DISCUSSION PAPER SERIES

No. 6572

SOME SOCIAL WELFARE IMPLICATIONS OF BEHAVIORAL PREFERENCES

Andrea Gallice

INDUSTRIAL ORGANIZATION and PUBLIC POLICY

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP6572.asp**

ISSN 0265-8003

SOME SOCIAL WELFARE IMPLICATIONS OF BEHAVIORAL PREFERENCES

Andrea Gallice, Ludwig-Maximilians-Universität München and ICER, Turin

Discussion Paper No. 6572
November 2007

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INDUSTRIAL ORGANIZATION and PUBLIC POLICY**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

CEPR Discussion Paper No. 6572

November 2007

ABSTRACT

Some Social Welfare Implications of Behavioral Preferences*

We investigate how the assumption that individuals are characterized by some recent forms of behavioral preferences changes the analysis of an otherwise classical welfare problem, namely the optimal allocation of a scarce resource among a finite number of claimants. We consider two preference specifications: inequity aversion and reference dependence. In the latter case we also study the implications of the claimants displaying a self-serving bias when setting their reference point. Using standard welfare criteria, we compute the optimal allocations that a benevolent social planner should implement in the various scenarios. Results are often remarkably different with respect to traditional (i.e., rational preferences) analysis. We discuss the policy implications and the role of a social planner.

JEL Classification: D01 and D61
Keywords: inequity aversion, optimum allocation, reference dependence, self-serving bias and social welfare

Andrea Gallice
International Centre for Economic
Research (ICER)
Villa Gualino
Viale Settimio Severo, 63
10133, Turin
Italy
Email: andrea.gallice@eui.eu

*I would like to thank Pascal Courty, Alessandro Innocenti, Klaus Schmidt, Sven Rady and Botond Koszegi for helpful suggestions as well as seminar participants at the University of Munich, Florence (Experimental Economics Workshop), Mannheim (European Symposium on Economics and Psychology) and Siena (Labsi International Conference on Political Economy and Public Choice) for useful comments. All errors are mine.

Submitted 07 November 2007

1 Introduction

In the last couple of decades, behavioral economics successfully challenged and enriched the traditional neoclassical analysis of individual behavior. This success is confirmed by a massive number of empirical and experimental studies that show how some of these behavioral contributions rationalize regularities that were previously labeled as puzzles. Nowadays terms such as "inequity aversion", "hyperbolic discounting", "reference dependent preferences" and "self-serving bias" belong to the vocabulary of many economists and there is little doubt that the phenomena indicated by these words do actually affect human behavior in many different contexts.

Acknowledging their importance at an individual level, one should also consider the effects that these same phenomena may have on the collectivity. In particular it seems interesting to study the social welfare and policy implications that these kinds of behavioral preferences may have with respect to the standard neoclassical preferences paradigm. The idea is that if behavioral models provide a better description of human behavior vis-à-vis traditional models, then welfare analysis must be retuned on these new forms of preferences as this can lead to a more realistic picture and to the implementation of more accurate policies. Some recent papers already proceed in this direction. For instance, O'Donoghue and Rabin (2006) investigate the issue of optimal taxation on sin goods (such as unhealthy food) when some consumers suffer from self-control problems as captured by time preferences based on quasi-hyperbolic discounting. In an analogous framework, Gruber and Koszegi (2001 and 2004) study how taxes can counteract individuals' addiction to cigarettes. Other papers (O'Donoghue and Rabin, 2001, Sunstein and Thaler, 2003, Choi *et al.*, 2003) consider various forms of "paternalistic" taxation on various kinds of goods.[1]

While sharing a similar approach, this paper does not study the issue of optimal taxation but focuses instead on a different problem of welfare economics, the problem of a benevolent social planner who must allocate a scarce, homogeneous and perfectly divisible resource

[1] Bernheim and Rangel (2005) provide a detailed overview of this recent literature about behavioral welfare analysis, focusing especially on problems involving saving, addiction and public goods.

among a finite number of claimants. Many 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 has to allocate humanitarian aid to different villages hit by a natural disaster.

The paper analyzes this allocative problem under the assumption that the claimants are characterized by various forms of behavioral preferences.[2] In particular, we consider three scenarios: the case of the claimants having inequity averse preferences (à la Fehr and Schmidt, 1999), the case of the claimants having reference dependent preferences (as formalized in Koszegi and Rabin, 2006) and the case in which reference dependent preferences are combined with a self-serving bias (discussed for instance in Babcock *et al.*, 1995). This last case is particularly interesting. On one hand, it captures situations that are very likely to arise and whose welfare/policy implications are unusual and remarkable. On the other hand, by formally linking the issue of how agents set reference points with the pervasive phenomenon of self-serving biases, it is also interesting from a methodological point of view.

This paper conveys many results. The more general lesson is that, even in this context, behavioral preferences can have important welfare implications, as optimal allocations are usually different with respect to the standard case with rational preferences. In the course of the analysis we will also be able to state more specific results that answer the following questions: if agents are inequity averse, does a social planner need to know individuals' actual preferences or shall he worry about how to define social welfare? (The answer is no in both cases). If individuals characterized by reference dependent preferences are self-serving biased and a first best solution is not achievable, is it more efficient to disappoint (a little) all the claimants or is it better to please some of them and disappoint (a lot) the remaining ones? If the second approach is superior (as we show it to be), how should the planner choose who to privilege? (The answer is that it does not matter). We also show how a

[2] The case with rational preferences has been carefully analyzed. See, among many others, Steinhaus (1948) and Dubins and Spanier (1961), or much more recently, Brams and Taylor (1996) and Maccheroni and Marinacci (2003).

model of biased claimants characterized by reference dependent preferences can rationalize the regularity of observing exceedingly high claims in litigations.

The remainder of the paper is organized as follows: Section 2 formalizes the problem and presents the traditional solution. Sections 3 considers the case in which the claimants are inequity averse. Section 4 studies the situation in which agents are characterized by reference dependent preferences starting with the case of agents with no self-serving bias and moving to the case where such a bias exists. In the latter case some strategic implications and the role of a social planner are also discussed. Section 5 concludes.

2 The problem and its standard solution

A social planner must allocate a homogeneous and perfectly divisible good (whose amount we normalize to 1) among $N \geq 2$ 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 $i \in \{1, ..., N\}$. Feasible allocations are the ones for which $x_i \in [0, 1]$ for any i and $\sum_i x_i \leq 1$. Each individual i has a utility function $u_i(x)$. Notice that this formulation allows the agent's utility to depend not only on x_i (as it happens with neoclassical preferences) but possibly also on other components of the vector x (as it happens with some kinds of behavioral preferences). In general it will be the case that $\frac{\partial u_i(x)}{\partial x_i} > 0$ for any $x_i \in [0, 1]$ such that claimants do not have any feasible satiation point. The vector $u = (u_1(x), ..., u_N(x))$ collects individual utilities.

The social planner wants to maximize social welfare. His objective function is given by a social welfare function (SWF) that takes the form $W(u) = W(u_1(x), ..., u_N(x))$, i.e., a function 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 all the agents. More precisely we consider three welfare functions: the utilitarian SWF, the maxmin SWF and what we call the fair SWF.

The utilitarian SWF has a very long tradition in welfare economics (starting with Ben-

tham, 1789) and prescribes that the social planner implements the allocation that maximizes the sum of individual utilities.

- *Utilitarian SWF*: $W_{ut}(u) = \sum_i u_i(x)$

At the opposite, a social planner who adopts the maxmin or Rawlsian (from Rawls, 1971) SWF wants to maximize the welfare of the worst-off individual.[3]

- *Maxmin SWF*: $W_{mm}(u) = \min\{u_1(x), ..., u_N(x)\}$

Finally, we introduce a function that selects the "fairest" available allocation where, in accordance with the literature on fair divisions (see for instance Varian, 1974 or Brams and Taylor, 1996), an allocation is called fair if it is Pareto-efficient and envy-free. Envy-freeness (as defined in Foley, 1967) means that no claimant envies the amount of the good received by the other agents: $u_i(x_i) \geq u_i(x_j)$ for any i and any j.[4]

- *Fair SWF:* $W_{fa}(u) = \frac{1}{1+\sum_i \sum_j \max\{u_i(x_j)-u_i(x_i),0\}}$ s.t. $\sum_i x_i = 1$

These three social welfare functions are clearly inspired by different motivations and will in general lead to different solutions. 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 $w \in \{ut, mm, fa\}$. In line with the benevolent nature of the social planner, a common feature of these three optimal allocations is that they all satisfy the Pareto principle.[5] More precisely, $\hat{x}_{ut}$, $\hat{x}_{mm}$ and $\hat{x}_{fa}$ are such that

[3] The utilitarian and the maxmin SWFs are the extremes of a family of functions captured by the so-called generalized utilitarian SWF. This function is given by $W(u) = \sum_i g(u_i(x_i))$ where g is a concave function. The more g is concave the more the resulting allocation will be equitable (see Mas-Colell *et al.*, 1995, or Moulin, 2003).

[4] The concepts of fair division and envy-freeness have also been applied to problems that somehow differ with respect to our basic framework. For instance, Tadenuma and Thompson (1995) investigate the case where the good is indivisible while Brams *et al.* (2006) and Dall'Aglio and Maccheroni (2007) consider the case of a non-homogeneous good. Another important strand of the literature focuses on procedures (for instance "moving-knife" protocols like "divide and choose") whose goal is to lead to fair allocations even when the planner does not know the claimants' preferences. See, for instance, Abreu and Sen (1990) and Bag (1996).

[5] There are situations in which the social acceptability of the Pareto principle may be disputable (see for instance Sen, 1977 and 1979, who considers the case of agents with illiberal or antisocial preferences). Still, these criticisms do not seem to apply to our simple allocation problem, leaving the Pareto principle as a valid objective a social planner should pursue.

$\sum_i \hat{x}_i^w = 1$. The first two SWFs automatically satisfy the Pareto principle given that an agent's utility is increasing in x_i. The third function is forced to be Pareto optimal by the constraint. Consider, for instance, the allocation $x = (0, ..., 0)$. This allocation is trivially envy-free but it does not qualify as being fair given that it is not efficient.

To actually find the optimal allocations identified by the three SWFs, one needs to know how claimants' utility functions are defined. Traditional neoclassical analysis postulates each agent i to have preferences that are exclusively defined on x_i and that lead to continuous, increasing and concave utility functions. More formally, $u_i(x) = u_i(x_i)$ with $\frac{\partial u_i(x_i)}{\partial x_i} > 0$ and $\frac{\partial^2 u_i(x_i)}{\partial x_i^2} < 0$ for any i. In such a situation, the utilitarian SWF selects $\hat{x}_{ut} = (\hat{x}_1^{ut}, ..., \hat{x}_N^{ut})$ with $\frac{\partial u_i(x_i)}{\partial x_i} \mid_{x_i = \hat{x}_i^{ut}} \equiv k$ for any i. In fact, the function $W_{ut}(u)$ is concave (it is the sum of N concave functions) and is maximized by the allocation that equalizes agents' marginal utilities. If, on the other hand, the social planner adopts the maxmin SWF, the optimal allocation is the one that equalizes individuals' actual utilities, i.e., $\hat{x}_{mm} = (\hat{x}_1^{mm}, ..., \hat{x}_N^{mm})$ such that $u_i(\hat{x}_i^{mm}) \equiv \gamma$ for any i. Finally, the allocation selected by the fair SWF is the one that equalizes individuals' endowments: $\hat{x}_{fa} = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$.

Alternatively, another common formulation of rational utility functions is the linear one such that $u_i(x) = u_i(x_i)$, $\frac{\partial u_i(x_i)}{\partial x_i} > 0$ and $\frac{\partial^2 u_i(x_i)}{\partial x_i^2} = 0$ for any i. Such a formulation can actually be considered as an approximation of concave functions for the cases in which the admissible range of x_i (in our case the size of the pie before normalization) is small enough to make the marginal decreases in utility negligible.[6] Optimal allocations with linear utility functions are then given by $\hat{x}_{ut} = (\hat{x}_1^{ut}, ..., \hat{x}_N^{ut})$ with $\hat{x}_i^{ut} = 1$ for the i (assumed to be unique) such that $\frac{\partial u_i(x_i)}{\partial x_i} > \frac{\partial u_j(x_j)}{\partial x_j}$ for any $j \neq i$, $\hat{x}_{mm} = (\hat{x}_1^{mm}, ..., \hat{x}_N^{mm})$ such that $u_i(\hat{x}_i^{mm}) \equiv \gamma$ for any i and $\hat{x}_{fa} = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$.

Figures 1.*a* and 1.*b* provide graphical examples for the case with two claimants. In each diagram the utility function of agent 1 is displayed from left to right. The utility of agent 2 goes from right to left. The length of the horizontal axis is fixed to 1 and represents the total amount of the good that the planner must allocate.

[6] Because of this, linear utility functions are often implicitly assumed in many low stakes experimental studies about strategic interactions (Ultimatum game, Dictator game, public goods games...).

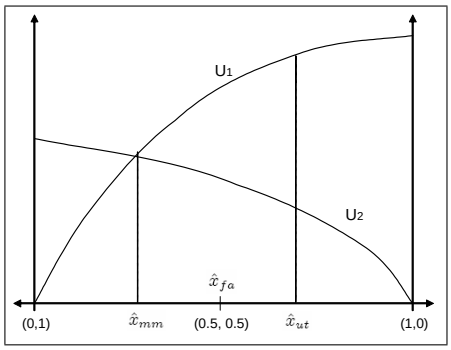

Fig. 1.a: concave utility functions.

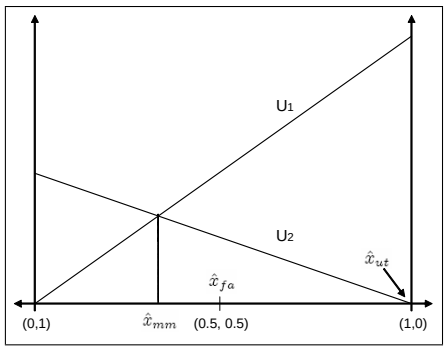

Fig. 1.b: linear utility functions.

Notice that, in general, the utilitarian, the maxmin and the fair SWFs lead to different allocations. Indeed, the three welfare functions select the same allocation (namely the egalitarian one, such that $\hat{x}_i^w = \frac{1}{N}$ for any $i \in \{1, ..., N\}$ and any $w \in \{ut, mm, fa\}$) only when all the agents are perfectly symmetric. In the following sections we consider how these results change when claimants are characterized by some kinds of behavioral preferences.

3 Inequity averse preferences

Neoclassical preferences implicitly assume that agents only care about their own payoff without being influenced by the payoffs to others in some appropriate reference group. In other words, agents are assumed to be totally selfish. Real life evidence, as well as a large number of experimental studies, show that this assumption often does not hold, as an agent's utility is usually affected by intra-group comparisons. In the last few years various papers modeled these kinds of "social preferences".[7] Particularly successful have been models of "inequity aversion" (prominent examples are Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000 and Charness and Rabin, 2002). An inequity averse agent is an agent whose utility, holding fixed what he gets, decreases with the degree of inequality that arises in the reference group. Inequity aversion therefore captures feelings such as envy (the agent gets less than

[7] The state of the art of these theories as well as their empirical evidence is carefully reviewed in Fehr and Schmidt (2005).

others) and guilt/embarrassment (the agent gets more).

In the context of our allocative problem, these issues seem to be very likely to affect an individual's ex-post assessment (and thus his utility) about the allocation implemented by the social planner.[8] Therefore, we study if and how the assumption of claimants being inequity averse changes traditional results. More precisely, we adopt the widely used specification introduced by Fehr and Schmidt (1999), where the utility function of individual $i \in \{1, ..., N\}$ is assumed to have the following form:

$$u_i = x_i - \alpha_i \frac{1}{N-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{N-1} \sum_{j \neq i} \max\{x_i - x_j, 0\}$$

In the original article the authors assume that $\beta_i \leq \alpha_i$ and that $0 \leq \beta_i < 1$. The first restriction implies that individuals cannot suffer less from disadvantageous inequality than from advantageous inequality. The second restriction rules out the existence of individuals that would be so inequity averse that they would be willing to heavily harm themselves for the sake of equity. Notice that these restrictions allow α_i and β_i to be equal to 0 so that some individuals could still have (linear) purely selfish preferences.

Given that issues of equity are already embedded into individuals' preferences, the fact that optimal allocations will be strongly egalitarian is not surprising. Indeed, all three SWFs select the symmetric allocation, such that $\hat{x}_{ut} = \hat{x}_{mm} = \hat{x}_{fa} = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$ and $u_i(x) = \frac{1}{N}$ for any i. Perhaps more surprising is the fact that this result is not affected at all by the specific values of the individual parameters α_i and β_i. In fact, a single claimant j being strictly inequity averse (i.e., with $\beta_j > 0$) is sufficient to make the utilitarian and the maxmin solutions collapse into the fair allocation. Proposition 1 formalizes and proves this claim.

Proposition 1 *If claimants have inequity averse preferences then* $\hat{x}_w = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$ *for any* $w \in \{ut, mm, fa\}$ *and any* α_i *and* β_i *as long as there is at least one agent* j *such that* $\beta_j > 0$.

[8] On the other hand, another important family of "other regarding preferences", namely those captured by models of intention-based reciprocity (Rabin, 1993, Dufwenberg and Kirchsteiger, 2004), do not apply to our framework where claimants simply must accept the chosen allocation and cannot reciprocate the social planner.

Proof. Consider the symmetric egalitarian allocation given by $\hat{x} = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$ such that $W_{ut}\left(u\left(\hat{x}\right)\right) = 1$, $W_{mm}\left(u\left(\hat{x}\right)\right) = \frac{1}{N}$ and $W_{fa}\left(u\left(\hat{x}\right)\right) = 1$. Now implement the best possible deviation from $\hat{x}$: take ϵ away from player j where j is such that $\alpha_j = \min\left\{\alpha_1, ..., \alpha_N\right\}$ and give it to agent k where k is such that $\beta_k = \min\left\{\beta_1, ..., \beta_N\right\}$. Call this new allocation $\tilde{x}$. Utilitarian welfare would now be:

$$W_{ut}\left(u\left(\tilde{x}\right)\right) = \sum_i \left(\tfrac{1}{N}\right) \; -\epsilon + \epsilon \underbrace{-\alpha_j \frac{1}{N-1}\left[2\epsilon + (N-2)\,\epsilon\right]}_{\text{ineq. av. of } j \text{ wrt } k \text{ and } l \neq k} \underbrace{-\beta_k \frac{1}{N-1}\left[2\epsilon + (N-2)\,\epsilon\right]}_{\text{ineq. av. of } k \text{ wrt } j \text{ and } l \neq j}$$
$$\underbrace{-\sum_{l \neq j,k}\left(\alpha_l \frac{1}{N-1}\epsilon + \beta_l \frac{1}{N-1}\epsilon\right)}_{\text{ineq. av. of } l \neq j,k \text{ wrt } j \text{ and } k}$$

which simplifies to: $W_{ut}\left(u\left(\tilde{x}\right)\right) = 1 - \left[\alpha_j \frac{N}{N-1}\epsilon + \beta_k \frac{N}{N-1}\epsilon + \sum_{l \neq j,k}\left(\alpha_l \frac{1}{N-1}\epsilon + \beta_l \frac{1}{N-1}\epsilon\right)\right] < 1$ given that the terms in the square brackets cannot be negative and at least one of them is strictly positive. Therefore, $W_{ut}\left(u\left(\tilde{x}\right)\right) < W_{ut}\left(u\left(\hat{x}\right)\right)$. Considering the maxmin SWF, welfare is given by $W_{mm}\left(u\left(\tilde{x}\right)\right) = u_j\left(\tilde{x}_j\right) = \frac{1}{N} - \epsilon - \alpha_j \frac{1}{N-1}\left[2\epsilon + (N-2)\,\epsilon\right] < \frac{1}{N}$ such that $W_{mm}\left(u\left(\tilde{x}\right)\right) < W_{mm}\left(u\left(\hat{x}\right)\right)$. Finally $\tilde{x}$ is not envy-free because the condition $u_j\left(\tilde{x}_j\right) = u_j\left(\tilde{x}_k\right)$ cannot be verified by valid parameters (i.e., the equality holds if $\alpha_j = \beta_k - 1$ but this cannot happen given the constraints $\alpha_j > 0$ and $\beta_k < 1$). It follows that $W_{fa}\left(u\left(\tilde{x}\right)\right) < W_{fa}\left(u\left(\hat{x}\right)\right)$. Therefore, any deviation from the egalitarian outcome strictly decreases social welfare under all three specifications. ■

Proposition 1 indicates that the task of a social planner who is facing inequity averse agents is pretty simple. In fact, the planner does not need to worry about guessing or eliciting the parameters of individuals' utility functions; indeed, he does not even have to care about how to solve the trade-off between efficiency and equity in deciding which social welfare function to adopt. By sharing the good equally among all the claimants the social planner is sure to maximize welfare, no matter how this is defined. Figure 2 depicts the situation with two claimants. Agent 1's utility function (solid line) goes from left to right and it is such that $0 < \beta_1 < 0.5$. Agent 2's utility function (dashed line, from right to left) is instead such that $0.5 < \beta_2 < 1$.

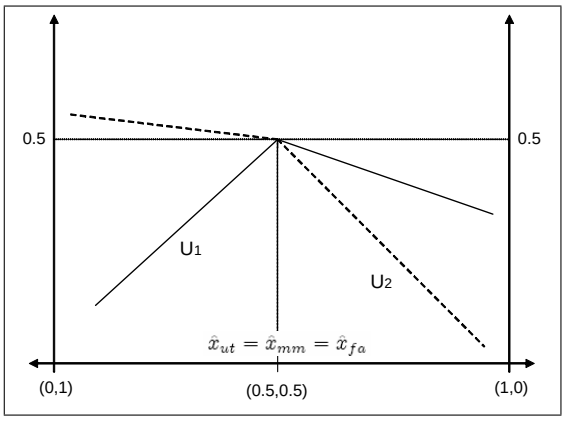

Figure 2: inequity averse preferences.

4 Reference dependent preferences

Another important family of behavioral preferences is labeled reference dependent preferences. With respect to the standard neoclassical formulation, these preferences explicitly acknowledge the fact that an agent's perception of a given outcome may be influenced by its comparison with a certain 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 the negative utility associated with a loss is higher than the positive utility associated with a gain of the same size.

A utility function that formally captures this intuition has been recently introduced by Bowman *et al.* (1999) and Koszegi and Rabin (2006). Such a utility function can be expressed as $u_i = m(x_i) + \mu(x_i - r_i)$ where, in the context of our allocative problem, x_i is a component of the vector $x = (x_1, ..., x_N)$ and indicates the amount of the good that the social planner assigns to claimant i. The function $m(x_i)$ is a traditional utility term that captures the direct effect that the possession/consumption of x_i has on total utility u_i. The function $\mu(x_i - r_i)$ is a "universal gain-loss function" and reflects the additional effects that perceived gains and losses (defined with respect to the individual's *ex ante* reference point r_i) have on u_i. Notice that such a specification allows for heterogeneity in agents' reference

points but adopts common functions for what concerns $m(\cdot)$ and $\mu(\cdot)$. On one hand, this is an important restriction that we will later partly relax by considering the implications of assuming an heterogeneous function $\mu_i(\cdot)$. On the other hand, such a representative agent formulation looks more realistic, as it assigns a lower informational burden to the social planner. In fact, while it seems reasonable to assume that the planner could reasonably infer (by eliciting or asking) individuals' reference points, the task of eliciting individuals' actual functions appears to be prohibitive because of informational constraint and implementation costs.

We now introduce and discuss a more specific functional form for the functions $\mu(\cdot)$ and $m(\cdot)$. In accordance with the properties discussed in Kahneman and Tversky (1979) and closely following Koszegi and Rabin (2006), the gain-loss function $\mu(\cdot)$ is assumed to satisfy the following requirements:

A1: $\mu(x_i - r_i)$ is continuous, strictly increasing in $(x_i - r_i)$ and such that $\mu(0) = 0$.

A2: $\mu(x_i - r_i)$ is twice differentiable with respect to x_i for $x_i \neq r_i$.

A3: $\frac{\partial^2 \mu(x_i - r_i)}{\partial x_i^2} > 0$ if $x_i < r_i$ and $\frac{\partial^2 \mu(x_i - r_i)}{\partial x_i^2} < 0$ if $x_i > r_i$.

A4: if $x_i' > x_i > r_i$ then $\mu(x_i' - r_i) + \mu(r_i - x_i') < \mu(x_i - r_i) + \mu(r_i - x_i)$.

A5: $\lim_{x_i \to r_i^-} \frac{\partial \mu(x_i - r_i)}{\partial x_i} / \lim_{x_i \to r_i^+} \frac{\partial \mu(x_i - r_i)}{\partial x_i} \equiv \lambda > 1$.

Some brief comments on the assumptions that look less transparent: A3 states that the function $\mu(\cdot)$ is convex for values of x_i that are below r_i (domain of losses) and concave for values of x_i that are above r_i (domain of gains). It also implies that the marginal influence of these perceived gains and losses is decreasing. A4 means that for large absolute values of x_i the function $\mu(\cdot)$ is more sensitive to losses than to gains. A5 implies the same result for small values of x_i: $\mu(\cdot)$ is steeper approaching the reference point from the left (losses) rather than from the right (gains). Taken together, these last two assumptions capture the loss aversion phenomenon. Finally, notice that A1 implies that $\mu(\cdot)$ is decreasing in r_i; in other words, for any given allocation x_i, agent i enjoys more utility if he has a lower reference point r_i.

For what concerns $m(x_i)$, the non-behavioral component of the utility function u_i, we set $m(x_i) = x_i$. This assumption has two important implications. First, it makes the analysis more comparable with the inequity aversion case studied in the previous section, where again the core part of the utility function was given by x_i. Second and more importantly, the linear form of $m(\cdot)$ implies that the properties of the function $\mu(\cdot)$ directly translate into equivalent properties of the utility function $u_i(\cdot)$.[9]

We know quite a few things about the function $u_i = x_i + \mu(x_i - r_i)$. What we still do not know is how an individual sets his reference point r_i. 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 (in line with the traditional status quo formulation of Kahneman and Tversky, 1979), to what they expect (as proposed by Koszegi and Rabin, 2006) or to what they think they deserve, just to name a few possibilities.

In this paper we relate the issue of how individuals set their reference points with another widespread behavioral regularity: self-serving bias. Self-serving bias is a pervasive phenomenon that influences people's behavior in various ways. For example, individuals tend to over-estimate their own merits, to favorably acquire and interpret information, to give biased judgments about what is fair and what is not, to inflate their claims and contributions. Research in psychology and sociology provides many nice examples of the existence of such a bias. For instance, the overwhelming majority of subjects (Svenson, 1981, reports 93%) declare to belong to the top 50% of drivers. More general survey studies show that more than 50% of respondents believe they are in the top 50% of any given category. Possibly more in line with the present paper, there is the robust finding that shows how estimates of the two members of married couples that are asked to indicate their own contribution to various household tasks usually sum up to more than 100% (Ross and Sicoly, 1979).

Self-serving bias can also have important 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). Even if the importance of such a

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

bias is widely acknowledged in the economic literature, a proper formalization of the concept and the analytical study of its implications are still missing. By inserting self-serving bias into the framework of reference dependent preferences, we think we are taking a step in this direction.

In the context of our allocative problem, the existence of a self-serving bias is likely to affect agents' reference points. In fact, everything else being equal, a self-serving biased claimant will have the tendency to set a higher reference point with respect to a claimant who is not biased. Consider the situation of the latter example. An unbiased agent should set his reference points as if he were "behind the veil of ignorance" (for instance, before knowing the kind of relationship he has with the planner and the other claimants or the nature of the good). Given that the agent knows that there are N claimants (including himself), he expects to get a fair portion of the good, i.e., $r_i = \frac{1}{N}$. This implies that reference points are mutually compatible whenever all the agents are unbiased ($\sum_i r_i = 1$). At the opposite, self-serving biased agents will set $r_i > \frac{1}{N}$. It follows that, whenever some of the agents have such a bias, reference points are no longer compatible ($\sum_i r_i > 1$).[10]

In the analysis that follows, we differentiate between two extreme cases. First, we study the allocation problem in a situation where no claimant is biased. Then we analyze the same problem under the assumption that all the claimants are biased. This second case is particularly interesting as it captures situations that are very likely to arise and that often have important economic implications. Think, for instance, about a divorcing couple who cannot agree on how to share their properties given that both persons claim more than 50%. Similarly, consider the case of heirs fighting over how to divide a bequest, or the situation of economic partners arguing about how to share the profits of a joint venture. Whenever reference points are not compatible and no agent is willing to concede, then no agreement can be reached on any feasible allocation. Therefore the claimants have to ask some external actor (a judge, an authority, or in our case, the planner) to solve their dispute.

[10] Notice that an individual can set $r_i > \frac{1}{N}$ without being biased if he actually deserves more. However, if all the agents set $r_i > \frac{1}{N}$ then at least some individual is surely biased. We do not consider the less common situation of agents displaying a self-defeating bias such that they may set $r_i < \frac{1}{N}$.

4.1 The case of agents with no self-serving bias

Agents that are not self-serving biased aim to get a fair share of the pie. Their utility function is given by $u_i = x_i + \mu(x_i - \frac{1}{N})$. The solution to the planner's problem is captured by the following proposition:

Proposition 2 *If claimants have reference dependent preferences and no self-serving bias, then* $\hat{x}_w = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$ *for any* $w \in \{ut, mm, fa\}$ *provided that* $\frac{|\mu(-\epsilon)|}{\mu\left(\frac{\epsilon}{N-1}\right)} > (N-1)$ *for any* $\epsilon \in \left(0, \frac{1}{N}\right]$.

Proof. The symmetric allocation $\hat{x} = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$ leads to $W_{ut}\left(u\left(\hat{x}\right)\right) = 1$, $W_{mm}\left(u\left(\hat{x}\right)\right) = \frac{1}{N}$ and $W_{fa}\left(u\left(\hat{x}\right)\right) = 1$. Now consider $W_{ut}\left(u\left(\tilde{x}\right)\right)$ and $W_{ut}\left(u\left(\breve{x}\right)\right)$ where $\tilde{x}$ is obtained from $\hat{x}$ by taking $\epsilon \in \left(0, \frac{1}{N}\right]$ away from agent j and redistributing it equally to the other $N-1$ agents while $\breve{x}$ is obtained from $\hat{x}$ by taking $\frac{\epsilon}{N-1} \in \left(0, \frac{1}{N(N-1)}\right]$ away from each of the $N-1$ $k \neq j$ agents and the resulting ϵ is reallocated to agent j. Because of the assumptions on the gain-loss function $\mu(\cdot)$, $\tilde{x}$ and $\breve{x}$ dominate all the other possible asymmetric allocations. At $\tilde{x}$ utilitarian welfare is $W_{ut}\left(u(\tilde{x})\right) = 1 + \mu(-\epsilon) + (N-1)\mu\left(\frac{\epsilon}{N-1}\right)$ such that $W_{ut}\left(u(\tilde{x})\right) < W_{ut}\left(u\left(\hat{x}\right)\right) = 1$ if and only if $\frac{|\mu(-\epsilon)|}{\mu\left(\frac{\epsilon}{N-1}\right)} > (N-1)$. At $\breve{x}$ welfare is $W_{ut}\left(u(\breve{x})\right) = 1 + (N-1)\mu\left(-\frac{\epsilon}{N-1}\right) + \mu\left(\epsilon\right)$ and $W_{ut}\left(u(\breve{x})\right) < 1$ as $(N-1)\mu\left(-\frac{\epsilon}{N-1}\right) + (N-1)\mu\left(\frac{\epsilon}{N-1}\right) < 0$ by A4 and A5 and $(N-1)\mu\left(\frac{\epsilon}{N-1}\right) > \mu\left(\epsilon\right)$ by A3. According to maxmin welfare we have $W_{mm}\left(u\left(\tilde{x}\right)\right) < W_{mm}\left(u\left(\breve{x}\right)\right)$ given that $-\frac{\epsilon}{N-1} > -\epsilon$ and A1. Moreover $W_{mm}\left(u\left(\breve{x}\right)\right) = \frac{1}{N} - \frac{\epsilon}{N-1} + \mu(-\frac{\epsilon}{N-1}) < \frac{1}{N} = W_{mm}\left(u\left(\hat{x}\right)\right)$ given that $\mu(-\frac{\epsilon}{N-1}) < 0$ and A1. Finally, under fair welfare maximization we have $W_{fa}\left(u\left(\tilde{x}\right)\right) < W_{fa}\left(u\left(\breve{x}\right)\right) < W_{fa}\left(u\left(\hat{x}\right)\right) = 1$ given that $u_j\left(\tilde{x}_k\right) - u_j\left(\tilde{x}_j\right) > u_k\left(\breve{x}_j\right) - u_k\left(\breve{x}_k\right) > 0$ for any $k \neq j$. ■

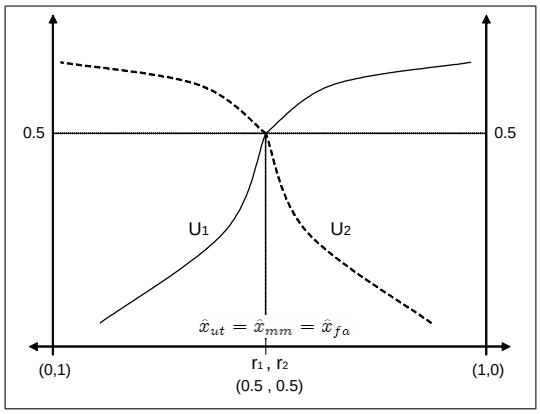

Figure 3: reference dependent preferences and no self-serving bias.

To sum up, the maxmin and fair SWFs always select $\hat{x} = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$ while the utilitarian SWF selects $\hat{x}$ provided that $\frac{|\mu(-\epsilon)|}{\mu\left(\frac{\epsilon}{N-1}\right)} > (N-1)$. For any given $\mu(\cdot)$, this condition is more easily fulfilled when N is low. Indeed, with $N = 2$, it surely holds as $|\mu(-\epsilon)| > \mu(\epsilon)$ by A5. The case with two claimants is depicted in Figure 3.

Whenever $\hat{x}_w = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$ for any $w \in \{ut, mm, fa\}$, the situation may seem analogous to the case of inequity averse individuals: the optimal allocation is the symmetric one no matter which SWF the social planner may use. Still, notice that, with reference dependent preferences, the utilitarian SWF selects the egalitarian outcome despite neither the planner nor the claimants having any explicit preference for equity. In fact, it is true that claimants are implicitly stating a preference for the egalitarian outcome by setting $r_i = \frac{1}{N}$. However, any individual i is actually indifferent to whether the implemented allocation is egalitarian or not, as long as he gets $x_i = \frac{1}{N}$. With respect to the inequity aversion case, agents do not compare what they get with what the others get but with what they were expecting to get.

Notice also that the results of Proposition 2 are less general as they rely on the assumption of agents having the same gain-loss function $\mu(\cdot)$ and thus basically the same utility function $u_i(\cdot)$. Allowing for heterogeneity in the gain-loss function (i.e., $\mu_i(\cdot)$), results become less clear-cut. The maxmin SWF always selects $\hat{x}_{mm} = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$ given that this is the point at

which claimants' utility functions intersect no matter the specific shapes of $\mu_i(\cdot)$. Similarly, the fair SWF keeps selecting $\hat{x}_{fa} = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$ because this remains the only envy-free and Pareto-efficient allocation. What may change is the solution selected by the utilitarian SWF, which can easily be asymmetric. For instance, by using the slope of the utility functions rather than the absolute values (as in Proposition 2), if there are two agents such that $\lim_{x_j \to \left(\frac{1}{N}^+\right)} \frac{\partial \mu_j(x_j - \frac{1}{N})}{\partial x_j} > \lim_{x_k \to \left(\frac{1}{N}^-\right)} \frac{\partial \mu_k(x_k - \frac{1}{N})}{\partial x_k}$ then $\hat{x}_{ut}$ will be such that $x_j > \frac{1}{N} > x_k$. Also, if $\lim_{x_j \to (1^-)} \frac{\partial \mu_j(x_j - \frac{1}{N})}{\partial x_j} > \sum_k \lim_{x_k \to (0^+)} \frac{\partial \mu_k(x_k - \frac{1}{N})}{\partial x_k}$ for any $k \neq j$ then agent j will be allocated the entire pie.

4.2 The case of agents with self-serving bias

In this section we analyze the widespread situation in which claimants, in setting their reference point, suffer from a self-serving bias. In particular we restrict the analysis to the common case of the social planner having to allocate the good among just two agents. An appropriate example is the one of a judge that must settle the dispute of a divorcing couple who, because both partners are self-serving biased, cannot agree on how to split the common belongings.[11]

Claimants' utility function is given by $u_i = x_i + \mu(x_i - r_i)$ with $r_i > \frac{1}{2}$ for any $i \in \{1, 2\}$ (as we did in the previous section, we will later discuss the case of $\mu_i(\cdot)$). Starting with the utilitarian criterion, total welfare is given by $W_{ut}(u(x)) = 1 + \mu(x_1 - r_1) + \mu(x_2 - r_2)$. The optimal (internal) allocation is then defined as $\hat{x}_{ut} = (\hat{x}_1^{ut}, \hat{x}_2^{ut})$ such that $\frac{\partial(1+\mu(x_1-r_1)+\mu(1-x_1-r_2))}{\partial x_1}\Big|_{x_1=\hat{x}_1^{ut}} = 0$ subject to $\frac{\partial^2(1+\mu(x_1-r_1)+\mu(1-x_1-r_2))}{\partial x_1^2}\Big|_{x_1=\hat{x}_1^{ut}} < 0$, $\hat{x}_1^{ut} + \hat{x}_2^{ut} = 1$ and $\hat{x}_1^{ut} \in [0, 1]$. Notice that the second order condition is now necessary as the function $W_{ut}(u(x))$ is not guaranteed to be concave. In fact, given that the allocation $x = (r_1, r_2)$ is unfeasible, the planner is forced to disappoint at least one of the claimants. Because of the assumptions about the gain-loss function $\mu(\cdot)$ (see page 11), this implies that $W_{ut}(u(x))$ is either the sum of one concave and one convex function or the sum of two

[11] There is quite a vast literature in sociology and psychology (see for instance Schriber *et al.*, 1985, Gray and Silver, 1990, and Schuutz, 1999) that shows how the existence of self-serving biases is a very important cause of conflicts and divorces in married couples.

convex functions. In the discussion that follows we define three intervals to which the share of the good that the planner allocates to agent 1 can belong (see Figure 4 for a graphical analysis). Given that the planner will certainly implement a Pareto optimal allocation, the focus on x_1 implies no loss of generality.

A) $1 - r_2 < x_1 < r_1$

B) $1 - r_2 < r_1 \leq x_1$

C) $x_1 \leq 1 - r_2 < r_1$

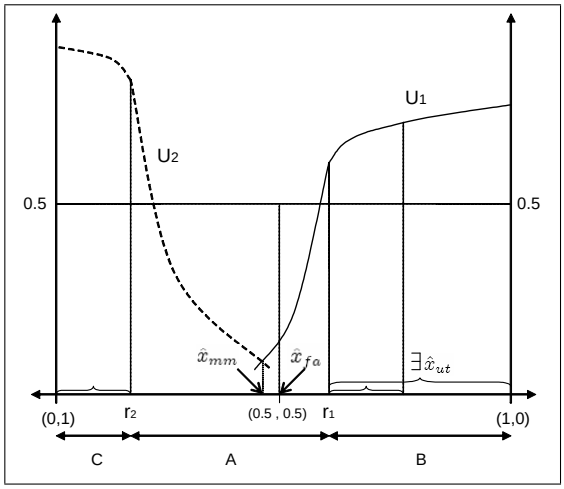

Figure 4: reference dependent preferences and self-serving bias.

As a first result we prove that the optimal utilitarian allocation cannot belong to the intermediate interval A.

Proposition 3 *If claimants have reference dependent preferences and are self-serving biased then* $\hat{x}_{ut} = (\hat{x}_1^{ut}, \hat{x}_2^{ut})$ *is such that* $\hat{x}_1^{ut} \notin (1 - r_2, r_1)$.

Proof. By contradiction. Assume $\hat{x}_1^{ut}$ is such that $\hat{x}_1^{ut} \in (1 - r_2, r_1)$; i.e., $\hat{x}_i^{ut} < r_i$ for both $i \in \{1, 2\}$. By assumption A3 about $\mu(\cdot)$, we know that $\frac{\partial^2 \mu(x_i - r_i)}{\partial x_i^2} \Big|_{x_i = \hat{x}_i^{ut}} > 0$ such that the

functions $u_i(\cdot)$ are convex at $\hat{x}_i^{ut}$. This implies that the function $W_{ut}(u(\cdot))$, being the sum of two convex functions, is also convex, which contradicts the second order necessary condition for the maximization of $W_{ut}(u(\cdot))$. ■

Indeed, any allocation that falls in interval A is dominated in terms of utilitarian welfare by the allocation that matches the reference point of one agent (say agent 1, i.e., $x_1 = r_1$) and leaves the other agent as the residual claimant ($x_2 = 1 - r_1$). In fact, the welfare associated with the allocation $(r_1, 1 - r_1)$ is given by $W_{ut}(u(r_1, 1 - r_1)) = r_1 + \mu(r_1 - r_1) + (1 - r_1) + \mu(1 - r_1 - r_2) = 1 + \mu(1 - r_1 - r_2)$. On the other hand, any allocation $(\tilde{x}_1, 1 - \tilde{x}_1)$ with $\tilde{x}_1 \in (1 - r_2, r_1)$ leads to $W_{ut}(u(\tilde{x}_1, 1 - \tilde{x}_1)) = 1 + \mu(\tilde{x}_1 - r_1) + \mu(1 - \tilde{x}_1 - r_2)$. Given that μ is convex in the negative orthant, we have that $1 + \mu(\tilde{x}_1 - r_1) + \mu(1 - \tilde{x}_1 - r_2) < 1 + \mu(\tilde{x}_1 - r_1 + 1 - \tilde{x}_1 - r_2)$ where the right-hand side simplifies to $1 + \mu(1 - r_1 - r_2)$ such that $W_{ut}(u(\tilde{x}_1, 1 - \tilde{x}_1)) < W_{ut}(u(r_1, 1 - r_1))$.

Therefore, from a utilitarian point of view, it is better to disappoint (a lot) an agent ($x_j = 1 - r_i$) while giving the other what he expects ($x_i = r_i$) rather than to disappoint (a little) both claimants. The natural question is then how to decide who is the agent to disappoint. The perhaps surprising answer (remember that r_1 and r_2 can be different) is that this does not matter, as both possibilities lead to the same welfare. In fact, the welfare associated with $(1 - r_2, r_2)$ is given by $W_{ut}(u(1 - r_2, r_2)) = (1 - r_2) + \mu(1 - r_2 - r_1) + r_2 + \mu(r_2 - r_2) = 1 + \mu(1 - r_2 - r_1)$ which is equivalent to $W_{ut}(u(r_1, 1 - r_1))$. Such an equivalence result remains valid for what concerns parts of the two intervals B and C. The following proposition shows that if a utilitarian maximum falls in the smaller of the two intervals (interval C in Figure 4), this maximum is not unique.

Proposition 4 *If claimants have reference dependent preferences, are self-serving biased and there exists* $\hat{x}_{ut} = (\hat{x}_1^{ut}, \hat{x}_2^{ut})$ *such that* $\hat{x}_1^{ut} \in [r_1, r_1 + \min\{1 - r_1, 1 - r_2\}]$ *then there exists* $\hat{x}'_{ut} = (\hat{x}_1'^{ut}, \hat{x}_2'^{ut})$ *with* $\hat{x}_i'^{ut} \neq \hat{x}_i^{ut}$ *and* $\hat{x}_i'^{ut} \neq \hat{x}_j^{ut}$ *such that* $W_{ut}(u(\hat{x}'_{ut})) = W_{ut}(u(\hat{x}_{ut}))$.

Proof. Assume $\hat{x}_{ut} = (\hat{x}_1^{ut}, \hat{x}_2^{ut})$ is such that $\hat{x}_1^{ut} \in [r_1, r_1 + \min\{1 - r_1, 1 - r_2\}]$. Utilitarian welfare is given by $W_{ut}(u(\hat{x}_{ut})) = 1 + \mu^+(\hat{x}_1^{ut} - r_1) + \mu^-(1 - \hat{x}_1^{ut} - r_2)$ where we use $\mu^+(\cdot)$

(resp. $\mu^-(\cdot)$) to indicate the positive and concave (resp. negative and convex) part of the function $\mu(\cdot)$. Now define $\hat{x}'_{ut} = (\hat{x}'^{ut}_1, \hat{x}'^{ut}_2)$ with $\hat{x}'^{ut}_1 = 1 - \hat{x}^{ut}_1 + (r_1 - r_2)$ such that $\hat{x}'^{ut}_1 \in [\max\{0, r_1 - r_2\}, 1 - r_2]$. Utilitarian welfare is given by $W_{ut}(u(\hat{x}'_{ut})) = 1 + \mu^-(\hat{x}'^{ut}_1 - r_1) + \mu^+(1 - \hat{x}'^{ut}_1 - r_2)$. By substituting $\hat{x}'^{ut}_1$ we get $W_{ut}(u(\hat{x}'_{ut})) = 1 + \mu^-(1 - \hat{x}^{ut}_1 + (r_1 - r_2) - r_1) + \mu^+(1 - (1 - \hat{x}^{ut}_1 + (r_1 - r_2)) - r_2)$ which simplifies to $W_{ut}(u(\hat{x}'_{ut})) = 1 + \mu^-(1 - \hat{x}^{ut}_1 - r_2) + \mu^+(\hat{x}^{ut}_1 - r_1)$. Therefore $W_{ut}(u(\hat{x}'_{ut})) = W_{ut}(u(\hat{x}_{ut}))$. Notice that $\hat{x}'^{ut}_i \neq \hat{x}^{ut}_j$ (i.e., $1 - \hat{x}^{ut}_1 + (r_1 - r_2) \neq 1 - \hat{x}^{ut}_1$) whenever $r_1 \neq r_2$. More generally, for any $\tilde{x}_1 \in [r_1, r_1 + \min\{1 - r_1, 1 - r_2\}]$ there exists an $\tilde{x}'_1 \in [\max\{0, r_1 - r_2\}, 1 - r_2]$ such that $W_{ut}(u(\tilde{x}')) = W_{ut}(u(\tilde{x}))$; i.e., in the two intervals the function $W_{ut}(\cdot)$ is symmetric with respect to $x = \frac{1+(r_1-r_2)}{2}$. ■

The utilitarian criterion can easily lead to two solutions that are very different in terms of actual allocation of the good. In particular one allocation favors agent 1 $\left(\hat{x}^{ut}_1 \geq r_1 > \frac{1}{2}\right)$ while the other favors agent 2 $\left(\hat{x}'^{ut}_1 \leq 1 - r_2 < \frac{1}{2}\right)$. The fact that $\hat{x}'^{ut}_i \neq \hat{x}^{ut}_j$ whenever $r_1 \neq r_2$ implies that if a maximum is reached at, say, $(0.7, 0.3)$ then the other maximum is not identified by $(0.3, 0.7)$; for instance, if $r_1 = 0.6$, $r_2 = 0.8$ and $\hat{x}_{ut} = (0.7, 0.3)$ then $\hat{x}'_{ut} = (0.1, 0.9)$. These two allocations are equivalent in terms of welfare such that a purely utilitarian planner should be indifferent between the two.[12]

Propositions 3 and 4 imply that the only part of the unit interval in which the utilitarian SWF could potentially identify a unique maximum is given by $x_i \in (1 - (r_j - r_i), 1]$ for the i such that $r_i < r_j$. Such a maximum, if it exists, is identified by first- and second-order conditions, or it emerges as a corner solution if the optimal allocation is such that the entire pie has to be given to a single agent.

To conclude the discussion of the utilitarian SWF, notice that the optimal allocation(s) lead to a welfare that is surely smaller than 1. In fact, welfare at $\hat{x}_{ut}$ (where, by proposition 3, $\hat{x}_{ut}$ is such that $\hat{x}^{ut}_i \geq r_i$ for only one $i \in \{1, 2\}$) is given by $W_{ut}(u(\hat{x}_{ut})) = 1 + \mu(\hat{x}_i - r_i) +$

[12] However, notice that whenever $r_1 \neq r_2$, one of the two maxima (in the specific example this would be $\hat{x}_{ut} = (0.7, 0.3)$) leads to lower inequality both in endowments and in utilities with respect to the other. Therefore, a social planner with lexicographic preferences defined over utilitarian welfare and equity would choose the former.

$\mu(1-\hat{x}_i-r_j)$. Because of the assumptions about $\mu(\cdot)$, we have that $|\mu(r_i-\hat{x}_i)| > \mu(\hat{x}_i-r_i)$ and $|\mu(1-\hat{x}_i-r_j)| > |\mu(r_i-\hat{x}_i)|$ because $1-\hat{x}_i-r_j < r_i-\hat{x}_i$ given that $r_i+r_j > 1$. Therefore $W_{ut}(u(\hat{x}_{ut})) < 1$. Utilitarian welfare is obviously larger ($W_{ut} \geq 1$) when the social planner has the possibility to match both claims, i.e., when agents are unbiased as shown in section 4.1. Self-serving bias is welfare detrimental since reference points that are not compatible make unavoidable the fact that someone must be disappointed: if everyone is ambitious, someone will be frustrated.

Consider now what happens if the social planner adopts the maxmin or the fair SWF. As usual, the maxmin criterion selects the allocation for which the utility functions intersect while the fair criterion selects the egalitarian allocation. Both these allocations surely fall in the intermediate interval A. The following special case is particularly striking:

Proposition 5 *If claimants have reference dependent preferences, are self-serving biased and perfectly symmetric then* $\hat{x}_{mm} = \hat{x}_{fa} = \arg\min_x W_{ut}(u(x)) = \left(\frac{1}{2}, \frac{1}{2}\right)$.

Proof. If claimants are symmetric then $r_1 = r_2$ and the only feasible and efficient allocation that equalizes their utility is the egalitarian one. It follows that $\hat{x}_{mm} = \hat{x}_{fa} = \left(\frac{1}{2}, \frac{1}{2}\right)$. Symmetry also implies that $\left(\frac{1}{2}, \frac{1}{2}\right)$ is the unique allocation for which the FOCs of $W_{ut}(u(x))$ are satisfied $\left(\frac{\partial u_1(x_1)}{\partial x_1} = \frac{\partial u_2(x_2)}{\partial x_2}\right)$ and $x_1 + x_2 = 1$ holds. But being in the interval in which both individual functions are convex then $W_{ut}(u(x))$ is also convex such that $\left(\frac{1}{2}, \frac{1}{2}\right)$ identifies the minimum of the utilitarian SWF. ■

Therefore, when claimants are perfectly symmetric, the egalitarian allocation is supported by both the maxmin and the fair SWFs. Indeed, possibly also because of its uniqueness with respect to the utilitarian solution (see Prop. 4) and 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, in the case of reference dependent preferences and self-serving biased agents, 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.

Some of the results presented in this section easily generalize to the case in which the two agents are endowed with heterogeneous gain-loss functions $\mu_i(\cdot)$. In particular, the fair criterion always selects the symmetric allocation while the maxmin criterion keeps selecting an intermediate solution such that $x_i < r_i$ for any $i \in \{1, 2\}$. This can never happen if the social planner adopts the utilitarian approach as $\hat{x}_i^{ut} \geq r_i$ for one $i \in \{1, 2\}$. Which agent will get the more favorable allocation depends on the specific functional forms. In particular, what matters is the relative slope of the $\mu_i(\cdot)$ functions in the domain of gains (better to appreciate more the surplus $x_i - r_i$) as well as in the domain of losses (better to suffer more the loss $r_i - x_i$).

To sum up, the main result of this section is that, when claimants are characterized by reference dependent preferences and they also display a self-serving bias, the optimal allocations selected by the three SWFs do not generally coincide. This is an important difference with respect to the cases of inequity aversion (Section 3) and reference dependence with no self-serving bias (Section 4). It implies a more decisive role for the social planner, as his choice about which welfare criterion to follow leads to very different allocations of the good. Moreover, it suggests the possibility that claimants may behave strategically. We now briefly discuss this issue.

4.2.1 Some strategic considerations

The basic formulation of our allocative problem did not present any strategic aspect: the social planner considers individual preferences, and implements a certain allocation according to the specific welfare criterion that he is following. The claimants have basically no active role in the process. Still, the mere fact that a social welfare function is a function of individuals' 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 a common regularity (namely, the fact that the claims of agents involved in disputes often appear to be exaggerated) is rationalized by a model in which claimants with reference dependent preferences strategically announce their reference point. Such a behavior

also reveals claimants' rational beliefs about the social welfare criterion that the planner will use.

Consider the already familiar case of two claimants having utility functions $u_i = x_i + \mu(x_i - r_i)$ and focus on the maxmin criterion. We saw that a maxmin social planner aims to equalize utilities so that, in the common case of biased individuals $\left(r_i > \frac{1}{2}\right)$, the selected allocation certainly falls into the intermediate interval where $x_i < r_i$ for any $i \in \{1, 2\}$. More precisely, the planner implements the allocation $\hat{x}_{mm} = (\hat{x}_1^{mm}, \hat{x}_2^{mm})$ that solves $\hat{x}_1^{mm} + \mu(\hat{x}_1^{mm} - r_1) = \hat{x}_2^{mm} + \mu(\hat{x}_2^{mm} - r_2)$ subject to $\hat{x}_1^{mm} + \hat{x}_2^{mm} = 1$. We want to study the effects that r_i has on $\hat{x}_i^{mm}$ for any given r_j. Focusing on agent 1, the last expression can be rewritten as $F(\hat{x}_1^{mm}, r_1) = 2\hat{x}_1^{mm} + \mu(\hat{x}_1^{mm} - r_1) - \mu(1 - \hat{x}_1^{mm} - r_2) - 1 = 0$. This is an implicit function that satisfies the assumptions of the implicit-function theorem. In fact, property A2 of the gain-loss function $\mu(\cdot)$ (see page 11) ensures that partial derivatives $\frac{\partial F(\hat{x}_1^{mm}, r_1)}{\partial \hat{x}_1^{mm}}$ and $\frac{\partial F(\hat{x}_1^{mm}, r_1)}{\partial r_1}$ are continuous and that $\frac{\partial F(\hat{x}_1^{mm}, r_1)}{\partial \hat{x}_1^{mm}} \neq 0$ for any $x_1 < r_1$.

By totally differentiating $F(\hat{x}_1^{mm}, r_1)$ one gets

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

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

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

where the numerator of the ratio is negative (by A1) while the denominator is positive; in particular, the second term is positive (again by A1) while the third one is negative given that $\hat{x}_2^{mm}$ decreases as $\hat{x}_1^{mm}$ increases. It follows that $\frac{\partial \hat{x}_1^{mm}}{\partial r_1} > 0$. Therefore an agent, even though he anticipates that he will get $x_i < r_i$, still should purposely inflate his (possibly already biased) claim r_i.[13]

Many are the real-life situations in which conflicting interests have to be settled by some kind of an authority and players have the possibility to ex-ante declare what they are

[13] The Nash equilibrium is such that both players announce $r_i = 1$ and the social planner implements $\hat{x}_{mm} = \left(\frac{1}{2}, \frac{1}{2}\right)$. Such a situation resembles the famous problem of King Solomon who suggested to split a baby in half in front of two women that were both claiming to be his mother.

expecting to get: divorces, reimbursements for damages, bargaining, political negotiations, lobbying. In all of these cases it is common to observe agents asking for extremely high claims. The above analysis shows that such a behavior is rational when the agents believe that the planner is going to implement an intermediate allocation such as the maxmin one where the final allocation is positively anchored to the initial claim.[14] Indeed, as it has already been noted, intermediate allocations appear to be the most common solutions implemented in reality.

4.2.2 A digression on paternalism

The study of the allocation problem in the case of self-serving biased agents with reference dependent preferences raises another interesting issue which, this time, has to do with the behavior of the social planner. More precisely, what should be the role of a benevolent planner in front of biased individuals? Shall the planner behave in a paternalistic way and try to debias the claimants? Or shall he simply maximize welfare, taking agents' biased preferences as given?

In general, a firm principle of traditional welfare analysis is that policies should not be paternalistic, i.e., the social planner/government should not substitute the preferences of the individuals with his own. Indeed, such a principle is hardly criticizable under the assumption that agents are fully rational. Still, recent behavioral contributions show that the alternative assumption of limited rationality is often more realistic, as individuals' preferences are characterized by biases and inconsistencies and their choices are plagued by mistakes. In such a situation the active intervention of a more rational social planner can improve welfare without being too distortive. Camerer (1999, p. 10577) provides a clear example for such an approach:

"*Relaxing rationality assumptions therefore permits reasoned argument about how people can be helped. For example, if people weight the future hyperbolically rather than exponentially,*

[14] This would not be the case with the fair SWF, which selects an intermediate allocation (the symmetric one) that does not depend on r_i. For what concerns utilitarian welfare, the sign of $\frac{\partial \hat{x}_i^{ut}}{\partial r_i}$ cannot be univocally assessed, as we saw (Proposition 4) that the solution is often not unique.

they will impulsively buy goods they will soon regret having bought. A good policy to help those who weight the future hyperbolically is a mandatory "cooling off" period that permits "hot" consumers to renege on purchase decisions for a short period of time, such as 3 days. (Many states have such policies). Cooling-off policies exemplify "conservative paternalism" - they will do much good for people who act impulsively and cause very little harm (an unnecessary 3-day wait) for those who do not act impulsively; thus, even conservatives who resist state intervention should find them appealing".

Indeed, some forms of such a "conservative or asymmetric paternalism" have been recently advocated in many influential papers that study the social welfare implications of behavioral models (see for example Camerer *et al.*, 2003, Gruber and Koszegi, 2004, O'Donoghue and Rabin, 2006). In line with the quotation above, these papers show that lifetime welfare of biased consumers can be larger when their behavior is constrained by some *ad-hoc* policies (for instance, taxes on the consumption of addictive goods). Moreover, unbiased consumers are shown to suffer very little from the introduction of such policies; this kind of government's intervention appears to be welfare improving even if the proportion of biased individuals is small.

These arguments are convincing but do not apply to the context of our simple one-period allocative problem in which agents characterized by reference dependent preferences suffer from a self-serving bias. In fact, in this case, welfare is surely larger when the planner takes agents' biases as given. An alternative scenario would be one of a repeated version of the one shot problem: in such a situation a planner that is interested in long term or average welfare can find it optimal to initially disappoint biased claimants with the goal of moving their reference point from a biased one ($r_i > \frac{1}{N}$) to an unbiased one ($r_i = \frac{1}{N}$). Our analysis showed in fact that, under all three SWF specifications, the welfare associated with the optimal allocation is usually larger in the case of unbiased claimants (see Section 4.1) than in the case of biased ones (see Section 4.2). The time horizon of the social planner is then crucial in order to justify or not the adoption of a paternalistic approach.[15]

[15] As an example, consider the situation of a self-serving biased baby who thinks he deserves to watch

5 Conclusion

We investigated the social welfare implications that some important classes of behavioral preferences have on the classical welfare problem of allocating a scarce resource among a finite number of claimants. In particular we considered the cases of inequity aversion preferences and reference dependent preferences. In the latter case we also studied the impact of agents displaying a self-serving bias when setting their reference points.

The analysis showed that optimal allocations are often remarkably different with respect to the traditional "rational preferences" case. It follows that the optimal policies that a social planner should implement are also different. Moreover, despite having used welfare criteria that are based on the concept of individuals' utility, some of the results of the paper have some strong positive, and not just normative, implications. For instance, we showed that in the case of inequity averse individuals it is usually not necessary for the planner to exactly know agents' utility functions, as the optimal allocation remains the same for a wide range of individuals' parameters as well as for different welfare specifications. In the case of reference dependent preferences, we showed how the presence of biased individuals gives rise to a severe trade-off between efficient versus equitable allocation and how and why this trade-off is usually solved in favor of equitability. In talking about the positive aspects of the analysis, an obvious question is how to decide which are the relevant preferences that have to be assumed in different contexts. The specific problem under study may possibly indicate the solution: a planner who has to divide a resource among a group of friends may safely assume inequity aversion. At the opposite, we mentioned the case of a couple facing a rough divorce as an example where reference dependent preferences and self-serving biases are likely to play a role.

This study could certainly be extended in many ways. For instance, we restricted the analysis to cases of inequity aversion and reference dependent preferences because we think

TV till 10 pm while, given his age and habits, he should actually go to sleep at 9 pm. A parent who cares about the long-run health and balance of her child should override the baby's biases and send him to sleep at 9 pm. However, a baby-sitter who simply wants to minimize the baby's crying may find it optimal to accommodate his request.

these to be the families of preferences more relevant in the context of our allocative problem. Still, the impact of other kinds of behavioral preferences could be analyzed in slightly different problems. For instance, models of reciprocity can matter if claimants have the possibility to reward or punish the planner while hyperbolic discounting can play a role if the game is repeated over time. Additionally, a mixture of various kinds of preferences could be allowed. Finally, at a more general level, another direction to go would be to perform a welfare analysis that is based on ordinal preference relations rather than on cardinal utility functions.

Despite the above-mentioned limitations, we feel that our analysis captures the main ingredients of many real-life allocative problems ranging from litigations to political lobbying, from impasses in bargaining to divorces and that, more generally, it contributes to the recent literature about the public-policy implications of research in behavioral economics.

References

[1] Abreu, D. and A. Sen (1990), "Subgame Perfect Implementation: A Necessary and Almost Sufficient Condition", *Journal of Economic Theory*, Vol. 50, No. 2, pp. 285-299.

[2] Babcock, L., Loewenstein, G., Issacharoff, S. and C. Camerer (1995), "Biased Judgments of Fairness in Bargaining", *American Economic Review*, Vol. 85, No. 5, pp. 1337-1343.

[3] Babcock, L. and G. Loewenstein (1997), "Explaining Bargaining Impasse: The Role of Self-Serving Biases", *Journal of Economic Perspectives*, Vol. 11, No. 1, pp. 109-126.

[4] Bag, P. K. (1996), "Efficient Allocation of a Pie: Divisible Case of King Solomon's Dilemma", *Games and Economic Behavior*, Vol. 12, no. 1, pp. 12-41.

[5] Bentham, J. (1789), *Introduction to the Principles of Morals and Legislation*. London: Athlone.

[6] Bernheim, B. D. and A. Rangel (2007), "Behavioral Public Economics: Welfare and Policy Analysis with Non-Standard Decision Makers", in Diamond, P. and H. Vartiainen (eds.): Behavioral Economics and its Applications, Princeton University Press: Princeton, NY.

[7] Bolton, G. E. and A. Ockenfels (2000), "ERC: a Theory of Equity, Reciprocity and Competition", *American Economic Review*, Vol. 90, No. 1, pp. 166-193.

[8] Bowman, D., Minehart, D. and M. Rabin (1999), "Loss Aversion in a Consumption-Savings Model", *Journal of Economic Behavior and Organization*, Vol. 38, No. 2, pp. 155-178.

[9] Brams, S. J. and A. D. Taylor (1996), *Fair Division: From Cake-Cutting to Dispute Resolutions*, Cambridge University Press, Cambridge, MA.

[10] Brams, S. J., Jones, M. A. and C. Klamler (2006), "Better Ways to Cut a Cake", *Notices of the American Mathematical Society*, Vol. 53, No. 11, pp. 1314-1321.

[11] Camerer, C. (1999), "Behavioral Economics: Reunifying psychology and economics", *Proceedings of the National Academy of Sciences*, Vol. 96, No. 19, pp. 10575-10577.

[12] Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T. and M. Rabin (2003), "Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism"", *University of Pennsylvania Law Review*, Vol. 151, No. 3, pp. 1211-1254

[13] Charness, G. and M. Rabin (2002), "Understanding Social Preferences with Simple Tests", *Quarterly Journal of Economics*, Vol. 117, No. 3, pp. 817-869.

[14] Choi, J., Laibson, D., Madrian, B. and A. Metrick (2003), "Optimal Defaults", *American Economic Review (Papers and Proceedings)*, Vol. 93, No. 2, pp. 180-185.

[15] Dall'Aglio, M. and F. Maccheroni (2007), "Disputed Lands", The Carlo Alberto Notebooks, No. 58.

[16] Dubins, L. E. and E. H. Spanier (1961), "How to Cut a Cake Fairly", *American Mathematical Monthly*, Vol. 68, No. 5, pp. 1-17.

[17] Dufwenberg, M. and G. Kirchsteiger (2004), "A Theory of Sequential Reciprocity", *Games and Economic Behavior*, Vol. 47, No. 2, pp. 268-298.

[18] Fehr, E. and K. Schmidt (1999), "A Theory of Fairness, Competition and Cooperation", *Quarterly Journal of Economics*, Vol. 114, No. 3, pp. 817-868.

[19] Fehr, E. and K. Schmidt (2006), "The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories", forthcoming in *Handbook on the Economics of Giving, Reciprocity and Altruism*, eds: S.-C. Kolm and J. M. Ythier. Elsevier.

[20] Foley, D. (1967), "Resource Allocation and the Public Sector", *Yale Economics Essays*, Vol. 7, pp. 45-98.

[21] Gray, J. and R. Silver (1990), "Opposite Side of the Same Coin: Farmer Spouses' Divergent Perspectives in Coping with their Divorce", *Journal of Personality and Social Psychology*, Vol. 59, No. 6, pp. 1180-1191.

[22] Gruber, J. and B. Koszegi (2001), "Is Addiction "Rational"? Theory and Evidence", *Quarterly Journal of Economics*, Vol. 116, No. 4, pp. 1261-1303.

[23] Gruber, J. and B. Koszegi (2004), "Tax Incidence When Individuals are Time Inconsistent : The Case of Cigarette Excise Taxes", *Journal of Public Economics*, Vol. 88, No. 9-10, pp. 1959-1987.

[24] Kahneman, D. and A. Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, Vol. 47, No. 2, pp. 263-291.

[25] Koszegi, B. and M. Rabin (2006), "A Model of Reference-Dependent Preferences", *Quarterly Journal of Economics*, Vol. 121, No. 4, pp. 1133-1165.

[26] Maccheroni, F. and M. Marinacci (2003), "How to Cut a Pizza Fairly: Fair Division with Decreasing Marginal Evaluations", *Social Choice and Welfare*, Vol. 20, No. 3, pp. 457-465.

[27] Mas-Colell, A., Whinston, M. D. and J. R. Green (1995), *Microeconomic Theory*, Oxford University Press, New York, NY.

[28] Moulin, H. J. (2003), *Fair Division and Collective Welfare*, The MIT Press, Cambridge, MA.

[29] O'Donoghue, T. and M. Rabin (2001), "Choice and Procrastination", *Quarterly Journal of Economics*, Vol. 116, No. 1, pp. 121-160.

[30] O'Donoghue, T. and M. Rabin (2006), "Optimal Sin Taxes", *Journal of Public Economics*, Vol. 90, No. 10-11, pp. 1825-1849.

[31] Rabin, M. (1993), "Incorporating Fairness into Game Theory and Economics", *American Economic Review*, Vol. 83, No. 5, pp. 1281-1302.

[32] Rawls, J. (1971), *A Theory of Justice*. Cambridge, Mass: Harvard University Press.

[33] Ross, M. and F. Sicoly (1981), "Egocentric Biases in Availability and Attribution", *Journal of Personality and Social Psychology*, Vol. 37, No. 3, pp. 322-336.

[34] Schriber, J., Larwood, L. and J. Peterson (1985), "Bias in the Attribution of Marital Conflict", *Journal of Marriage and the Family*, Vol. 47, No. 3, pp. 717-721.

[35] Schuutz, A. (1999), "It was your Fault! Self-Serving Biases in Autobiographical Accounts of Conflicts in Married Couples", *Journal of Social and Personal Relationships*, Vol. 16, No. 2, pp. 193-208.

[36] Sen, A. (1977), "On Weights and Measures: Informational Constraints in Social Welfare Analysis", *Econometrica*, Vol. 45, No. 7, pp. 1539-1572.

[37] Sen, A. (1979), "Personal Utilities and Public Judgements: Or What's Wrong with Welfare Economics?", *Economic Journal*, Vol. 89, No. 355, pp. 537-558.

[38] Steinhaus, H. (1948), "The Problem of Fair Division", *Econometrica*, Vol. 16, No. 1, pp.101-104.

[39] Sunstein, C. and R. H. Thaler (2003), "Libertarian Paternalism", *American Economic Review (Papers and Proceedings)*, Vol. 93, No. 2, pp. 175-179.

[40] Svenson, O. (1981), "Are We All Less Risky and More Skillful than our Fellow Drivers?", *Acta Psychologica*, Vol. 47, No. 2, pp. 143-148.

[41] Tadenuma, K. and W. Thompson (1995), "Games of Fair Division", *Games and Economic Behavior*, Vol. 9, No. 2, pp. 191-204.

[42] Varian, H. (1974), "Equity, Envy and Efficiency", *Journal of Economic Theory*, Vol. 9, No. 1, pp. 63-91.