all of them? 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 SSB exacerbates the trade-off between equity and efficiency of the final allocation. Finally, by allowing the claimants to strategically announce their reference points, we also rationalize the regularity of observing exceedingly high claims in disputes and litigations.

[2] As such, the paper 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.

2 Reference dependent preferences and self-serving bias

Koszegi and Rabin (2006) recently 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. 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 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 self-serving bias is likely to affect 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 implies 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 the interests of $n \geq 2$ agents whose preferences can be captured by RDP are in conflict. Examples include bargain-

ing problems, principal-agent relations, claimants fighting over a scarce resource, political lobbying. Let the utility function of agent $i \in \{1, ..., n\}$ be $u_i(x_i, r_i) = m(x_i) + \mu(m(x_i) - m(r_i))$.[3] 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 amount of resource to be shared. Normalizing this amount to one, this last condition reads as $\sum_i r_{i(unbiased)} = 1$ with $r_{i(unbiased)} \in [0, 1]$ for every i.[4] The situation 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 $\sum_i r_i > 1$ then at least some of the agents are self-serving biased.*

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

Proposition 1 *Given n 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 $\sum_{i=1}^{k} r_i > 1$ and $\sum_{i=1}^{k-1} r_i \leq 1$.*

[3] 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.

[4] We do not consider the situation of $\sum_i r_i < 1$ as this would imply that some agents display a self-defeating bias, an hypothesis whose empirical support is much weaker.

Proof. Consider the set $N = \{1, ..., n\}$ with $r_1 \leq r_2 \leq ... \leq r_n$. If $\sum_{i=1}^{n} r_i > 1$, then, by Definition 1, at least one player is biased. Imagine that n is the only biased agent. Moreover, imagine that his bias is extreme, i.e., $r_{n(unbiased)} = 0$. Now consider the set $N \setminus \{n\} = \{1, ..., n-1\}$. If $\sum_{i=1}^{n-1} r_i > 1$ 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 $N \setminus \{k, ..., n\} = \{1, ..., k-1\}$ with $\sum_{i=1}^{k} r_i > 1$ and $\sum_{i=1}^{k-1} r_i \leq 1$. 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 $n = 4$. 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 an allocation problem

Consider the problem of a social planner who must allocate a homogeneous and perfectly divisible good (whose amount we normalize to 1) among $n \geq 2$ claimants.[5] Let $N = \{1, ..., n\}$ denote the set of claimants. The notation $x = (x_1, ..., x_n)$ indi-

[5] 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.

cates 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 $w \in \{ut, mm\}$.

3.1 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 $u_i'(\hat{x}_i^{ut}) = \lambda$ for any $i \in N$. 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 $u_i(\hat{x}_i^{mm}) = \gamma$ for any $i \in N$. Alternatively, another common formulation

of rational utility functions is the linear one.[6] 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 $u_i\left(\hat{x}_i^{mm}\right) = \tau$ for any $i \in N$.

3.2 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 likely to be 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 usually affected by the self-serving bias.

Let claimants have preferences à la Koszegi and Rabin (2006). In particular, assume that $m(x_i) = x_i$ such that $u_i(x_i, r_i) = x_i + \mu(x_i - r_i)$. This assumption has one important implication. The linear form of $m(\cdot)$ implies in fact that the properties of the function $\mu(\cdot)$ directly translate into equivalent properties of the utility function $u_i(\cdot)$.[7] 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

[6] 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).

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

the claimants, i.e., to allocate $x_i < r_i$ to some $i \in N$. This implies that 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 2 *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 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 a 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 leave the others as residual claimants. In other words, it is more efficient to satisfy some agents and disappoint a lot some others 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 3 *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 3 is unique whenever $r_{n-1} < \sum_{i=1}^{n} r_i - 1 \leq r_n$. If this is not the case then there could be other optimal allocations. Still, $\hat{x}_i^{ut}$ always belongs to the set of optimal solutions.

Example 2 *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 3 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 3 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 same welfare $W_{ut} = 1 + \mu(-0.4)$.*[8]

Proposition 3 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 3, the answer can be positive if and only if

[8] 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 decrease in welfare associated with further disappointing the residual claimant $\tilde{\imath}$ (in both scenarios of Example 2 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 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 4 *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 4 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.

3.2.1 Some strategic considerations

The basic formulation of the allocation problem does not involve any strategic aspect: the social planner elicits agents' preferences and reference points, chooses a welfare criterion and finally implements the optimal solution. Claimants play no active role in the process. Still, the mere fact that a social welfare function uses as inputs individual utility functions suggests the possibility that agents may try to influence the final allocation by strategically disguising their real preferences.

In this section we show how the common regularity of observing exceedingly high claims in bilateral disputes can be rationalized by a model in which claimants expect the social planner to adopt a maxmin SWF and strategically announce their reference point. As such, the analysis applies to all those cases in which conflicting interests must be settled by an external player and litigants have the possibility to ex-ante declare what they expect to get. Examples include trials, divorces, reimbursements for damages and political negotiations.

As before let claimants' utility functions be given by $u_i = x_i + \mu(x_i - r_i)$ and let $i \in \{1, 2\}$. Assume that it is common knowledge that the planner will adopt a maxmin SWF and let the planner directly ask the claimants to announce their reference points. We indicate with $r_i^a \in [0, 1]$ the reference point announced by

agent i. Notice that r_i^a may differ from r_i, i.e., what an agent declares to expect (r_i^a) may differ from what he actually expects (r_i). The latter remains unknown to the planner who thus relies on the vector (r_1^a, r_2^a). As such, he implements the allocation $\hat{x}_{mm} = (\hat{x}_1^{mm}, \hat{x}_2^{mm})$ that solves $\hat{x}_1^{mm} + \mu(\hat{x}_1^{mm} - r_1^a) = \hat{x}_2^{mm} + \mu(\hat{x}_2^{mm} - r_2^a)$ subject to $\hat{x}_1^{mm} + \hat{x}_2^{mm} = 1$. We want to study the effects that r_i^a has on $\hat{x}_i^{mm}$ for any given r_i, r_j and r_j^a. Focusing on agent 1, the previous expression can be rewritten as:

$$F(\hat{x}_1^{mm}, r_1^a) = 2\hat{x}_1^{mm} + \mu(\hat{x}_1^{mm} - r_1^a) - \mu(1 - \hat{x}_1^{mm} - r_2^a) - 1 = 0$$

This is an implicit function that satisfies the assumptions of the implicit-function theorem. In fact, given that the planner considers $r_1^a = r_1$, property P2 of the gain-loss function $\mu(\cdot)$ ensures that partial derivatives $\frac{\partial F(\hat{x}_1^{mm}, r_1^a)}{\partial \hat{x}_1^{mm}}$ and $\frac{\partial F(\hat{x}_1^{mm}, r_1^a)}{\partial r_1^a}$ are continuous and different from zero for any $x_1 < r_1^a$. Total differentiation of $F(\hat{x}_1^{mm}, r_1^a)$ leads to:

$$\frac{\partial \mu(\hat{x}_1^{mm} - r_1^a)}{\partial r_1^a} + \left(2 + \frac{\partial \mu(\hat{x}_1^{mm} - r_1^a)}{\partial \hat{x}_1^{mm}} - \frac{\partial \mu(1 - \hat{x}_1^{mm} - r_2^a)}{\partial \hat{x}_1^{mm}}\right) \frac{\partial \hat{x}_1^{mm}}{\partial r_1^a} = 0$$

such that $\frac{\partial \hat{x}_1^{mm}}{\partial r_1^a}$ can be expressed as:

$$\frac{\partial \hat{x}_1^{mm}}{\partial r_1^a} = -\frac{\frac{\partial \mu(\hat{x}_1^{mm} - r_1^a)}{\partial r_1^a}}{2 + \frac{\partial \mu(\hat{x}_1^{mm} - r_1^a)}{\partial \hat{x}_1^{mm}} - \frac{\partial \mu(1 - \hat{x}_1^{mm} - r_2^a)}{\partial \hat{x}_1^{mm}}}$$

The numerator of the ratio is negative (by property P1) and the denominator is positive. In particular, the second term is positive (again by P1) while the third one is negative given that $\hat{x}_2^{mm}$ decreases as $\hat{x}_1^{mm}$ increases. It follows that $\frac{\partial \hat{x}_1^{mm}}{\partial r_1^a} > 0$. And given that utility $\hat{u}_1^{mm} = \hat{x}_1^{mm} + \mu(\hat{x}_1^{mm} - r_1)$ is strictly increasing in $\hat{x}_1^{mm}$ for any actual reference point r_1, the agent, even though he anticipates that he will get

$\hat{x}_1^{mm} < r_1^a$, should purposely inflate his (possibly already biased) claim and announce $r_1^a > r_1$.[9]

4 Conclusions

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. As an application, we analyzed the classical welfare problem of allocating a scarce resource among a finite number of claimants. The analysis highlighted the existence of a severe trade-off between the efficiency and the equity of the final allocation and it rationalized the regularity of observing exceedingly high claims in trials and litigations.

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, more in general, contributes to the recent literature about the public policy implications of research in behavioral economics.

[9] In the Nash equilibrium both players announce $r_i^a = 1$ and the planner implements $\hat{x}_{mm} = \left(\frac{1}{2}, \frac{1}{2}\right)$. Such a solution resembles king Solomon's proposal of splitting a baby in half in front of two women that were both claiming to be his mother.

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

[5] 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.

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

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

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

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

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

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

[12] 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.

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

DEPFID WORKING PAPERS

DIPARTIMENTO DI POLITICA ECONOMICA, FINANZA E SVILUPPO
UNIVERSITÀ DI SIENA
PIAZZA S. FRANCESCO 7 I- 53100 SIENA
http://www.depfid.unisi.it/WorkingPapers/
ISSN 1972 - 361X