NEGATIVE EXTERNALITIES AND SEN'S LIBERALISM THEOREM

DONALD G. SAARI AND ANNE PETRON

Abstract. Sen's seminal, negative theorem about minimal liberalism has had a profound effect on economics, philosophy, and the social sciences. To address concerns raised by his result, we show how Sen's assumptions *must* be modified to obtain positive conclusions; e.g., one resolution allows an agent to be decisive only if his choice does not impose "strong negative externalities" on others. We also uncover an alternative interpretation of Sen's societal cycles: beyond describing the rights of individuals to choose, the cycles identify when these choices impose difficulties on others. Other ways to address Sen's difficulties come from game theory.

1. Introduction

In 1859, John Stuart Mill succinctly articulated the issue with his declaration, "There is a circle around every human being which no government, be it of one, of a few, or of the many, ought to be permitted to overstep... [T]hat there is, or ought to be, some space in human existence thus entrenched around... no one who professes the smallest regard to human freedom or dignity will call in question." Mill's belief underscores a basic tenet of liberalism, which asserts that certain issues and choices naturally belong within an individual's sphere of influence. Restated in terms of the structure of society, the right of individuals to decide constitutes a natural "decentralization" of a portion of societal processes. The admirable intent of this timeless objective is generally assumed to be true, so it is understandable why Sen's "Minimal Liberalism Theorem" (Sen 1970a, b), which asserts that no such decentralization exists, has generated considerable concern. More precisely, Sen proved that no procedure allows individuals to take these actions while protecting society against the discord caused by decision cycles. (For links between economics and philosophy that are created by Sen's result, see Sen (1987), Broome (1991), Hausman and McPherson (1996), and Kolm (1996).)

Yet, seemingly in direct conflict with Sen's result, individuals do make personal decisions on a daily basis, and these decisions need not cause societal problems. This reality suggests that something deeper must occur. To address this puzzle, we examine the structure of Sen's result to determine not what *might* be done, but what *must* be done to circumvent the perplexing difficulties Sen raised. Adding support to what we discover, from a different methodological perspective, is the theory of decentralization and mechanism design (which characterizes the organizational ways to achieve a specified societal goal) introduced by Hurwicz (1960): by using this theory, we indicate why our way to address Sen's Theorem is almost mandated. Among our other conclusions is a new interpretation of Sen's result, another (which addresses the above puzzle) is that

The research of D. Saari was supported by a NSF grant. Our thanks to P. Hammond for suggesting some useful references. Some of our joint results were reported in Brunel (now Petron) (1998).

the ways we develop to sidestep Sen's problem have several parallels in daily practice. While arguments will be developed in a different paper to support Sen's (1970a) suggestion that "the ultimate guarantee for individual liberty may rest not on rules for social choice but on developing individual values that respect each other's personal choice," our approach identifies societal mechanisms that have similar consequences.

What significantly aids our program is that the source of Sen's negative conclusion now is understood (Saari 1997, 1998, 2001). The surprisingly simple explanation shows that when a procedure satisfying Sen's condition of minimal liberalism (ML) is used with a sufficiently heterogeneous society, the merger effectively eliminates the crucial assumption that the individuals have transitive preferences.[1] Obviously, if individual rationality is ignored, then Sen's conclusion must be anticipated.

What makes this comment unexpected is that Sen's Theorem explicitly requires the individuals to have transitive preferences. The problem, as described in Sect. 3, is that ML inadvertently requires procedures to treat certain transitive profiles as though the data comes from individuals with cyclic preferences. But if cyclic, rather than the actual transitive preferences, are being mistakenly serviced, it is readily understandable why cyclic societal outcomes occur. Even more: beyond illustrating a conflict between individual rights and Pareto optimality, Sen's result demonstrates how natural assumptions can, unintentionally but effectively, jettison other crucial assumptions that we mistakenly believe are being used.

According to this description, in order to obtain positive conclusions, we must discover how to incorporate the intended transitivity information. There are many ways to do this: one offered here (Sect. 4) extracts a particular informational aspect about transitive preferences that, when used to modify ML, converts Sen's negative conclusion into a positive assertion. This condition is based on a new interpretation for Sen's societal cycles that significantly shifts the emphasis about who is being wronged. Rather than describing a person's right to select, the cycles can be viewed as reflecting societal conflicts where the actions of decisive agents strongly impose upon the interests and wishes of others.[2] To avoid this infringement, the informational facet we identify, a "strong negative externality constraint," determines when a decisive agent can be decisive. Because the spirit of our condition mimics what happens in society, this result can be viewed as providing theoretical insights into centuries old, pragmatic practices that have evolved to avoid these concerns. Then we indicate why these types of constraints are almost demanded by decentralization theory.

A different perspective comes from Fine's observation (1975) that Sen's (1970a) well known Prude and Lascivious example can be identified with the Prisoner's Dilemma game. This fine observation introduced a new research direction: instead of the traditional direct mechanism approach where individuals report their sincere preferences, Sen's structures can be described in terms of games with an emphasis on strategies and finding a choice; e.g., see the articles in Arrow, Sen, and Suzumura (1996) such as Hammond (1996) and Pattanaik (1996).

[1]A similar argument (Saari 2001) explains Arrow's Impossibility Theorem (Arrow, 1951).

[2]All of the papers about Sen's Theorem that we have seen emphasize an individual's sphere of influence and the rights of the decisive agent. Thus, this new explanation is unexpected.

It is easy to show (as we do) that associated with *any* example illustrating Sen's result is a game, and, conversely, any game satisfying some simple conditions generates an example illustrating Sen's Theorem. This connection suggests examining whether game theory can provide ways to address the problems raised by Sen's Theorem. The connection we stress is that, with appropriate assumptions (e.g., an appropriate discount rate, etc.), the 'Tit-for-Tat" strategy for infinitely repeated games leads to cooperative solutions for the Prisoner's Dilemma. As we show, this strategy always is applicable within Sen's structure, and it has parallels with commonly used practices.

2. Sen's Theorem

Sen's Theorem requires individual preferences to be complete, transitive and unrestricted over the alternatives. The only condition imposed on the societal outcome is that it does not admit cycles. As for the decision procedure, Sen assumes the (weak) Pareto condition: if the individuals are unanimous in their ranking of any pair, that is the pair's societal ranking. The remaining condition is a minimal version of the rights of individuals as associated with liberalism.

Definition 1. Minimal Liberalism *(ML) is where at least each of two individuals is assigned at least one pair of alternatives. These individuals are "decisive" over the assigned pair in that the way they rank the pair is the pair's societal ranking.*

Although these conditions seem reasonable and even innocuous, Sen proves the surprising conclusion that with three or more alternatives, no decision procedure always satisfies them. As described in Sen's writings, this outcome suggests a fundamental divide between welfarism and liberalism. In an interesting paper Gaertner, Pattanaik, and Suzumura (1991) expand on this point by commenting that "this problem persists under virtually every plausible concept of individual rights that [they] can think of." Our explanation for Sen's Theorem adds support to their comment because it requires a similar conclusion for all choices of individual rights with certain natural properties.

Notice how each individual is to rank even those pairs of alternatives that, by being assigned to decisive individuals, presumably belong to someone else's personal sphere of influence: as such, Sen's formulation requires externalities. One might wonder whether we could sidestep Sen's problems by excusing individuals from ranking those pairs of alternatives assigned to other decisive individuals. We dismiss this direction because it fails to address the essence of Sen's argument and because societal cycles can still arise. A more interesting reason is that because externalities are an economic and political reality, Sen's Theorem provides a natural setting to examine their consequences; e.g., externalities play a central role when describing Sen's Theorem from the perspective of decentralization.

3. The source of Sen's result

It is customary to prove Sen's theorem, while demonstrating its relevancy, by creating examples similar to what occurs in daily life. The ways the individuals in these examples make personal decisions force cyclic societal rankings. Subsequent to Sen's proof, others have found particularly clever examples that generate not one, but several societal cycles, e.g., Brunel and Salles (1998) and Salles (1997) do so by nicely extending Sen's "Prude and Lascivious" story.

A newly developed approach (Saari 2001) converts the construction of examples from an art form that requires deep insights about human interactions into a simple constructive process. With this approach, it now is easy to generate any number of examples that illustrate Sen's theorem with any number of societal cycles that are intertwined in any desired manner, as well as examples that do, or do not, use the Pareto condition. Aspects of this argument that are central to our discussion are outlined next.

Start with any desired societal outcome where AB means that alternative A is preferred to alternative B. Assign the societal cycle as the (temporary) preferences for each individual: initially, then, individuals have cyclic preferences. To illustrate by creating examples where the societal outcome has the two cycles AB, BC, CD, DA and BC, CE, EA, AB, this assignment defines the following "informational table." With the unanimity of individuals, the Pareto condition mandates the specified societal outcome.

individual	$\{A,B\}$	$\{B,C\}$	$\{C,D\}$	$\{A,D\}$	$\{C,E\}$	$\{A,E\}$
Anne	AB	BC	CD	DA	CE	EA
Barb	AB	BC	CD	DA	CE	EA
Connie	AB	BC	CD	DA	CE	EA
Outcome	AB	BC	CD	DA	CE	EA

(3.1)

The next step is to assign pairs of alternatives to decisive individuals. The only condition imposed on this assignment rule is that

> *for each individual and each societal cycle, there is at least one pair where* another *individual is decisive.*

Notice that this condition holds for *each individual* whether or not the individual is decisive over some pair. In our example, BC occurs in both cycles, so a way to satisfy this condition for both cycles and for both Barb and Connie is to let Anne be decisive over $\{B, C\}$. It remains to satisfy this conditioin for Anne; i.e., pairs from the first and second societal cycles, which differ from $\{B, C\}$, must be assigned to other decisive agents. Because AB is in both cycles, we could let Barb be decisive over $\{A, B\}$. But, to illustrate the flexibility of the approach, let Barb be decisive over $\{A, D\}$ and Connie over $\{C, E\}$. These assignments are reflected in the next information table where the dashes indicate that a person's ranking is irrelevant because the outcome is determined by a decisive agent. With unchanged individual preferences, the societal outcome must also remain unchanged. The only difference is that the BC, DA, EA societal outcomes are now justified by minimal liberalism rather than the Pareto condition.

Person	$\{A,B\}$	$\{B,C\}$	$\{C,D\}$	$\{A,D\}$	$\{C,E\}$	$\{A,E\}$
Anne	AB	BC	CD	–	CE	–
Barb	AB	–	CD	DA	CE	–
Connie	AB	–	CD	–	CE	EA
Outcome	AB	BC	CD	DA	CE	EA

(3.2)

The important point is that the Eq. 3.2 information table also arises if, instead of the cyclic preferences, Anne, Barb, and Connie have, respectively, the transitive preferences

$$ABCDE, \quad CDEAB, \quad CDEAB. \tag{3.3}$$

So the third step is to construct transitive rankings that generate the same information table: as explained later, this always can be done. When presenting the end-product as an example with "two societal cycles" that illustrates Sen's theorem, ignore the initial cyclic preference and use only the constructed Eq. 3.3 transitive rankings.

To further illustrate, instead of the above assignment process, adopt the original option where Barb is decisive over $\{A, B\}$. The information table now becomes

Person	$\{A,B\}$	$\{B,C\}$	$\{C,D\}$	$\{A,D\}$	$\{C,E\}$	$\{A,E\}$
Anne	–	BC	CD	DA	CE	EA
Barb	AB	–	CD	DA	CE	EA
Connie	–	–	CD	DA	CE	EA
Outcome	AB	BC	CD	DA	CE	EA

(3.4)

where associated choices of transitive preferences for Anne, Barb, and Connie could be

$$BCDEA, \quad CDEAB, \quad CDEAB. \tag{3.5}$$

Again, by using the derived transitive preferences rather than the initial cyclic ones, we have a different "two societal cycles" example that illustrates Sen's result.

The reason transitive preferences always can be created at the end of this process is that the way decisive agents are assigned to pairs makes it irrelevant how each individual ranks at least one pair from each societal cycle. But, reversing the ranking of one pair in a cycle (from the initial preferences) creates a transitive ranking.

This approach offers a simple and easy way to construct a wide variety of examples illustrating Sen's theorem. Even stronger, the construction makes it arguable that, at least for those examples that can be created in this manner, the cyclic societal outcomes reflect an intent to satisfy the needs of the original cyclic preferences rather than the later constructed transitive preferences. After all, the societal cycle for the initial setting reflects the original unanimity among the group, and the construction makes it clear that the initial cyclic setting cannot be distinguished from the transitive preferences selected at the end of the process.

These last comments lead to one of our central points. As asserted next, *all possible examples illustrating Sen's Theorem can be constructed in this manner.* This includes Gibbard's (1974) troubling extension, and any other choice of rights that we have examined. (Also see the "colored shirt" example in Gaertner, Pattanaik, and Suzumura (1991).) This assertion makes it arguable that, for *all possible examples* illustrating Sen's theorem, the procedure attempts to respect the wishes of individuals with associated cyclic preferences rather than the intended individuals with transitive preferences.

Theorem 1 (Saari 2001)**.** *Any example illustrating Sen's theorem, which is based on the rankings of decisive agents and unanimity among individuals, can be constructed in the above manner by first assigning the societal cyclic rankings to each person. In this way, the Pareto condition requires the societal outcome to be as specified.*

The important message of Thm. 1 is that Sen's conclusion occurs because minimal liberalism emasculates the assumption that individuals have rational preferences. In turn, this means that to find resolutions of Sen's problem, we must find ways to allow procedures to use the explicitly required information that the preferences are transitive. This search is started in the next section.

4. Regaining transitivity through negative externalities

To sidestep the problems identified by Sen's Theorem, we must find ways to allow a procedure to differentiate transitive from cyclic rankings. One approach uses the notion of a "strong preference" given in Saari (1995, 2001) and used in Brunel (1998). ("Strong" is in the sense of *ordinal rankings.*) Others (e.g., Luce and Raiffa 1957) have used versions of this natural notion, but our usage differs significantly because we use it to distinguish unrelated binary rankings from those with some transitive structure.

The idea is simple: when listing a binary ranking coming from a transitive ranking, also specify how many other alternatives separate the two choices. Thus the $ABCDE$ ranking has a $[BC, 0]$ binary ranking because no alternative separates B and C, but a stronger $[AE, 3]$ ranking because three alternatives separate A and E. When binary rankings are not intended to be related, pairs are not separated. For instance, if $AE, AC, CE, AD, DE, CB, ...$ are binary rankings with no intended relationship among them (even though $\{A, C, E\}$ accidentally satisfy transitivity), we have $[AE, 0]$.

We use these $[XY, \alpha]$ terms, where α measures the *intensity of the binary XY ranking*, to distinguish transitive rankings from general binary ones. This distinction occurs because a transitive ranking always has some positive intensity values, while α always equals zero for unrelated binary rankings. To illustrate, each binary ranking for each cyclic preference in the Eq. 3.1 information table has an $\alpha = 0$ intensity level. But there is a difference when the Eq. 3.2 table reflects the Eq. 3.3 transitive rankings: because Anne is decisive, she can determine the societal $\{B, C\}$ outcome with her weak $[BC, 0]$ preference even though Barb and Connie disagree as manifested by their strongly opposing $[CB, 3]$ rankings. Similarly, Barb determines the $\{A, D\}$ outcome with her $[DA, 1]$ preference, but Anne disagrees with her stronger[3] $[AD, 2]$ views. Finally, Connie's decisive agent's choice of EA from her $[EA, 0]$ preferences affects Anne's opposite and stronger $[AE, 3]$ views. A similar pattern emerges with the Eq. 3.4 information table and the Eq. 3.5 transitive preferences: Barb's decisive $[AB, 0]$ selection is opposed by Anne's strongly opposing $[BA, 3]$ views, and Anne's decisive $[BC, 0]$ choice counters Barb's strongly opposing $[CB, 3]$ views.

4.1. **Dysfunctional societies.** This transitivity information, then, suggests that a decisive agent's choice can create a strong, negative externality for someone else. The observation that a decisive agent's choice can be met with disagreement is not new; it has been recognized by many including Fine (1975), Saari and Brunel as reported in (Brunel 1998), and Hillinger and Lapham (1998). What does appear to be new (as shown below) is that this disagreement must occur in *all possible examples* that illustrate Sen's Theorem; that beyond someone disagreeing with the outcome, in *each cycle someone strongly* disagrees with the decisive agent's choice; that this strong disagreement signals

[3]Remember, these intensity comparisons are being made with ordinal, not cardinal, rankings.

that ML is obscuring whether the individual preferences are, or are not, transitive; and that it identifies how to resolve these difficulties. For purposes of this paper, this central term is defined as follows.

Definition 2. *For any pair of alternatives $\{X, Y\}$, a decisive agent's choice of X creates a strong, negative externality if another agent's sincere ranking is $[YX, \alpha]$ with a positive α intensity.*

To achieve our objective of finding a way to allow the societal outcome to reflect the rationality of the individuals, we change ML so that a decisive agent can make a decision only if the choice does not impose a strong negative externality on someone else. This change in ML, in which an aspect of the transitivity of individual preferences is imbedded, converts Sen's negative conclusion to a positive one. It is easy to loosely connect this conclusion with the spirit of what we experience in daily life. For instance, consider those common noise abatement laws that limit how loudly music can be played in public. But even when music is being played loudly, these laws are enforced only should someone complain. Because enforcement requires a complaint, the personal cost of reporting an infringement tacitly determines whether the negative externality is strong. As another example, the color of a shirt a person wears should be a personal decision even if others disagree. A possible exception is if this choice causes strong negative externalities, as in some large cities where certain colors indicate support for a rival gang: here a reaction manifesting a strong negative disagreement can be lethal.

Theorem 2. *Suppose a decisive agent can determine the societal outcome of an assigned pair only when the choice does not create a strong negative externality for some other agent. The pairwise outcomes determined by the decisive agents and the Pareto condition do not generate cycles.*

What adds value is that this theorem is not targeted toward assisting only one or two individuals: a surprising fact is that the consequences affect everyone. This is because for any societal cycle in any example illustrating Sen's Theorem, *each person* suffers at least one strong negative externality. Sen's societal cycles, then, require *everyone to be strongly and adversely affected by the actions of another.*

Theorem 3. *In each societal cycle caused by Pareto and the choices of decisive agents, the choice from at least two pairs of alternatives create a strong negative externality for someone. Indeed, for each cycle,* each individual *suffers at least one strong negative externality.*

These two theorems uncover an alternative explanation for Sen's Theorem: the societal cycles capture dysfunctional settings where *everyone* is negatively and strongly affected by what someone else does. While discussions of Sen's Theorem have traditionally focussed on the rights of individuals to decide within their personal spheres of influence, Thms. 2 3 promote a distinctly different message. Namely, maybe these personal spheres are not as "personal" as normally assumed if the actions taken can strongly (in our intensity sense) affect others. It is our reading of Mill's statement, for instance, that the circle about each human does not include actions that strongly and negatively affect others. Thus, rather than concentrating on decisive individuals, maybe a way to analyze Sen's result is to direct more thought toward the rights of the victims. This is the spirit of Thm. 2 and those described in Sect. 5.

It is easy to refine Thm. 2. According to Thm. 3, for example, at least two pairs in each cycle cause strong negative externalities, but only one ranking needs to be reversed. So different criteria—perhaps a condition comparing how many individuals are negatively affected in each pair—can be used to make the selection.[4] Also, cardinal rankings could be used. But rather than exploring these refinements, our main interest in Thm. 2 is to demonstrate how the structural source of Sen's Theorem helps to identify natural solutions for these problems. Indeed, as described next, the externalities required by Sen's formulation come close to mandating that any resolution reflects the spirit of Thm. 2.

4.2. **Decentralization.** To explain the last comment, building on the theory of decentralization, as started by Hurwicz (1960), arguments motivated by results in Hurwicz, Reiter, and Saari (1978) were developed in Saari (1984) to characterize all organizational ways to achieve a specified societal outcome (for certain classes of problems). As we should expect, the admissible forms of "decentralization" are governed by what is being modeled. For instance, an agent can have unfettered rights to make decisions only if the consequences of this agent's actions are separated from that of all others. While the admissible kinds of "separation" can be surprisingly subtle, natural examples have the $h(x)g(y_1, y_2)$ or $h(x) + g(y_1, y_2)$ form (where the choice of x has consequence $h(x)$, that of y_1, and of y_2 is $\mathrm{g}(\mathrm{y}_1, y_2)$, etc.). A point to notice is that the "separability" occurs because the $h(x)$ consequences are not constrained, in any mannner, by the acts of others. But, as discussed next, Sen's "no societal cycles" condition prohibits this separability. Consequently, as true in Thm. 2, the actions taken by "decisive agents" must be constrained.

To avoid introducing complicated technicalities, we illustrate the basic idea with a simple example where two agents select, respectively, x and y values with the respective consequences $u_1(x) = x^2$, $u_2(y) = y^3$. So far, the agents' actions are unrestricted. But if a constraint must be satisfied, such as $u_1(x) + u_2(y) \leq 1$, then the act of one person obviously limits what the other can do.

A strikingly similar situation arises in Sen's framework. When considering only a particular pair, or when no constraints are imposed on the societal ranking, the actions of the decisive agents are unfettered. But just as the model problem changes by imposing the $u_1(x) + u_2(y) \leq 1$ constraint, the choice theory setting is changed by imposing Sen's "no societal cycle" condition. This is because, according to Thms. 1, 3, the rankings for at least two pairs in each societal cycle are determined by different decisive agents. To avoid or break the cycle, one of these rankings must be reversed. Thus, to achieve a non-cyclic outcome, the actions of these decisive agents must be coordinated in some manner. (This comment extends; e.g., replacing the "no cycle" condition with any other constraint on societal outcomes that requires coordination of pairwise rankings creates another "impossibility theorem." For instance, this kind of argument also explains why IIA causes Arrow's conclusion. What adds considerable support to the earlier Gaertner, Pattanaik, and Suzumura (1991) comment about individual rights

[4]This approach is being developed by I. Li from UCI.

is that if a definition of "individual rights" involves, in some manner or structure, an unconstrained consequence, then conflict will arise.) As true with Thm. 2, the message of this decentralization argument is that to resolve the difficulties raised by Sen's Theorem, there must be a coordination among the decisive agents' actions.

Theorem 2 demonstrates one way to reintroduce information about the transitivity of preferences to achieve this coordination: other approaches have been proposed. While they were developed independent of our structural argument, they clearly reflect it as they require coordination. As examples we point to the interesting contractual solutions explored by Hardin (1998), Sugden (1978, 1985) and others. While it is true that the resulting mechanisms can experience other difficulties, such as with incentives (e.g., see Sen's (1986) comments about the contractual approach), the first objective should be to understand what needs to be done to get around Sen's perplexing problems before addressing other consequences. This exploration is continued with game theory.

5. Seeking help from games

A problem with the "no strong externality" condition of Thm. 2 is that it probably must be enforced through laws. But, are there self-enforcing methods that are in the spirit of Thm. 2? To find some, we turn to elementary aspects of game theory.

After Fine (1975) established a connection between the Prisoner's Dilemma (PD in what follows) and Sen's Lascivious and Prude example, game theory has become a valued venue for studying individual rights; e.g., see Hammond (1996), Pattanaik (1996) and their references. These papers adopt a game theoretic approach: an overly simplified description is that in Sen's original version, the agents report their sincere preferences, but the generalizations permit other strategic behavior. But with our goal to understand how to resolve the problems (the strong, negative externalities everyone suffers) from Sen's original formulation, our use of game theory appears to differ.

As a player's moves in a game constitute the actions of a decisive agent, it is easy to establish connections between multiplayer games and Sen's Theorem. To indicate how to do so with examples, consider only the part of the Eq. 3.4 information table that involves the AB, BC, CD, DA cycle and the two decisive agents Anne and Barb: this creates the abridged information table

Person	$\{A,B\}$	$\{B,C\}$	$\{C,D\}$	$\{A,D\}$
Anne	–	BC	CD	DA
Barb	AB	–	CD	DA
Outcome	AB	BC	CD	DA

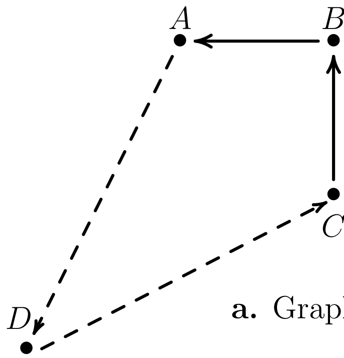
(5.1)

Letting solid vertical and horizontal arrows indicate, respectively, the preferred alternative in each pair where Anne and Barb are decisive, and letting dashed arrows indicate a Pareto improved choice, the information table of Eq. 5.1 is described by Fig. 1a.

The Fig. 1b game matrix mimics the direction of the arrows: the decisive moves are either horizontal or vertical reflecting which player can make these moves. Next, substitute appropriate values for A, B, C, D in the Fig. 1 matrix that, in the appropriate coordinate, satisfy the designated ranking inequality; the dashes indicate where the entry is immaterial. The particular choice made here,

$$\mathcal{G}_1 = \begin{pmatrix} - & 1,1 & 4,0 \\ - & - & 3,3 \\ 2,2 & - & - \end{pmatrix},$$

is not of the PD form. Using the Eq. 3.2 information table, or restoring Connie to this example, leads to a three player game



a. Graph

$$\begin{array}{cc} & \text{Barb} \\ \text{Anne} & \begin{pmatrix} - & A & B \\ - & - & C \\ D & - & - \end{pmatrix} \end{array}$$

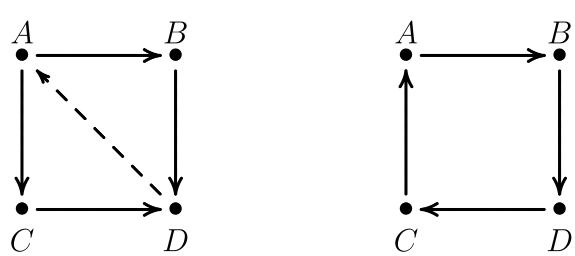
b. Associated game

Fig. 1. From a Sen example to a game

To go from a game to a Sen example, select matrix entries that create a cycle, and then draw solid and dashed lines in the above manner. Two illustrations follow where $\mathcal{G}_2$ is the PD and $\mathcal{G}_3$ is a game with a mixed strategy solution. Anne and Barb are, respectively, the row and column players, and the generic representation for the matrix is $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$.

$$\mathcal{G}_2 = \begin{pmatrix} 5,5 & -1,6 \\ 6,-1 & 0,0 \end{pmatrix}, \quad \mathcal{G}_3 = \begin{pmatrix} 3,0 & 0,2 \\ 1,3 & 2,1 \end{pmatrix} \tag{5.2}$$

Figure 2 shows each game's arrow structure that is used to construct the associated information tables.



$\mathcal{G}_2$; Prisoner's Dilemma $\mathcal{G}_3$; Mixed strategy

Fig. 2. From games to Sen examples

For the PD game $\mathcal{G}_2$, Anne and Barb have, respectively, preferences $CADB$ and $BADC$. The information table for Sen's framework is

	$\{A, B\}$	$\{B, D\}$	$\{A, D\}$	$\{A, C\}$	$\{C, D\}$
Anne	–	DB	AD	CA	–
Barb	BA	–	AD	–	DC
Outcome	BA	DB	AD	CA	DC

(5.3)

For the mixed strategy $\mathcal{G}_3$ game, Anne and Barb have, respectively, preferences $ADCB$ and $CBDA$. Notice that the associated Sen societal cycle in Eq. 5.4 does not use the Pareto condition.

	$\{A, B\}$	$\{B, D\}$	$\{C, D\}$	$\{A, C\}$
Anne	–	DB	–	AC
Barb	BA	–	CD	–
Outcome	BA	DB	CD	AC

(5.4)

While there is a mapping between games and Sen's Theorem, there are far more significant differences that complicate a transfer of insights from game theory to choice theory. While game theory seeks a specific solution, Sen's result seeks a societal ranking. Moreover, a strategy choice in game theory requires other options to disappear; e.g., if Anne selects the middle row in the Fig. 1 game, Barb faces a triplet of options, but only these options. In Sen's framework, agents make decisions for each pair separately.

Nevertheless, certain solution concepts, such as a mixed strategy, do transfer for special settings of Sen's Theorem. To identify the kinds of solution concepts that can be useful, recall from Thm. 2 that we need to incorporate information about individual rationality, as manifested by the strongly negative externality condition, into the analysis. To illustrate with the "Tit-for-Tat" strategy from an infinitely repeated PD, observe that "Tit-for-Tat" is applicable for PD because while a player can create a strong negative externality for his opponent, the repeated structure of the game allows his opponent to retaliate—at times this strategy eventually encourages cooperation. On the other hand, this strategy may seem not to be applicable to Sen's setting because, as illustrated, most examples (and it is easy to make this precise) illustrating Sen's result do not have the PD game structure. But Thm. 3 establishes the needed connection by asserting that, for each societal cycle, any decisive agent who imposes a strong externality on someone else *is vulnerable* to the punishment of having a strong negative externality imposed on him. Consequently, the structure needed to impose a "Tit-for-Tat" strategy *always exists.* (With several decisive agents, implementing this strategy may involve coordinated action among several agents.)

Theorem 4. *If a decision problem with decisive agents is infinitely repeated, a "Tit-for-Tat" strategy, where action is taken to impose a strong negative externality on an agent who imposes a strong negative externality on someone else, always is applicable.*

Whether the "Tit-for-Tat" strategy enforces cooperation depends on how individuals discount the future and the values they assign to their choice and negative externalities, but the analysis is a standard exercise. Our main point is that this strategy always exists, so it identifies a way that differs from Thm. 2 to achieve the needed coordination among agents described earlier. Namely, is shows how to address the strong negative externalities (by retaliating on another issue) that we now know are characteristic of any example illustrating Sen's Theorem. Clearly, the retaliatory behavior has parallels in practice.

To further explore the connection between Sen's result and daily life, consider those situations where actions of individuals impose upon others, but the decisive agents are immune from retaliation. An example from the 1980s is the choice to smoke in a restaurant. In terms of the structure of Sen's Theorem, smokers were immune from sanctions because the non-smokers did not have sufficient power or rights. But once the non-smokers gained power, a cyclic societal outcome arose as manifested by claims and counter claims and even legislative debates. In other words. the societal cycles from Sen's theorem may manifest a transitory state during the emergence of retaliatory power (Thm. 4) from agents who previously suffered from the acts of the unrestrained decisive agents. Brunel (1998) explores related issues.

A weakness with "Tit-for-Tat" is that it assumes a continual (infinite) interaction with players. Thus, we now discuss another natural way to counter the strong negative externalities that cause Sen's result. Rather than using a "Tit-for-Tat" strategy to avoid the PD consequences, we follow the spirit of mechanism design to allow the players to take actions that change the game. Namely, a way to implement some version of Thm. 2 is to create the appropriate societal structures.

Our goal is to understand whether the spirit of Thm. 2 can be captured without appealing to formal societal requirement. This suggests examining what happens in the criminal world. Here we rely on insights developed by Sieberg's (2001) in her use of game theory to explain criminal activities,; e.g., she shows that many settings, such prostitutes without a pimp, dealers without a gang, can be captured by the PD. To explain with her prostitution example, if a customer pays in advance, the prostitute may not provide the intended service; if he is to pay after, he may leave without paying. By being involved in illegal activity, neither agent can appeal to authorities. To avoid these problems, a surrogate government—an enforcement policy policed by a pimp to ensure that customers will not be bilked (so they will return) and prostitutes are ensured payment—is created. To avoid the strong negative externalities, then, the players, as indicated in Eq. 5.5 (from page 66 of Sieberg 2001), take actions to convert the PD into a game where the Nash and Pareto optimal point agree.

$$\begin{pmatrix} (1,1) & (-\beta,\alpha) \\ (\alpha,-\beta) & (0,0) \end{pmatrix} \rightarrow \begin{pmatrix} (1-C,1-C) & (-\beta+P,\alpha-J) \\ (\alpha-J,-\beta+P) & (0,0) \end{pmatrix} \tag{5.5}$$

In Eq. 5.5, select $\alpha > 1, \beta > 0$ to create a PD. The changed environment is modeled by $0 < C < 1$, which is an enforcement charge levied on each party, $J > \alpha$, which is the penalty imposed on the cheating party, and $P > \beta$, which is a payment to the injured party. The first PD matrix creates a Sen cycle given by the Eq. 5.3 information table, while (with the realistic assumption that $-\beta + P < 1 - C$) the second matrix defines the Anne and Barb preferences as, respectively, $ABDC$ and $ACDB$. The associated information table is

	$\{A, B\}$	$\{B, D\}$	$\{A, D\}$	$\{A, C\}$	$\{C, D\}$
Anne	–	BD	AD	AC	–
Barb	AB	–	AD	–	CD
Outcome	AB	BD	AD	AC	CD

(5.6)

with the associated *transitive* societal ranking of $ABCD$. While this surrogate institutional approach does not fulfill Sen's prediction that "individual values [will emerge] that respect each other's personal choice," it is close in the sense that enforcement approaches, either surrogate or legally established, guarantee it. Actions are taken by the players to avoid the strong, negative externalities caused by others.

6. Conclusion

Although well examined for over three decades, Sen's Theorem rightfully remains a source of considerable discussion and insight. As demonstrated here, the structure that explains the source of Sen's Theorem is surprisingly rich; e.g., it demonstrates that rather than describing whether an individual has a right to make a personal decision within his "sphere of influence," the societal cycles reflect a "conflict among rights," a dysfunctional society. This dysfunctional society arises because the actions of the decisive agents create strong negative externalities for the others, and nobody is immune from these difficulties. Once this notion is identified, ways to address these conflicts are forthcoming.

7. Proofs

Proof. [Thm. 2] Assume that the conclusion is false. This means that an example illustrating Sen's example can be created with a societal cycle, but where no individual strongly disagrees with any decisive agent's choice. According to Theorem 1, this example can be created by using the approach described in Sect. 3; this is where each individual starts with cyclic preferences and where, if necessary, the individual's ranking of a pair assigned to a different decisive individual is reversed to convert the ranking from the initial cyclic one to a transitive one. But if two alternatives are adjacent in a transitive ranking, reversing them keeps the ranking transitive. As at least the ranking of one pair from each cycle for each individual needs to be interchanged to generate a transitive ranking, and as this pair has to be one assigned to a decisive agent, for each individual and each cycle the transitive ranking has to have at least one strong disagreement. This contradiction proves the theorem. □

Proof. [Thm. 3] The above proof also proves Thm. 3. This is because to go from the original cyclic preferences to transitive ones, for each cycle each individual has to reverse the binary ranking for a pair that is assigned to a decisive individual. But, reversing this pair makes it a strong negative externality. Hence, for each cycle, each individual suffers at least one positive externality. Also, for each cycle, the construction requires at least two people to be decisive. □

Proof. [Thm. 4] This follows from the above discussion. □

References

[1] Arrow, K. 1951 (2nd ed. 1963), *Individual Values and Social Choice*, Wiley,

[2] Arrow, K., A. Sen, and K. Suzumura, 1996, *Social Choice Re-examined*, Vol. 2, St. Martin's Press.

[3] Broome, J., 1991, *Weighting goods*, Oxford, Basil Blackwell

[4] Brunel, A., 1998, *Contribution à lánalyse des droits en théorie du choix social*, Ph.D. Thesis, Université de Caen.

[5] Brunel, A. and M. Salles, 1998, Interpretative, Semantic and Formal Difficulties of the Social Choice Rule Approach to Rights, p 101-111 in *Freedom in Economics, New Perspectives in Normative Analysis* ed., J-F Laslier, M. Fleurbaey, N. Laslier, and A. Trannoy; Routledge.

[6] Fine, B., 1975, Individual Liberalism in a Paretian Society, *Journal of Political Economy,* **83**, 1277-1281.

[7] Gaertner, W., P. Pattanaik, and K. Suzumura, 1991, Individual rights revisited, *Economica* **59**, 161-177.

[8] Gibbard, A., 1974, A Pareto-Consistent Libertarian Claim, *Journal of Economic Theory* **7**, 388-410.

[9] Hammond, P., 1996, Game Forms versus Social Choice rules as Models of Rights, in *Social Choiced re-examined*, ed. by K. Arrow, A. Sen, K. Suzumura, St. Martin's Press, 82 - 95.

[10] Hardin, R.,1998, *Morality Within the Limits of Reason*, University of Chicago Press

[11] Hausman, D.M. and M.S. McPherson, 1996, *Economic Analysis and Moral Philosophy*, Cambridge University Press, Cambridge.

[12] Hillinger and Lapham, Cpt. 11 *Freedom in Economics: New Perspectives in Normative Analysis* London and New York: Routledge, 1998, pp. 139-157.

[13] Hurwicz, L. , 1960, Optimality and informational efficiency in resource allocation processes, p. 27-46, in: K. J. Arrow, S. Karlin, P. Suppes, eds. *Mathematical Methods in the Social Sciences* Stanford University Press, 1960.

[14] Hurwitz L., S. Reiter, D. Saari, 1978, *On constructing mechanisms with message spaces of minimal dimension for smooth performance functions*, mimeo, Northwestern University, University of Minnesota.

[15] Kolm S.C. 1996, *Modern Theories of Justice*, the MIT Press, Cambridge, Mass.

[16] Luce, R. D. and H. Raiffa, 1957, *Games and Decisions*, New York, John Wiley.

[17] Pattanaik, P., 1996, On modelling individual rights: Some conceptual issues, in *Social Choiced re-examined*, ed. by K. Arrow, A. Sen, K. Suzumura, St. Martin's Press, pp 100-128.

[18] Saari, D. G., 1984, A method for constructing message systems for smooth performace functions, *J. Economic Theory*, **33** 249-274.

[19] Saari, D. G. 1995, *Basic Geometry of Voting*, Springer-Verlag.

[20] Saari, D. G., 1997, Are individual rights possible? *Mathematics Magazine* **70**, 83-92.

[21] Saari, D. G., 1998, Connecting and resolving Sen's and Arrow's Theorems; *Social Choice & Welfare* **15**, 239-261.

[22] Saari, D. G. *Elections and Decisions; Explaining the Unexpected*, Cambridge University Press, Cambridge, England, 2001.

[23] Salles, M., 1997, On Modelling Individual Rights: Some Conceptual Issues: Discussion; p 129-133 in *Social Choice Re-examined, Vol. 2;* Ed. by K. J. Arrow, A. K. Sen, and K. Suzumura, St Martin's Press New York.

[24] Sen, A., 1970a, The impossibility of a Paratian Liberal, *Journ of Political Economy* **78**, 152-157.

[25] Sen, A. 1970b, *Collective Choice and Social Welfare*, Holden-Day,

[26] Sen, A., 1983, Liberty and Social Choice *The Journal of Philosophy* **80**, 5-28.

[27] Sen, A., 1986, Foundations of the Social Choice Theory: An Epilogue, In *Foundations of Social choice Theory* ed. J. Elster and A. Hylland, Cambridge University Press.

[28] Sen, A., 1987, *On Ethics and Economics*, Oxford, Basil Blackwell

[29] Sieberg, K., 2001, *Criminal Dilemmas*, Springer-Verlag, New York.

[30] Sugden, R., 1978, Social Choice and Individual Liberty, In *Contemporary Economic Analysis*, ed. M. Martis,and A. Norbay, Croom Helm, London

[31] Sugden, R., 1985, Liberty, preference and choice, *Economics and Philosophy* **1**, 213-219.

Institute for Mathematical Behavioral Science, University of California, Irvine, CA 92697-5100; GEMMA, Université de Caen, 14032 Caen, France

E-mail address: dsaari@uci.edu, petron@econ.unicaen.fr