Some Social Welfare Implications of Behavioral Preferences

Andrea Gallice*

Abstract

We reconsider the standard welfare problem of allocating a scarce resource among a number of claimants in light of some recent behavioral contributions. In particular we study the alternative scenarios of the claimants being characterized by inequity averse preferences, reference dependent preferences and self serving biases. For each case we compute the social planner's optimal allocations identified by the utilitarian, the maxmin and the fair criteria. Results are often at odds with respect to the traditional "neoclassical preferences" case. We discuss the policy implications.

Keywords: social welfare, optimal allocation, inequity aversion, reference dependence, self serving bias.

JEL classification: D00, D61.

*University of Munich and ICER, Turin. Contact: andrea.gallice@eui.eu. I would like to thank Klaus Schmidt for helpful discussion as well as seminar participants at the University of Munich, Florence (Experimental Economics Workshop) and Mannheim (European Symposium on Economics and Psychology) for useful comments.

1 Introduction

In the last couple of decades behavioral economics successfully challenged and enriched traditional neoclassical analysis of individuals' behavior. This success is testified by a massive number of empirical and experimental studies that show how (some of) these behavioral contributions are able to rationalize regularities otherwise previously labelled as puzzles. Nowadays terms like "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 labelled by these terms do actually affect the behavior of many people in many different contexts.

Taking for granted their importance at an individual level, one should then consider the effects that these same phenomena may have on the collectivity as a whole. 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. Some recent papers already proceed in this direction. For instance O'Donoghue and Rabin (2006) study the issue of optimal taxation on sin goods (like unhealthy food) when some consumers suffer from self-control problems as captured by time preferences based on quasi-hyperbolic discounting. Or yet, in a similar framework, Gruber and Koszegi (2001 and 2004) study how taxes can counteract individuals' addiction to cigarettes. And some other papers (O'Donoghue and Rabin, 2001, Sunstein and Thaler, 2003, Choi et al., 2003) also consider various forms of "paternalistic" taxation on some kind of goods.[1]

Still sharing a similar approach, this paper does not study the issue of optimal taxation but it focuses instead on a different problem of welfare economics. The problem is the one of a (benevolent) social planner who has to allocate a scarce resource among a number of claimants. Many are the possible examples for such a situation: a parent who has to divide a chocolate bar among her children, a boss who wants to share a monetary bonus among

[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.

his subordinates, a judge called to decide upon 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 such a problem under the assumption of the claimants being characterized by various forms of behavioral preferences. In particular three are the scenarios that we consider: the case of the claimants having inequity averse preferences (à la Fehr and Schmidt, 1999), the case with reference dependent preferences (as formalized in Koszegi and Rabin, 2006) and the case where reference dependent preferences are combined with a self serving bias (such a bias is 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.

The analysis of this simple allocative problem in a context that differs from the rational paradigm leads to many remarkable results. The more general lesson is that behavioral preferences can have important welfare implications. The paper shows in fact that optimal solutions do differ with respect to the standard "neoclassical preferences" case. This implies that different are also the policies that a social planner should implement. In the course of the analysis we will also be able to state more specific results that answer to the following questions: if agents are inequity averse shall a social planner worry about how to define social welfare? And if individuals are self serving biased and a first best solution is not achievable, is it more welfare enhancing to disappoint (a little) all the claimants or is it better to please some of them and disappoint (a lot) the remaining ones? And, if this is the case, how should the planner choose who are the individuals to privilege?

The remainder of the paper is organized as follows: Section 2 formalizes the problem and it presents its 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 then moving to the case where such a bias exists. Finally Section 5 concludes.

2 The problem and its standard solution

A social planner must allocate a homogeneous and perfectly divisible good (whose amount is normalized 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 social 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 agent i has a utility function $u_i(x)$. Notice that this formulation allows individual utility to depend not only on x_i (as it happens with neoclassical preferences) but possibly also on the other components of the vector x (as it happens with some kinds of behavioral preferences). Still in general it will be the case that $u_i'(x_i) > 0$ for any $x_i \in [0, 1]$ such that claimants do not have any feasible satiation point.

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 the social planner not to be biased towards any particular claimant and therefore we only consider symmetric SWF that give equal weight to all the agents.

More precisely we consider three specific welfare functions. The first one is the utilitarian SWF. This function, which has a very long tradition in welfare economics (starting with Bentham, 1789), prescribes the social planner to implement the allocation that maximizes the sum of individual utilities.

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

The second function is the equally well known maxmin (or Rawlsian from Rawls, 1971)

SWF. A social planner who adopts the maxmin SWF wants to maximize the welfare of the worst-off individual.[2]

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

Finally we also consider what we call the fair SWF. This is a function that selects the "fair" 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 $j \neq i$).

- *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. It follows that, in general, they select different optimal allocations. Still, in line with the benevolent nature of the social planner, a common feature of these allocations is that they all satisfy the Pareto principle.[3] More precisely the optimal allocations $\hat{x}_{ut}$, $\hat{x}_{mm}$ and $\hat{x}_{fa}$ are such that $\sum_i \hat{x}_i = 1$. The first two SWF automatically satisfy the Pareto principle given that agents' utility functions are 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 SWF 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,

[2] The utilitarian and the maxmin SWF 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.

[3] 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 such that the Pareto principle remains a valid objective a social planner should pursue.

increasing and concave utility functions.[4] Formally this implies that $u_i(x) = u_i(x_i)$ with $u'_i(x_i) > 0$ and $u''_i(x_i) < 0$ for any i. With utility functions of this kind the utilitarian SWF selects $\hat{x}_{ut} = (\hat{x}_1, ..., \hat{x}_N)$ with $u'_i(\hat{x}_i) \equiv k$ for any i. In fact the function W_{ut} is concave (it is the sum of N concave functions) and it is maximized by the allocation that equalizes agents' marginal utilities. If, on the other hand, the social planner adopts the maxmin SWF then the optimal allocation is the one that equalizes individuals' actual utilities, i.e. $\hat{x}_{mm} = (\hat{x}_1, ..., \hat{x}_N)$ such that $u_i(\hat{x}_i) \equiv \gamma$ for any i. Finally, given that claimants' preferences are monotonic, the allocation selected by the fair SWF is the one that equalizes individuals' endowments: $\hat{x}_{fa} = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$.

Another benchmark case considers linear utility functions such that $u'_i(x_i) > 0$ and $u''_i(x_i) = 0$. This 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 such as to make the marginal decreases in utility negligible.[5] Optimal allocations with linear utility functions are then given by $\hat{x}_{ut} = (\hat{x}_1, ..., \hat{x}_N)$ with $\hat{x}_i = 1$ for the i (assumed to be unique) such that $u'_i > u'_j$ for any j, $\hat{x}_{mm} = (\hat{x}_1, ..., \hat{x}_N)$ such that $u_i(\hat{x}_i) \equiv \gamma$ 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 cases with two claimants. In each diagram the utility function of agent 1 is as usual displayed from left to right. The utility of agent 2 must be read from right to left. The length of the horizontal axis is fixed to 1, i.e. the total amount of the good the planner has to distribute.

[4] In particular the concavity requirement has always been a key ingredient of classical microeconomic theory. On one hand this requirement captures the standard assumption of diminishing marginal utility. On the other hand, by making first order conditions sufficient, it simplifies the maximization problem.

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

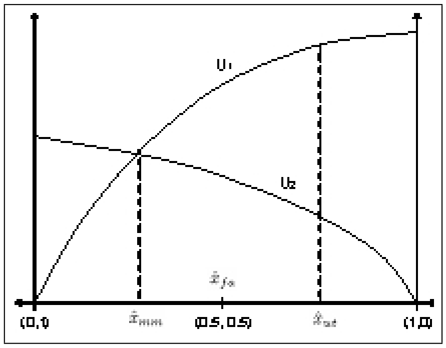

Figure 1.a: concave ut. functions

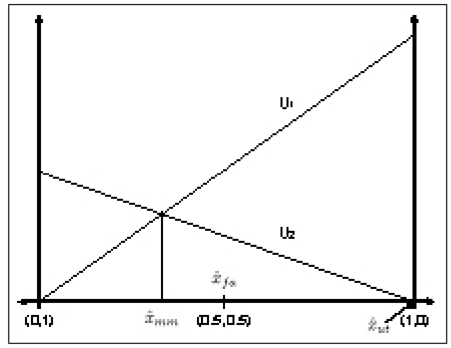

Figure 1.b: linear ut. functions

Notice that, in general, the utilitarian, the maxmin and the fair SWF lead to different allocations under both kinds of utility functions. Indeed, with concave utility functions, the three welfare functions select the same allocation (namely the egalitarian one such that $\hat{x}_i = \frac{1}{N}$ for any i) only when all the agents are perfectly symmetric. The same happens with linear preferences even though in this case a utilitarian social planner is actually indifferent among all the possible allocations. 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 the others (in some appropriate reference group) get. 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 does not always need to hold. In fact agents' utility is often affected by comparisons between what they get versus what the others get. In the last few years various papers modelled these kinds of "other regarding preferences".[6]

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

In this section we consider the social welfare implications of inequity aversion preferences in the context of our simple allocation problem. Prominent examples of inequity aversion models 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 like envy (the agent gets less than others) and guilt / embarrassment (he gets more). In the problem under study these issues seem to be very likely to affect an individual's ex-post judgement (and thus his utility) of the allocation implemented by the social planner.[7]

More precisely we adopt the widely used specification introduced by Fehr and Schmidt (1999) such that 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 (they get less than others) than from advantageous inequality (they get more than others). The second one rules out the existence of individuals that would be so much inequity averse such as to be willing to heavily harm themselves for the sake of equity. Notice that the assumptions allow α_i and β_i to be equal to 0 such that some individuals could still have (linear) purely selfish preferences.

Given that issues of equity are already embedded into individuals' preferences, the fact that, in the context of our problem, optimal allocations will be strongly egalitarian is not

[7] On the other side another important family of other regarding preferences, namely those captured by models of intention based reciprocity (Rabin, 1993, Dufwenberg and Kirchsteiger, 2004) do not seem to apply in our framework. In fact our agents simply have to accept the allocation decided by the social planner and cannot react to it and reciprocate.

surprising. Indeed all the three SWF select the symmetric allocation such that $\hat{x}_{ut} = \hat{x}_{mm} = \hat{x}_{fa} = \left(\frac{1}{N}, ..., \frac{1}{N}\right)$ and $u_i = \frac{1}{N}$ 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. Indeed a single claimant i being "strictly" inequity averse (i.e. with $\beta_i > 0$) is sufficient to make the utilitarian and the maxmin optimal allocations collapse into the fair egalitarian allocation. Proposition 1 formalizes and proves this claim.

Proposition 1 *If claimants have inequity averse preferences and there is at least an agent with $\beta_i > 0$ then the utilitarian, the maxmin and the fair SWF all select the egalitarian allocation, no matter the parameters α_i and β_i.*

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

Considering the utilitarian SWF the social welfare would now be:

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

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

Proposition 1 indicates that the task of a benevolent 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. And indeed he does not even have to care about how to solve the trade off between total utility 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.

As it has been done in the previous section, 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 $0.5 < \beta_2 < 1$.

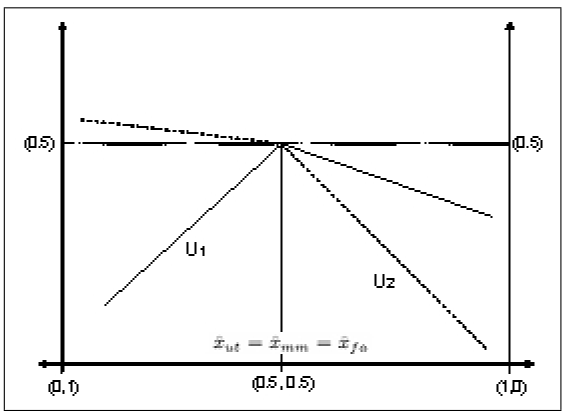

Figure 2: inequity averse preferences.

4 Reference dependent preferences

Another important family of behavioral preferences is labelled under the name of reference dependent preferences. With respect to the standard neoclassical formulation, these preferences explicitly acknowledge the fact that individuals' 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 in Bowman et al. (1999) and in Koszegi and Rabin (2006). In the context of our allocation problem, such a utility function can be expressed as $u_i = m(x_i) + \mu(x_i - r_i)$ where, as before, x_i indicates the amount of the good that the social planner allocates 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. The function $\mu(x_i - r_i)$ is a "gain-loss function", i.e. it captures the additional effects of perceived gains and losses defined with respect to the individual's ex ante reference point r_i. Notice that the function $\mu(\cdot)$ does not have the subindex i. In fact, for the sake of simplicity, we assume $\mu(\cdot)$ to be the same for all the individuals.

In accordance with the properties discussed in Kahneman and Tversky (1979) and closely following Koszegi and Rabin (2006), the function $\mu(\cdot)$ is assumed to satisfy the following requirements:

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

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

A3: $\mu''(x_i - r_i) \geq 0$ if $x_i < r_i$ and $\mu''(x_i - r_i) \leq 0$ 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: $\lim_{x_i \to r_i^-} \mu'(x_i - r_i) / \lim_{x_i \to r_i^+} \mu'(x_i - r_i) \equiv \lambda > 1$.

A brief comment 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 (i.e. losses) and concave for values of x_i that are above r_i (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 μ

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 (domain of losses) than from the right (domain of gains). Taken together these last two assumptions capture the loss aversion phenomenon.

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 simpler as well as more comparable with the inequity averse preferences studied in the previous section (where again the "core" part of the utility function was given by x_i). Second, and most importantly, the linear form of $m(x_i)$ implies that the properties of the function $\mu(\cdot)$ directly translate into equivalent properties of the utility function $u_i(\cdot)$.[8]

Therefore 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 yet 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, namely the self serving bias. Self serving bias is a pervasive phenomenon that influences people's behavior in various ways: agents tend to over estimate their merits, to favorably acquire and interpret informations, to give biased judgements about what is fair and what is not, to inflate their claims. Self serving bias can have important economic implications. For instance it is considered as one of the main causes of costly impasses in bargaining and negotiations (see for instance Babcock et al., 1995 and Babcock and Loewenstein, 1997). Even if the importance of such a bias is widely acknowledged in the literature, still there have a been only a few attempts to properly formalize it and study its implications in an analytical way (....).

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

In the context of our allocation problem self serving bias is likely to affect the way agents set their reference points. In particular, 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. In fact the expectations of the former agent are inflated by his own (biased) perception about how much he deserves.

Consider first the case of an unbiased agent. In this case the claimant should set his reference points as if he was "behind the veil of ignorance" (for instance before knowing the kind of relation he has with the other claimants or the nature of the good distributed by the planner). Given that each agent knows that there are N claimants (including himself) then an unbiased agent expects to get a fair portion of the good and thus he sets $r_i = \frac{1}{N}$. This implies that reference points are mutually compatible whenever all the agents are unbiased ($\sum_i r_i = \sum_i x_i = 1$). At the opposite an agent who is self serving biased will set $r_i > \frac{1}{N}$. It follows that, whenever some agents have such a bias, then reference points are no more compatible ($\sum_i r_i > \sum_i x_i = 1$).[9]

In the analysis that follows we differentiate between two extreme cases. First we study the allocation problem in the situation of no claimant being biased. Then we analyze the same problem when claimants do display a self serving bias. 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 that cannot find an agreement about how to share their property because both the persons claim more than half of it. Or similarly the case of heirs fighting over how to divide a bequest. Or yet 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 specific allocation. Therefore the claimants have to ask some external actor (a judge, an authority, in our case the planner) to solve their dispute.

[9] Notice that an individual can set $r_i > \frac{1}{N}$ without being biased if he actually deserves more. But if all the agents set $r_i > \frac{1}{N}$ then at least some individual is surely biased.

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, i.e. they set $r_i = \frac{1}{N}$. Their utility function is given by $u_i = x_i + \mu(x_i - \frac{1}{N})$. The solution to the planner's problem is then captured by the following proposition:

Proposition 2 *If the claimants have reference dependent preferences and no self serving bias then the utilitarian, the maxmin and the fair SWF all select the egalitarian allocation.*

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 imagine to take ϵ away from player j and give it to agent k. Call this new allocation $\tilde{x}$. According to the utilitarian SWF, welfare would now be $W_{ut}\left(u(\tilde{x})\right) = (N-2)\frac{1}{N} + \left(\frac{1}{N} - \epsilon + \mu(-\epsilon)\right) + \left(\frac{1}{N} + \epsilon + \mu(\epsilon)\right)$ which simplifies to $W_{ut}\left(u(\tilde{x})\right) = 1 + \mu(-\epsilon) + \mu(\epsilon)$. The relation $W_{ut}\left(u(\tilde{x})\right) < W_{ut}\left(u\left(\hat{x}\right)\right) = 1$ surely holds given that $\mu(-\epsilon) < 0$, $\mu(\epsilon) > 0$ (by A1) and $|\mu(-\epsilon)| > \mu(\epsilon)$ (by A5). Under the maxmin SWF, welfare at $\tilde{x}$ is given by $W_{mm}\left(u\left(\tilde{x}\right)\right) = u_j(\tilde{x}) = \frac{1}{N} - \epsilon + \mu(-\epsilon)$ such that $W_{mm}\left(u\left(\tilde{x}\right)\right) < W_{mm}\left(u\left(\hat{x}\right)\right) = \frac{1}{N}$ given that $\mu(-\epsilon) < 0$ (by A1). With the fair SWF $W_{fa}\left(u\left(\tilde{x}\right)\right) < W_{fa}\left(u\left(\hat{x}\right)\right) = 1$ because $u_j\left(\tilde{x}_j\right) < u_j\left(\tilde{x}_k\right)$ and thus $\tilde{x}$ is not envy free. Therefore social welfare is maximized at the egalitarian allocation under all the three SWF specifications. ■

The situation may thus seem analogous to the case of individuals endowed with inequity averse preferences: the optimal allocation is the egalitarian one no matter which SWF the social planner may use. Still notice that, with reference dependent preferences, the utilitarian SWF chooses the egalitarian outcome despite neither the planner nor the claimants having any explicit preference for equity. It is true that claimants are implicitly stating a preference for the egalitarian outcome by setting the reference point as if they were behind the veil of ignorance. But then any individual i is actually indifferent if the implemented allocation is egalitarian or not as far as he gets $x_i = \frac{1}{N}$. With respect to the inequity aversion case, agents here do not compare what they get with what the others get but with what they were expecting to get.

Moreover 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 $u_i(\cdot)$. Without this assumption 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 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. But if agents have very different gain-loss functions the utilitarian SWF may select asymmetric allocations.[10]

Figure 3 depicts the case with two claimants having the same $\mu(\cdot)$. In line with Proposition 2 all the three social welfare functions select the same allocation, namely the egalitarian one.

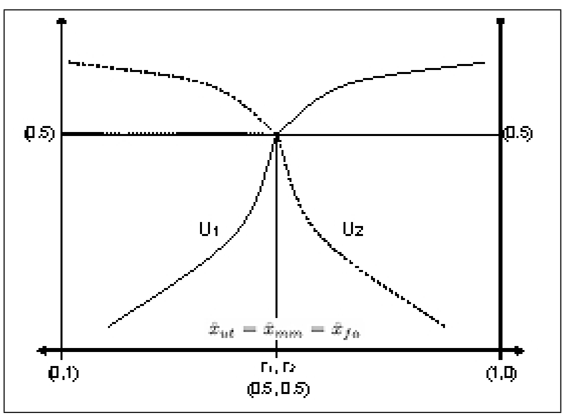

Figure 3: reference dependent preferences and no self serving bias.

[10] A necessary condition for this to happen is that there exists at least two agents such that $\lim_{x_j \to r_j^+} \mu_j'(x_j - r_j) > \lim_{x_k \to r_k^-} \mu_k'(x_k - r_k)$.

4.2 The case of agents with self serving bias

In this section we consider the social welfare implications of claimants having reference dependent preferences and being self serving biased. According to the definition introduced before self serving biased individuals set a reference point that is higher than the one they would have set behind the veil of ignorance. More formally the utility function of biased agents is given by $u_i = x_i + \mu(x_i - r_i)$ where $r_i > \frac{1}{N}$. In what follows, for the sake of tractability, we restrict the analysis to the common case of the social planner having to allocate the good among just two claimants.

First consider the situation of the social planner using the utilitarian SWF. Given the assumption of self serving bias $(r_1 + r_2 > 1)$ and the Pareto constraint $(x_1 + x_2 = 1)$ it follows that the allocation $x = (r_1, r_2)$ is not feasible and the planner has to disappoint at least one of the claimants. Therefore there will be at least an agent who receives less than his reference point such that $\mu(x_i - r_i) < 0$. Most importantly, being in the domain of losses, the function $\mu(\cdot)$ is convex and thus also the utility $u_i(\cdot)$ is convex. This implies that the social welfare function W_{ut} is not guaranteed to be concave such that first order conditions do not necessarily identify a maximum.

More specifically assume that first order conditions are satisfied by the allocation $(\dot{x}_1, \dot{x}_2)_n$, i.e. $\frac{\partial W_{ut}}{\partial x_1} (\dot{x}_1, \dot{x}_2)_n = 0$. The subscript $n \in \{1, 2, 3\}$ refers to the fact that $(\dot{x}_1, \dot{x}_2)_n$ can fall in one of the following intervals (we focus on $\dot{x}_1$ but this implies no loss of generality given that $\dot{x}_1 + \dot{x}_2 = 1$):

1) $1 - r_2 < \dot{x}_1 < r_1$

2) $1 - r_2 < r_1 < \dot{x}_1$

3) $\dot{x}_1 < 1 - r_2 < r_1$

In the first case both players receive less than their reference point. In this interval both utility functions are convex such that the welfare function W_{ut} is also convex. Therefore

$(\dot{x}_1, \dot{x}_2)_1$ identifies a minimum, and not a maximum, of W_{ut}. Given that $(\dot{x}_1, \dot{x}_2)_1$ is the minimizer of W_{ut}, it follows that utilitarian welfare is increasing moving away from it. Proposition 3 shows that social welfare is indeed higher if the social planner satisfies an agent (say agent 1 giving him $x_1 = r_1$) and let the residual part of the pie to the other ($x_2 = 1 - r_1$).

Proposition 3 *If* $1 - r_2 < \dot{x}_1 < r_1$ *then* $W_{ut}\,(\dot{x}_1, \dot{x}_2)_1 < W_{ut}(r_{1,}1 - r_1)$.

Proof. The social welfare associated with the allocation $(r_{1,}1 - r_1)$ is given by $W_{ut}(r_{1,}1 - r_1) = r_1 + \mu(r_1 - r_1) + (1 - r_1) + \mu(1 - r_1 - r_2)$ which simplifies to $W_{ut}(r_{1,}1 - r_1) = 1 + \mu(1 - r_1 - r_2)$. Under the $(\dot{x}_1, \dot{x}_2)_1$ allocation welfare is $W_{ut}\,(\dot{x}_1, \dot{x}_2)_1 = 1 + \mu(\dot{x}_1 - r_1) + \mu(1 - \dot{x}_1 - r_2)$. Given the convexity of μ, the following relation holds: $1 + \mu(\dot{x}_1 - r_1) + \mu(1 - \dot{x}_1 - r_2) < 1 + \mu(\dot{x}_1 - r_1 + 1 - \dot{x}_1 - r_2)$ where the right hand side simplifies to $1 + \mu(1 - r_1 - r_2)$, i.e. the welfare reached under $(r_1, 1 - r_1)$. Therefore $W_{ut}\,(\dot{x}_1, \dot{x}_2)_1 < W_{ut}(r_{1,}1 - r_1)$. ■

Proposition 3 states that, from a utilitarian point of view, it is better to disappoint (a lot) an agent and let the other indifferent rather than to disappoint (a little) both of them. The natural question is then how to decide who is the claimant to disappoint and who is the one to let indifferent. The perhaps surprising answer (remember that r_1 and r_2 can be different) is that both possibilities lead to the same welfare.

Proposition 4 $W_{ut}(r_{1,}1 - r_1) = W_{ut}(1 - r_2, r_2)$.

Proof. The welfare associated with $(r_1, 1 - r_1)$ is $W_{ut}\,(r_1, 1 - r_1) = 1 + \mu(1 - r_1 - r_2)$ (see Prop. 3). The welfare associated with $(1 - r_2, r_2)$ is given by $W_{ut}(1 - r_{2,}r_2) = (1 - r_2) + \mu(1 - r_2 - r_1) + r_2 + \mu(r_2 - r_2) = 1 + \mu(1 - r_2 - r_1)$. Therefore $W_{ut}(r_{1,}1 - r_1) = W_{ut}(1 - r_2, r_2)$. ■

In other words a purely utilitarian social planner is indifferent about to whom allocate r_i and therefore he can freely randomize.[11]

More in general notice that the welfare associated with an allocation at a reference point is surely smaller than 1. In fact $W_{ut}(r_1, 1-r_1) = W_{ut}(1-r_2, r_2) = 1 + \mu(1-r_1-r_2) < 1$ given that $r_1 + r_2 > 1$. Utilitarian welfare is obviously larger ($W_{ut} = 1$) when the social planner can match both claims, i.e. when agents are not self serving biased.

Proposition 3 showed that the utilitarian SWF cannot reach a maximum in the first interval of interest ($1 - r_2 < \dot{x}_1 < r_1$). We now move to the analysis of the second and the third intervals. Given the similarity of these two cases (an agent gets more than his reference point while the other gets less) we only consider Case 2, i.e. we assume first order conditions identify $(\dot{x}_1, \dot{x}_2)_2$ such that $1 - r_2 < r_1 < \dot{x}_1$. In such a situation the utilitarian SWF is the sum of a concave and a convex function and therefore its curvature cannot be univocally assessed as it will depend on the specific functional form of the individuals' utility functions. More precisely utilitarian welfare is given by $W_{ut}\left(\dot{x}_1, \dot{x}_2\right)_2 = 1 + \mu(\tilde{x}_1 - r_1) + \mu(1 - \tilde{x}_1 - r_2)$ where now the second term is positive and the third one is negative. Welfare is again strictly smaller than 1. In fact $|\mu(r_1 - \dot{x}_1)| > \mu(\dot{x}_1 - r_1)$ by the assumptions on the gain-loss function $\mu(\cdot)$, and $|\mu(1 - \dot{x}_1 - r_2)| > |\mu(r_1 - \dot{x}_1)|$ because $1 - \dot{x}_1 - r_2 < r_1 - \dot{x}_1$ since $r_1 + r_2 > 1$ given that both agents are self serving biased.

Still the proper comparison has to be made between $W_{ut}\left(\dot{x}_1, \dot{x}_2\right)_2 = 1 + \mu(\dot{x}_1 - r_1) + \mu(1 - \dot{x}_1 - r_2)$ and $W_{ut}(r_1, 1 - r_1) = 1 + \mu(1 - r_1 - r_2)$, the welfare realized at a reference point. No definitive conclusions can be stated about this comparison. In fact it is possible that $\mu(\dot{x}_1 - r_1) > -\mu(1 - \dot{x}_1 - r_2) + \mu(1 - r_1 - r_2)$ in which case the allocation $(\dot{x}_1, \dot{x}_2)_2$ is welfare maximizing. But if the last inequality does not hold, such that the extra utility agent 1 enjoys from getting more than what he was expecting is below the threshold level,

[11] Still notice that, assuming wlog that $r_1 > r_2$, then the allocation $(1 - r_2, r_2)$ leads to lower inequality, i.e. the difference between $u_1(\cdot)$ and $u_2(\cdot)$ is smaller than the one associated with the allocation $(r_1, 1 - r_1)$. Therefore a social planner with lexicographic preferences defined over utilitarian welfare and then equity will choose $(1 - r_2, r_2)$.

then a utilitarian social planner must implement the reference point allocation $(r_1, 1 - r_1)$.

Consider now the implications of the maxmin and the fair SWF. As usual the first criterion selects the allocation for which the utility functions intersect while the fair SWF selects the egalitarian allocation. Notice that both these allocations surely fall in the intermediate interval (case 1 described above) where both players get less than their reference point. The following special case is particularly striking:

Proposition 5 *If* $r_1 = r_2$ *then* $\max W_{mm} = \max W_{fa} = \min W_{ut} = \left(\frac{1}{2}, \frac{1}{2}\right)$.

Proof. If $r_1 = r_2$ then the two claimants are symmetric and therefore their utility is equalized only by the symmetric allocation. 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 FOC of W_{ut} are satisfied $\left(u_i'(x_i) = u_j'(x_j)\right)$ and $x_i + x_j = 1$. But being in the interval in which both functions are convex then W_{ut} is also convex and therefore $\left(\frac{1}{2}, \frac{1}{2}\right)$ identifies the minimum of the utilitarian SWF. ■

Proposition 5 shows that the maxmin and the fair social welfare functions select the worst possible outcome from a utilitarian point of view. Therefore the preferences of the social planner for the equitable and fair allocation has to be particularly strong to justify this selection.

Notice also that in the more general case for which $r_1 \neq r_2$, the size of the pie that agent i will receive according to a maxmin SWF is increasing in r_i (holding r_j constant). This can provide a rationale to explain why claimants, anticipating that the judge will choose an intermediate allocation, declare to expect a very large share of the pie. In other words players can strategically announce a reference point which is higher than their real reference point.

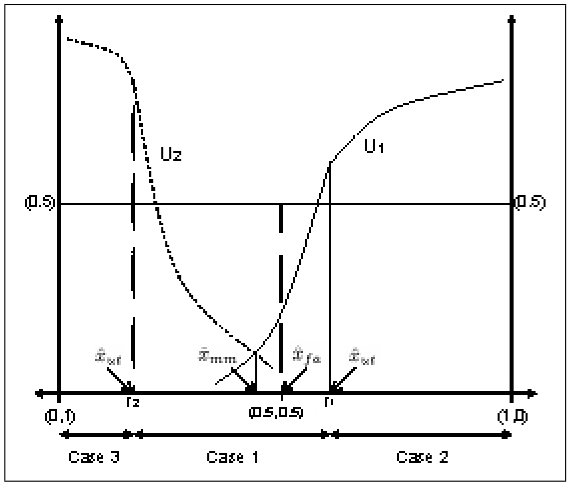

Figure 4: reference dependent preferences and

self serving bias.

5 Conclusion

The paper 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. The analysis showed that the optimal allocations selected by some standard social welfare functions can be remarkably different with respect to the traditional "neoclassical preferences" case. An open question is then 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. And also, on a more general level, which is the role of a benevolent social planner in front of biased individuals? Shall the planner behave in a paternalistic way and try to de-bias the claimants (and thus disappoint them, at least in the short run)? Or shall he focus on instant gratification, maximize short term welfare and possibly accept and please agents' biases? Again different situations seem to suggest different approaches. A parent who has the long run objective of educating her child should in general contrast the baby's biases. But a baby sitter who simply wants to minimize the baby's evening amount of crying may find less costly to accomodate his requests.

References

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

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

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

[9] 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.

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

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

[12] 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.

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

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

[15] 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.

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

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

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

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

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

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

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

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

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

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