Can Deterrence Lead to Fairness?

Riccardo Alberti and Atulya K. Nagar

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/54121

1. Introduction

Game theory has been applied, in the last five decades, to the resolution of strategic situations in order to analyse rational behaviour. Following the growth of the technical apparatus, many requests of finding adequate ethical basis to the theory have been made.

Schelling, in [20], reminds what kind of contributions game theory could yield to the study of ethical system, and in a similar fashion, what ethical basis should we include into game theory.

First of all, game theory provides a well defined set of mathematical tools to formalise known ethical problems, thus it allows people with different backgrounds to approach moral issues by means of rigorous procedures. Moreover, procedures themselves, in the sense of behavioural dynamics and solution concepts, could raise methodological insights to the study of ethics.

On the other hand, game theory has its own ethical basis deriving from utilitarian morality, whereby an action cannot be judged by itself, but only its consequences define its moral value. Though Schelling warns the researchers about this simplistic conception of morality, in its most speculative examples, game theory, embraces the definition of utility from the utilitarian tradition. One may argue, as Schelling does, that a moral calculation could be at the foundation of particular allocation of payoff values or could explain why psychological tests strongly diverge from theoretic result; therefore the issue of ethical foundations in game theory is not exhaustively resolved by the utilitarian view. It is nonetheless incontrovertible that game theory has no meaningful applications when the actors are not interested in the consequences of their actions or when the final purpose of a game goes beyond the information comprised in the payoff functions.

The type of ethics that emerges from this view is called "situation" ethics by Fletcher, for it strongly depends on specific circumstances i.e. the consequences of an action, rather than on universal principles. Situation ethics is also called by other authors "retributive" ([13]). If we take for granted that game theory is founded on a situation ethics, then individual rationality finds a complete moral justification as the mean to solve conflicting situations. However, in [20], Schelling refers to Rapoport's consideration that in many cases, some form of conscience, is superior to individual rationality. Can "situation" ethics resolve every ethical issue of game theory?

It is from the uncertainty on the philosophical foundations of game theory that the issues treated in this work derive. If rationality assumes that an agent should pursue his own selfish interest, how could they be deterred to choose inefficient outcomes? How can we define any notion of fairness? Is there any connection between deterrence and fairness from, at least, a game theoretic perspective? Throughout this chapter we try to answer these questions.

In the first paragraph, we introduce the theory of deterrence, from its ethical justifications and philosophical significance ([4, 5, 13, 19, 20, 22]), to some practical applications and results ([8, 19, 20, 22]).

In the second paragraph, we expose three different approaches in the formulation of theories of justice ([10, 14, 16]) in order to have a complete understanding of the implications of justice related to fairness.

The following two sections are dedicated to the analysis of some game theoretic models which were designed to grasp the key features of deterrence and fairness; in particular the third paragraph contains a conspicuous number of examples ([3, 7, 11, 22]) about how to include deterrence within game theory and, in the fourth paragraph, we show how different authors have managed to incorporate fairness into game theoretic forms ([2, 12, 15]).

Next we introduce our personal contribution to the subject. We provide the extension to general n-person games of our theory of temporised equilibria proposed in [1], and we cite some simple applications in computer science.

We then explore the connections between our model and some fundamental results in the social choice theory ([6, 18]).

Eventually, in the last section, we draw the conclusions on our research and we include some future developments.

2. Theory of deterrence

In its simplest form, deterrence is basically an attempt by party A to prevent party B from undertaking a course of action which A regards as undesirable, by threatening to inflict unacceptable costs upon B in the event that the action is taken [22].

Such a definition of deterrence has its origins from the numerous studies made by sociologists, game theorists, psychologists and theorists of nuclear strategy during the years of the Cold War. In that period, understanding the nature of deterrence was of primary importance to design policies capable to prevent the impact of a nuclear holocaust and, at the same time, maintaining own's strategic positions over the opponent party.

Though a great variety of specifications have been proposed, the common sense of the term is widely accepted to indicate the *enforcement over a number of opponents to refrain from specific unwanted course of action in anticipation of some retaliatory response*.

2.1. Ethical justification of deterrence

The doctrine of deterrence is strongly related to the idea of punishment that one can find in psychology and pedagogical tradition. In fact a deterrent threat is realised by a shrewd manipulation of others' beliefs, in regard to some state of the world, using the pressure of a potential punishment. For it to be effective the subject making the threat must have the capability of carrying it out, and, the threat must be credible and stable (it must not prompt the undesired behaviour) [22].

Most people have agreed that the institution of punishment in the sense of attaching penalties to the violation of legal rules, is acceptable. As Rawls underlines [13], the question that rises from trying to ethically justify punishment, rotates around the justification of punishment itself and not around whether or not punishment is justifiable. To better understand this diversity, he suggests to clearly distinguish the justification of a practice from the justification of a particular action falling under a practice; if such a separation is intended then utilitarianism can be used to provide moral judgement.

There exists two criteria to advocate a punishment: the retributive view for which *it is morally fitting that a person who does wrong should suffer in proportion to his wrongdoing* [13] and, the utilitarian view for which a punishment is justifiable only if it is capable of promoting the interests of the whole society. Without the separation proposed by Rawls, it is very difficult to comprehend any act of suppression. It is a common mistake to think that if a practice is justified by the utilitarian view then each action falling under such practice will follow the same view, hence to grasp the moral implications of behaviour it is necessary to formalise the conception of practice and, to include into common knowledge, the specifications of a practice. Within this framework questioning an action that lies in a practice is equivalent to questioning the practice itself.

In [20], the author reports a very significant example, borrowed from Piaget's studies on the moral abilities of children, that summarises the argumentation of this section: *In short, kids find truth socially useful; lying is bad because the children have freely and contractually adopted a rule against it; the purpose of punishment is to deter, and to reaffirm the rule.* Punishment is justified when it is used to deter individuals from trying to prevaricate the rules of an accepted practice. Moreover if we assume that the intentions towards the definition of a practice are driven by the utilitarian view or situation ethics [20], then the position held by Fischer in [5] becomes clear. In the investigations on collective irrationality he writes that it must be considered moral to implement violent acts when they prevent collective irrationality.

2.2. Difficulties in developing a unified theory of deterrence

The theory of deterrence has been extensively developed and refined since the introduction of the notion of deterrence in the analysis of conflicting situations but still, many problems arose when researchers tried to formalise the dynamics of rational behaviour, subject to threats, into a strong axiomatic apparatus. As Downs points out in [4] a strong theory must be capable of producing reliable predictions based on some well defined information set. In this case, a prediction requires the generation of models that specify benefits and costs, the shape of utility functions, the size of the stochastic component associated with assessments of the cost and the probability of winning [4].

The main problems in achieving this goal rely on the extreme difficulty to set up objective experiments i.e. finding an unbiased relevant population from which to draw a sample, defining the duration of the experiment, grasping some general acceptance thresholds, understanding the role of uncertainties. Moreover, deducing high principles from subjective responses might lead to distorted results; in fact all the abstract models of deterrence made the assumption that every individual is perfectly rational and calculates benefits and costs of each alternatives before making a choice. This is virtually impracticable in real life where factors like culture, emotions, personality etc. enter into the decision process.

Another potential problem for the experimental approach should arise in what we should call aggregate threats. The idea comes from the argumentation of Schelling about the issue of total disarmament [19]. In a totally disarmed world, the re-arming of a nation might be a threat for its neighbours. Such a threat is ineffective on its own unless it is corroborated by the commitment to use the reacquired military power; sometimes deterrence threats must be aggregate in order to make sense and, such aggregations are difficult to identify.

Downs, recalling the work of Achen and Snidal (*Rational Deterrence Theory and Comparative Case Studies*), distinguishes two kinds of approach to a rational theory of deterrence: a weak and a strong version. If, as we have already mentioned, a strong theory should be verifiable, self contained and general, a theory of deterrence is defined weak if it is limited to the simple view of the outcome of choices as a function of expected benefits and costs and it derives by induction on specific cases.

The ultimate objective of a coherent and powerful theory of deterrence is to guarantee an effective policy design founded on necessary conditions for a certain type of behaviour to take place; a strong theory of deterrence, if implementable in Downs' view, could fulfil this requirement.

2.3. Models of deterrence

The main production of models of deterrence comes from the '60s and '70s of the 20th century, during the years of the Cold War. Economic analysis became of primary importance in the examination of costs and benefits and it did take advantage from the utilisation of game theory to investigate strategy through deterrence. Within this scenario, Kahn, in the book *On Thermonuclear War*, distinguishes two forms of deterrence: deterring an enemy's first nuclear strike and, using threat of our own first strike to deter lesser aggressions [8].

As Schelling pointed out in many of his works [19, 20], deterrence is effective only in very restrictive circumstances, namely, only when each participant rationally calculates his benefits according to a consistent system of values. Thus each model underlies the assumption of perfect rationality.

Perhaps the simplest example of such models is depicted in [8]. The problem studied by David and Robert Levine is that of a friendly country (f) deterring one enemy's (e) inimical activity. Following the notation used in [8] we have:

- $a \in [0, \bar{a}]$ is the level of inimical activity;
- $u^e(a)$ indicates the non-decreasing level of utility for the enemy;
- $u^f(a)$ represents the non-increasing level of utility for the friendly country;

- $r(a)$ is the capability of retaliatory response of the friendly country to the inimical activity level a;
- $u^e(a) - r(a)$ is the overall enemy utility;
- $u^f(a) - cr(a)$ is the overall friendly country utility, where $c > 0$ is the marginal cost of retaliation.

Assuming perfect information, the optimal solution is for the friendly country to commit to a sufficiently high level of response to any inimical activity to deter all inimical activities.

In order to relax the assumption of perfect information one can use the quantal response model developed by McKelvey and Palfrey as depicted in [8]. By including the probability density for an inimical level activity to take place, the problem of the friendly country is to maximise the utility function (U) expressed by:

$$U^f = \int_0^{\bar{a}} |u^f(a) - cr(a)| \frac{e^{\lambda(u^c(a)-r(a))}}{\int_0^{\bar{a}} e^{\lambda(u^c(a')-r(a'))} da'} da \tag{1}$$

λ is intended to be the measure of "rationality" of the enemy, i.e. $\lambda = 0$ the enemy behaves completely randomly.

From the analysis of this new model, the optimal solutions are $r(a) = 0$ and $\bar{r}$, which mean respectively "do not retaliate" and "retaliate at the maximum possible level". Hence strategies of the type all-or-nothing are indeed optimal. However, as carefully pointed out in [8], this is not always the case in real contexts. In a more complex scenario with more than one enemy, increasing penalties might change the distribution of inimical activities, incrementing the occurrence of more dangerous actions, rather than deter the wrongdoing. By using Zagare's terminology, deterrence might loose its stability.

We report another game theoretic model of deterrence that is focused on describing the logical structure of mutual deterrence. This model has been proposed by Zagare in [22]. He starts with the consideration that in an anarchic world, stability is achieved by a balance of power maintained by relationships of mutual deterrence; hence the equilibrium dynamics depends on the connections between alternative outcome. If mutual deterrence must fulfil the requirements of capability, stability and credibility then the relationship among the possible outcomes of a mutual deterrence game generates the structure of a 2 × 2 Prisoner's Dilemma.

Zagare reconsiders the descriptivity of the Game of Chicken, that has been the most used model to describe mutual deterrence, in favour of the Prisoner's Dilemma. In the following, we propose the main steps of his argument: we shall indicate with A and B the two players, A plays rows and B plays columns. Table 1 represents the bimatrix of a 2 × 2 Prisoner's Dilemma.

(a_1, b_1)	(a_1, b_2)
(a_2, b_1)	(a_2, b_2)

Table 1. 2-players Prisoner's Dilemma

We designate (a_1, b_1) as the status quo, that is the outcome from which the unilateral defection of one player is undesirable to its opponent. This means that (a_1, b_2) and (a_2, b_1) are less preferred than the status quo, and hence:

$$For\ A,\ (a_1, b_1) > (a_1, b_2) \tag{2}$$

$$For\ B,\ (a_1, b_1) > (a_2, b_1) \tag{3}$$

In addition, there would be no need for mutual deterrence if each player preferred the status quo to the outcome it could induce unilaterally by departing from it. Thus we have the conditions:

$$For\ A,\ (a_2, b_1) > (a_1, b_1) \tag{4}$$

$$For\ B,\ (a_1, b_2) > (a_1, b_1) \tag{5}$$

and, putting 2, 3, 4 and 5 together we obtain:

$$For\ A,\ (a_2, b_1) > (a_1, b_1) > (a_1, b_2) \tag{6}$$

$$For\ B,\ (a_1, b_2) > (a_1, b_1) > (a_2, b_1) \tag{7}$$

Zagare identifies an important feature of a deterrent threat, that is capability. A threat is said to be capable if the outcome imposed to the recipient of the threat is less desirable than the one he can obtain by a unilateral deviation from the status quo. This further consideration brings to the relations:

$$For\ A,\ (a_2, b_1) > (a_2, b_2) \tag{8}$$

$$For\ B,\ (a_1, b_2) > (a_2, b_2) \tag{9}$$

Stability requires the recipient to prefer the original status quo to the outcome associated with the threat, therefore:

$$For\ both\ players,\ (a_1, b_1) > (a_2, b_2) \tag{10}$$

And eventually credibility requires:

$$B\ must\ perceive\ that\ for\ A\ (a_2, b_2) > (a_1, b_2) \tag{11}$$

$$A\ must\ perceive\ that\ for\ B\ (a_2, b_2) > (a_2, b_1) \tag{12}$$

Combining 6, 7, 8, 9, 10, 11 and 12 we obtain the following relations that indeed form the structure of the classical 2 × 2 Prisoner's Dilemma:

$$For\ A,\ (a_2, b_1) > (a_1, b_1) > (a_2, b_2) > (a_1, b_2) \tag{13}$$

$$For\ B,\ (a_1, b_2) > (a_1, b_1) > (a_2, b_2) > (a_2, b_1) \tag{14}$$

Though the simplicity of the model described, there are at least two reasons why mutual deterrence theorists have seldom incorporated the analysis of the Prisoner's Dilemma within their case studies:

- the Game of Chicken represents a worst-case scenario, thus it defines some lower bounds to the problem;
- the Prisoner's Dilemma is a pathological example that shows how individual rationality could bring to social inefficiency. Thus, for players to accept the status quo as the solution of the game will require the introduction of novel equilibrium concepts.

A broader critique on the utilisation of normal form games has also been advanced; in fact a normal form game requires players to make simultaneous decisions and assumes perfect information while in a mutual deterrence scenario actions are more likely to be sequential and conditional.

To overcome the last problem Brams and Wittman in their article *Nonmyopic Equilibria in* 2 × 2 *Games*, cited in [22], introduce the concept of nonmyopic equilibrium (quoting from [22]).

1. both players simultaneously choose strategies, thereby defining an initial outcome of the game, or alternatively, an initial outcome or status quo is imposed on the players empirical circumstances;
2. once at an initial outcome, either player can unilaterally switch his strategy and change that outcome to a subsequent outcome;
3. the other player can respond by unilaterally switching his strategy, thereby moving the game to yet another outcome;
4. these strictly alternating moves continue until the player with the next move chooses not to switch his strategy. When this happens the game terminates, and the final outcome is reached.

In such a practice, recalling Rawls, the nonmyopic equilibrium strategy for the Prisoner's Dilemma will be both stable and socially rational.

3. Fairness and justice

The fundamental idea in the concept of justice is that of fairness [14]; in order to understand the characteristics of fairness, it is, then, indispensable to have a clear comprehension of justice at least in its moral significance. Many philosophical apparatus have been developed to describe justice, but we identified three authors, the work of whom is considered to be seminal for subsequent refinements. These authors are Rawls [14], Nozick [21] and Otsuka [16]. Though their views strongly disagree on many levels, they all share the same assumption that justice is possible only through fairness; the object of fairness is what differentiates the three approaches.

3.1. Justice as fairness

Rawls in his article *Justice as fairness* [14] links justice to the notion of practice. The idea of practice, as we have seen in the paragraph on deterrence, plays a central role in Rawl's philosophy, for it allows the author to distinguish between two levels of moral reasoning. Thus fairness must be referred to the institutions that provides rights and duties rather than to the individual behaviour. In a similar fashion to the principle of equal consideration of interests, depicted by Singer in *Practical Ethics*, the object of Rawls' fairness is equality of opportunities. In fact, for a practice to ensure justice, the following principle must be met:

- every person who participates in a practice must possess the same amount of freedom;
- inequalities are arbitrary unless they serve to a common good.

In addition, a practice is fair if each participant would benefit from it or when all participants are acting as any other participant would do in similar conditions. The utilitarian influence is quite evident in Rawls' scheme and, as we shall see in the following sections, it has indeed found some interesting implementations in game theory, i.e. fairness equilibrium and Kantian equilibrium.

Since an individual is free when he makes the choice of participating in a practice, he voluntarily bounds his self-interests envisaging some higher common benefit; by accepting the condition of a practice, a participant is requested to behave fairly, in the sense that he cannot expect a principle not to be adopted by others if he is not himself disposed to renounce to it. This notion is called "fair play", and its application will become more evident when placed within a game theoretic framework.

Eventually, Rawls' conception of fairness as justice implies a radical view of a person; people's talents do not belong to them and, hence, one can benefit from his own talents only when every person in a practice can benefit from them. Such assumptions, as we explain in the next section, might represent a failure to treat people as equals, since the disadvantages have partial property rights in other people. If this is so, then the distribution of the products of talents might be unbalanced generating an unfair restriction on individual freedom.

3.2. Justice as individual rights

The apparent violation of people's freedom, resulting from Rawls' conception of a person, is morally unjustified in Nozick's view [10]. Nozick founds his philosophy on the idea of self-ownership and individual natural rights and argues that if a person own himself, then he is the owner of his talents and of the products deriving from the application of them. It is a strong form of libertarianism that provides absolute rights over one's properties. Those rights are indispensable for, by taking them away, one is limited in his options and in the pursuit of a self determined way of life.

Though it seems perfectly reasonable that each individual could have some rights on his properties, Nozick's conception may as well generate inequalities in the distribution of public goods that are unjustifiable on a moral level.

One fundamental aspect of Nozick's theory of justice, that plays and important role in linking deterrence to fairness, is the concept of state. A state is an organisation that threats its citizens to use violence if they do not follow the regulations. A night-watchman state, [21], is then morally legitimate to use force to deter its citizens. In this fashion, it is clear, why Nozick limits centralised justice to the restrictive use of force. In paragraph 6 we shall see how this conception of unbalanced distribution of rights on the use of force could be used to implement fairness from strategic situations.

3.3. Fairness as common ownership

To complete our exposition about the perspective adopted in the definition of a coherent theory of justice, we expose an intermediate position to Rawls' and Nozick's approaches: left-libertarianism. Treated by Otsuka in 1998, it has gained a growing consensus in the last decade [16]. Besides, individuals are self-owned, but they can exercise a property right on some natural resources only with the common consensus of others.

Some difficulties with this egalitarian view may arise when an individual with a scarce productivity would receive the rights to own a bigger part of the world; such a distribution of resources might damage the capability of high efficient producers that would need resources for their work. In this case the assumptions of a fair distribution are violated. One possible way out is to establish a flow of transaction from the capables to the disadvantages, in order to balance the inequalities. Risse, in [16], calls these transaction "solidarity", and utilise the concept as a moral justification for common property.

Eventually, in our view, common ownership and self-ownership together give the most comprehensive definition of fairness.

4. Game theory of deterrence

In the paragraph about deterrence, we have presented two game theoretic models that have been extensively used to study the dynamics of players' behaviour under the influence of some deterrent threat. As Myerson underlines in [11] games are to be considered as simplification of life, thus the understanding of real scenarios through game theoretic models must be accompanied by a process of interpretation. Players are intelligent in the sense that they have a prefect understanding of every aspect of the game and are rational in the sense that they will always choose the action that maximises their individual payoff.

For historical reasons, the strongest argumentations and the most comprehensive models of deterrence were first proposed during the years of the Cold War, hence it is not surprising that the great number of analogies and stories associated to such models are borrowed from peculiar scenarios of those years.

4.1. Strategic form games

Herman Kahn, in his book *On Thermonuclear War*, was the first to link a nuclear crisis to the strategic behaviour of players in the game of Chicken. Table 2 reports the matrix form of the game. Each player is given two choices "cooperate" and "defect". We assume player A plays rows and player B plays columns.

<Draw>	<A wins>
<B wins>	<Disaster>

Table 2. Game of Chicken

For each player i, preferences can be ordered as follows: (1) i wins, (2) Draw, (3) i loses, (4) Disaster. If both cooperate then the game will end up in a draw and if neither player is willing to cooperate than the result is an escalation to nuclear disaster. If only one player behaves cooperatively, it can be exploited by the opponent to his advantage.

As we mentioned in the first paragraph, Zagare criticises the usage of this game form as a model for deterrence and he proofs how the Prisoner's Dilemma might be more suitable. But either models cannot overcome the limits of their theoretic formulation. The critique, is not directly addressed to the game of Chicken or to the Prisoner's Dilemma, but rather to the application of the theory of non-cooperative normal form games to the issue of deterrence. It is the unrealistic assumption that once a strategy has been chosen, players are not given a chance to reconsider and to promote a more compromising attitude.

To overcome the discrepancy with real life, game theorists have developed models based on different type of game forms.

4.2. Extensive form games

The models based on the class of extensive form games allow the presence of sequential move. A common situation studied within this framework provides the presence of two players that we may call *challenger* and *defender*. In the simplest analysis, if the challenger challenges, the defender may resist or submit and if the challenger waits, the status quo continues. Again, a weak defender is one who prefers to submit rather than defend and a strong defender is one who prefers to defend rather than submit. For each type of defender there exists only one solution. When the defender is weak, a challenge is always followed by a submission; when the defender is strong, the challenger waits. Though its sequentiality, this model still has no deep insight into the problem, especially because it makes the assumption of perfect information.

One alternative is to employ a model inspired to the Cuban missile crisis in the 60's. Such a model, called Hawks and Doves, requires the presence of two type of behaviour: a "hawk" behaviour, for both challenger and defender, is assumed to escalate to war with disastrous consequences for both sides, and, a "dove" behaviour from the defender will yield a compromise with a "dove" challenger and a victory to a "hawk" challenger. Eventually a "hawk" defender will win against a "dove" challenger.

This game allow us to characterise the credibility of a threat by using a purely game theoretic argument. In fact, only the subgame perfect equilibria form a credible threat [7]. Figure 1 represents the sequential game of Hawks and Doves, with arbitrary payoff values.

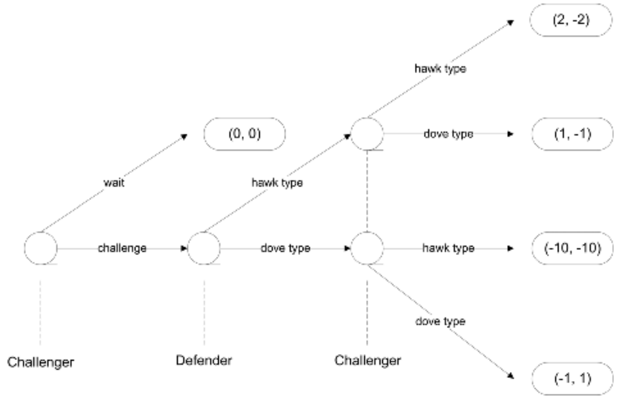

Figure 1. Hawks and Doves

One might complexify the situations represented by this model by adding uncertainty to the knowledge of the challenger about the defender's true preferences. In this way, the challenger never knows which type of defender he is facing, he can only form beliefs about it. The equilibria in the modified example are simply perfect Bayesian since strategies are optimal given beliefs and, beliefs are consistent with strategies on the equilibrium path. We shall say that deterrence has been successful when the challenger is deterred from making demands by the expectation that the defender will always resist so that he will have to backdown with some high probability.

In *Nuclear Deterrence Theory*, Robert Powell proposes another model that grasps the feature of deterrence as being a strategy that leaves something to chances. The idea behind this model is again borrowed from the Cuban missile crisis. In this model, Powell describes a crisis as a ladder of escalation steps that the two sides can in turn climb. Each time a further step is taken, an autonomous risk of nuclear escalation rises. At each step, players are provided with three choices: to surrender, to attack by conducting a full scale nuclear strike and, to take a further step on the escalation ladder. We shall take this example as a strong metaphor for the dynamics of deterrence.

4.3. Repeated Games

Deterrence is a major theme of game theory since, recalling the utilitarian view of rationality, a player's action is measured on the utility it brings rather than on *a-priori* ethical issue related to the action itself. Hence it is not surprising that only a rational decision maker is likely to be moved by deterrent threats.

The theory of repeated games might help us in the analysis of this utilitarian feature of players.

We shall go back to the bimatrix associated to the Prisoner's Dilemma 1 in paragraph 1. As already pointed out in [22] and again underlined in [11], though each player has a deterrent strategy (a_1, b_1) that motivates his opponents to act cooperatively, the issue of credibility is raised from the allocation of payoffs that occurs in this particular example. When B is cooperative (b_1), player A would get a_1 by doing the same as B, but, recalling equation 4, A could get a higher payoff by playing a defective strategy (a_2). Therefore A prefers not to follow his own deterrent strategy when B cooperates, unless A constrains B to follow A's deterrent strategy. But in a pure strategic environment where no communication is allowed, this may not happen thus the credibility of A deterrent threat is compromised. While Zagare bounds his analysis to the explanatory features of the Prisoner's Dilemma, Myerson, in [11], pushes the equilibrium analysis further, considering the implications related to the repeated game.

In such a scenario the concept of reputation becomes even more important. We shall define reputation as the attitude of a player towards a specific strategy, i.e. if A plays the cooperative strategy in a certain number of games, then he gains the reputation of using such strategy.

Let us suppose that A has the reputation of using the strategy "do same as B" against which B plays the cooperative strategy. This will lead to the Pareto efficient outcome. If A changes his strategy and loses his reputation, then the game will eventually end up in the Nash equilibrium which is worse off for both.

The main conclusion one can draw from the repeated game analysis is that the only way to maintain a deterrence threat credible is to keep a sufficiently high level of reputation for this will cause the solution of the game to be the deterrent strategy for both players, at least on the long run.

So far we have introduced classical game theory models and their implications to the theory of deterrence. In the next section we discuss a kind of games, in the class of qualitative games, that are expressly developed for the study of deterrence.

4.4. Games of deterrence

In [3], the authors introduce the concept of games of deterrence as inheriting its key structural characteristics from Isaac's attempt to develop a consistent theory of qualitative games. In fact, instead of dealing with real valued, continuous functions over the product space of individual strategy, each player is provided with a binary valued index that maps a point from the joint strategy space to the set $\{0, 1\}$. An outcome is unacceptable if it is labelled with 0 and it is acceptable if it is labelled with 1. Obviously a rational player will look for an acceptable outcome.

We can distinguish three types of strategies:

1. "no risk": a strategy that guarantees that a player ends up in an acceptable outcome, whether his opponent is rational or not;
2. "limited risk": a strategy that guarantees that a player ends up in an acceptable outcome, as long as his opponent is rational;
3. "high risk": a strategy that give a player an unacceptable outcome, whether his opponent is rational or not.

If a player has no "high risk" strategies, then his strategies are termed positively playable; if he has no positively playable strategies, then his strategies are said playable by default. Eventually, a strategy which is not positively playable nor playable by default is termed non playable. Table 3 represents a simple example of game of deterrence. We assume player E plays rows (e_1, e_2) and player R plays columns (r_1, r_2).

$(1, 0)$	$(1, 1)$
$(1, 1)$	$(1, 1)$

Table 3. Game of Deterrence

Both E's strategy and r_2 for R are "no risk" and playable, and r_1 for R is not playable.

The idea behind the solution concept for this type of games is rather simple: since a rational player will always select an available strategy of, at most, "limited risk", it follows that an equilibrium point is any pair of playable strategies. If we associate to a strategy x a positive playability index $J(x)$ such that, $J(x) = 1$ if x is positively playable and $J(x) = 0$ if not, a solution is, obviously, the set $J(x)$.

Another important feature, of this particular class of games, is the introduction of the *graphs of deterrence*. A graph of deterrence is a bipartite graph such that an arc with origin in x and extremity in y represents the fact that the player who selects x will obtain 0 in the joint strategy (x, y). In reference to 3, the graph of deterrence should be:

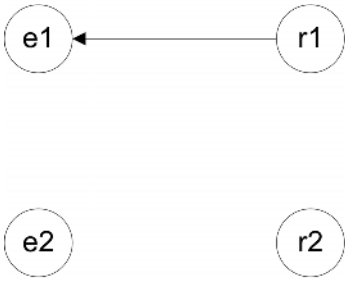

Figure 2. Graph of deterrence of game 3

The games of deterrence have been effectively used to model business processes and network congestion problems.

5. Games of fairness

Most economic models are based on the assumption that rational agents purse their own selfish interests and do not care about any common outcome. Such a behaviour is evident in various examples like the Tragedy of the Commons and the Prisoner's Dilemma.

The Standard View, as Ambrus-Lokatos calls it in [2], considers games as purely means to understand the dynamics of rationality in strategic situations, thus it is not surprising that inefficient solutions can (often) arise. However, as many psychological experiments have remarked, individuals, sometimes act in ways that diverge from the game theoretic expectations; emotions, culture and, in particular, altruism have received some consideration in the realisation of economic and psychological models.

5.1. Fairness as a psychological evidence

One approach to incorporate some sort of moral reasoning into game theory, is closely related to the studies on altruistic behaviour made in psychology. It appears that altruism is more complex than "acting as to benefit the well-being of the others"; in fact, people are inclined to adopt an altruistic behaviour with other who are altruistic to them and tend to hurt those who hurt them. In addition, individuals are more inclined to reduce their well-being in favour of some social goal when they believe that the other will do the same. Such a position implies a form of fairness that the game theoretic rational behaviour cannot incorporate. Rabin, in [12], develops a framework for including this kind of factors within a coherent mathematical model. His work is based on three assumptions that are corroborated by a long series of experiments:

- individuals are inclined to be kind with those who are kind with them;
- individuals are inclined to be aggressive with those who are aggressive with them;
- the smaller the cost of an "emotional behaviour" the greater the effect of such a behaviour on strategic choices.

To have a clearer view let us consider the strategic form of the Ultimatum Game.

A proposer offers a decider to split an amount of money X, if the decider is satisfied, than the money are split according to the offering, otherwise they both get no money. The canonical equilibrium for this game is for the proposer to always propose no more than the smallest unit of X (let us say 1 č) and, for the decider, to always accept it.

This solution is far from being fair, and the behaviour of the proposer is rather aggressive. Rabin's assumptions are capable to address this issue. In fact they do. As resulted from many experiments, [2], the response of the decider to the aggressiveness of the proposer is to not accept the offering and adopt an aggressive behaviour.

Moreover, when the value of the smallest unit of X increases (let us say from 1č to 1000č), then the effect of the "emotional" behaviour on the game outcome becomes almost insignificant. This result, is a *de facto* proof of the third assumption, for which an individual in more inclined to pursue his own interest in proportion to the profitability of doing so.

To formalise the concept of fairness in 2 × 2 games, Rabin's model includes elements from the Geanakoplos, Pearce and Stacchetti framework on psychological games. The utility function depends on players' beliefs as well as on their actions following the relation:

$$U_i(a_i, b_j, c_i) = \pi_i(a_i, b_j) + f_j(b_j, c_i)(1 + f_i(a_i, b_j)) \tag{15}$$

where:

- a_i is the strategy of player i;
- b_j represents his beliefs about what strategy j will play;
- c_i indicates i's beliefs about what strategy j believes that i will play;
- $\pi_i(a_i, b_j)$ is the "material payoff";
- f_l is the kindness function that measures how kind i is to j ($f_i(a_i, b_j)$) and how kind i perceives j to i himself ($f_j(b_j, c_i)$).

A pair (a_1, a_2) is defined a fairness equilibrium if a_i maximise $U(a_i, b_j, c_i)$ for each player and if $a_i = b_i = c_i$.

Though some authors, in particular Ambrus-Lakatos in [2], have pointed out that Rabin's consideration of players' beliefs are quite unrealistic, it might be useful to view how Rabin's model can be applied to the resolution of common games.

Recalling the Prisoner's Dilemma bimatrix from Table 1, and applying the theory of the games of fairness, we shall observe that the common Nash equilibrium (defect, defect) is also a fairness equilibrium. It is easy to show how the Pareto efficient outcome (cooperate, cooperate) is as well a fairness game. From [12], if it is common knowledge that both players are playing the Pareto efficient outcome, then each player is aware that the other is being kind to him for he renounces to a higher payoff he could get by unilaterally deviating from this strategy. Thus, by means of the first assumption, each player is inclined to play cooperatively. It can be shown that, as long as the gains from deviating is not too large, the strategy (cooperate, cooperate) is a fairness equilibrium.

Furthermore this conclusion seems to grasp a reasonable aspect of social interaction, for an individual is more inclined to cooperate when he is confident that others will not defect.

Another interesting aspect is how the behaviour of a player changes when he is responding to constrained choice. Rabin considers a degenerate version of the Prisoner's Dilemma to be as follows:

(a_1, b_1)
(a_2, b_1)

Table 4. Degenerate Prisoner's Dilemma

The outcome in this example are ordered, for player A, in accordance with 4. Player B will always defect, forcing the game to end up to the solution (defect, cooperate) which is the only fairness equilibrium. As pure altruism is excluded from this model, i.e. B does not choose to cooperate but he is forced to do so, the dynamic of A's response changes; A does not have any moral obligation towards B and hence he maximises his own payoff.

In our view, the idea of fairness equilibrium, could be employed as an additional feature of canonical solution concept. Let us now consider the game of Chicken. In table 5, we provide a numerical example of the game represented by 2; player A plays rows and player B plays columns and, each player can choose to dare (first row/columns) or chicken (second row/columns):

(-3X, -3X)	(2X, 0)
(0, 2X)	(X, X)

Table 5. Numerical example of the game of Chicken

This game has 2 pure strategy Nash equilibria (dare, chicken) and (chicken, dare) and it can be shown that both are not fairness equilibria [12].

As one may notice from the above solution, the set of fairness equilibria is not a subset nor a superset of the set of Nash equilibria. Hence we suggest that the concept of fairness equilibrium could be considered as a *ex-post* characterisation of the Nash equilibrium. In fact, a Nash equilibrium point that is also a fairness equilibrium, is the solution in which rational and "emotional" behaviour conciliate.

5.2. Fairness as a philosophical necessity

So far, we have associated the notion of fairness to the framework of game theory by means of psychological arguments on individual's behaviour. In [15] a completely different approach is proposed; indeed Roemer's idea is to describe fairness as a moral necessity rather than an "emotional" calculation. The key concept of this view is borrowed from Kant's categorical imperative: *one should take those actions and only those actions that one would advocate all others take as well*. Specifically, Roemer is interested in understanding how players make their choices when facing proportional alternative joint strategies.

Given these premises, the definition of a solution concept is quite straightforward, in fact, if no player has incentives in realising an outcome in which all players' strategies are altered by the same multiplicative factor, then the game is in a Kantian equilibrium.

For the interests of our argumentation, we shall investigate the Kantian solution when Roemer's concepts are addressed to the study of the Prisoner's Dilemma. In this manner the reader may have an exhaustive picture of different approaches to the same class of problems. For this purpose we shall use a modified version of Table 1.

$(1, 1)$	(a_1, b_2)
(a_2, b_1)	$(0, 0)$

Table 6. Modified Prisoner's Dilemma

with $a_1 = b_1 < 0, a_2 = b_2 > 1$.

Once again, the point of interest is the cooperative solution (1, 1). In [15], the author proves that that strategy is indeed a Kantian equilibrium if and only if the relation $(a_1 + b_2) \leq 2$ holds. Moreover, if this is the case, (cooperate, cooperate) is also the only non-trivial equilibrium.

The Kantian equilibrium framework has been used to model different classic problems in game theory, such as, public-good economies, oligopolist markets and redistributive taxation, however, we are more interested in understanding the motivations behind the concept. The Kantian solution differs, in principle, from the reasons behind the implementation of "pure altruism", i.e taking on the preferences of others; it is a rather individual moral necessity to behave in a way such that it maximise oneself payoff when all others behave in a similar fashion [15].

As we did for fairness equilibrium, we suggest to employ Roemer's equilibrium to calculate the moral efficiency of strategic solutions; in fact when a Nash equilibrium is also a Kantian equilibrium, the outcome has rational and ethical validity.

In the next section, we will propose our contribution to the field. After providing the motivations for our model, we will discuss the mathematical details of the concept of equilibrium.

6. Deterrence as a mean of fairness

In [1], we have presented the concept of *temporised equilibria* in the case of 2 × 2 games. Within this framework we claimed the possibility to incorporate fairness into game theory by means of deterrence. What we propose in this paragraph is the extension to our model to the general case of n-person games with an arbitrary (finite) number of strategies.

In its simplest form, the element of a temporised game are: a *judge* who provides the *deterrent* information and a *parametric function* from which each player derives his optimal behaviour. We recall the definition of judge, that resembles in many ways the idea proposed by Schelling in [19] of a centralised organisation as a democratic way to monitor nuclear illicit actions, or as envisaged by Nozick in [21] as the only rational sustainable type of state.

The judge is an entity that is external to the game; he is absolutely reliable and trustworthy and he provides the deterrent information to the players [1].

Our model is inspired by Freud's theory of mind. In his works, Freud considers the human mind as divided into three main element: the id, the superego and the ego. While the id and the superego are completely irrational, the ego serves to balance the demands of the id against those of the superego by realistically assessing the limits imposed by the real world. The ego serves an executive function to maximise the benefits to the whole person.

In our view, there are no substantial differences in the justifications that drove Rousseau to develop contractualism, Freud to define his model of mind, Roemer to produce the concept of the Kantian equilibrium and our framework. They all share the idea of the presence of a *meta* concept that does not directly interact with players, but radically changes, at least in some examples, their strategic behaviour. The novelty or our work is that of linking deterrence to fairness in a game theoretic model.

6.1. Mathematical structure

We provide the extension of the theory of temporised equilibrium to general coupled constrained n-person games. We use a mathematical apparatus in which the temporised solution to 2 × 2 k-game can be derived as a case of the general theory.

Rosen, in [17], shows that when every joint strategy lies in a compact region R in the product space of the individual strategies and that each player's payoff function ϕ_i, $i = 1, \ldots, n$, is concave in his own strategy, then an equilibrium always exists. Besides if the payoff functions satisfy some additional concavity properties then the equilibrium is also unique. These conclusions are valid both for orthogonal constraint sets (R is the direct product of the individual player's strategy spaces S) and for general coupled constrained sets (R is a subset of S).

In our framework, the coupled constrained set is defined by a system of constraints

$$C_j(\phi_j(x)/M_j, \phi_{j+1}(x)/M_{j+1}, k_j) = 0, \tag{16}$$

where:

- $k_j \in \mathbb{R}/0$;
- $M_j \neq 0$ is the maximum value for $\phi_j(x)$;
- $j = 1, \ldots, n-1$.

C_j expresses the weighted parametric difference between consecutive pairs of payoff functions. A game, in which all constraints are regular and have a minimum number of explicit forms that are continuous maps of compact subsets of $[0,1]^l$ into subsets of $[0,1]^{m-l}$, is called k-game.

6.1.1. Definitions and existence of the equilibrium points

The n-person k-game to be considered is described in terms of the individual strategy vector for each of the n players. A strategy profile $x_i \in \mathbb{E}^{m_i}$ is defined for each player $i = 1, \ldots, n$ and the vector $x \in \mathbb{E}^m$ denotes the concurrent strategies of all players where $\mathbb{E}^m$ is the product space $\mathbb{E}^m = \mathbb{E}^{m_1} \times \mathbb{E}^{m_2} \times \ldots \times \mathbb{E}^{m_n}$ and $m = \sum_{i=1}^{n} m_i$. The allowed strategies will be limited by the requirements that x must be selected from the graph of a continuous differential map which is the explicit form of the constraints C_j in some of their variables.

If the payoff function of the ith player is continuous in all the variables and is concave in the ith set of variables for fixed values of the other sets of variables, then these conditions should be satisfied in the convex compact set S, which is the product space of the projections of R onto the subspaces containing the variables of each player. It is evident that S contains R so that, in general, continuity and concavity are required also outside R, that is, outside the investigation region of interest. Figure 3 represents a two players coupled constrained strategy set.

The payoff function for the ith player depends on the strategies of all other players as well as on his own strategy: $\phi_i(x) = \phi_i(x_1, \ldots, x_i, \ldots, x_n)$. It will be assumed that for $x \in S$, $\phi_i(x)$ in continuous, differentiable and is quasi-concave in x_i for each fixed value of opponents strategies. With this formulation an equilibrium point of the n-person k-game is given by a point $x^0 \in R$ such that

$$\phi_i(x^0) = \max_{t_i}\{\phi_i(x_1^0, \ldots, t_i, \ldots, x_n^0) | (x_1^0, \ldots, t_i, \ldots, x_n^0) \in R\} \tag{17}$$

$$(i = 1, \ldots, n) \tag{18}$$

At such a point no player can increase his payoff by a unilateral change in his strategy in R.

The result to follow make use of the notation $x = \langle z_1, \ldots z_{m-n+1}, y_1, \ldots y_{n-1}\rangle$ to indicate the concurrent strategy vector. y_i are the elements of x that constitute a $(n-1)$-dimensional

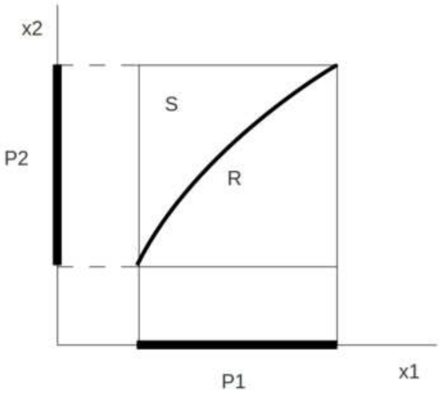

Figure 3. Coupled contrained set with $n = 2$

manifold. Obviously if $z_i \in Z$ and $y_i \in Y$ then $S \supset Z \times Y$. We now prove the equilibrium existence theorem for a n-person k-game.

Theorem 1. *An equilibrium point exists for every n-person k-game.*

Proof. The system of constraints C is an implicit vector-valued function which defines a map from $W \subseteq S \subset \mathbb{R}^{w+n-1} \longrightarrow \mathbb{R}^{n-1}$, following the expression:

$$C(x) : \begin{cases} C_1(x) = \phi_1(x)/M_1 - k_1\phi_2(x)/M_2 = 0, \\ \cdots \\ C_l(x) = \phi_l(x)/M_l - k_l\phi_{l+1}(x)/M_{l+1} = 0 \\ \cdots \\ C_{n-1}(x) = \phi_{n-1}(x)/M_{n-1} - k_{n-1}\phi_n(x)/M_n = 0 \end{cases} \tag{19}$$

By definition of k-game, C fulfils the requirements of Dini's theorem; thus $C = 0$, $C = (C_1, \dots, C_{n-1})$ is solvable in one of its variables.

Formally, if C is differentiable in W and the Jacobian $\partial(C_1, \dots, C_{n-1})/\partial(y_1, \dots, y_{n-1})|_{(z^0,y^0)} \neq 0$, then there are neighbourhoods U and V of $z^0 \in \mathbb{R}^w$ and $y^0 \in \mathbb{R}^{n-1}$, respectively, $U \times V = W$ and $w + n - 1 = m$, and a unique mapping $g : U \longrightarrow V$ such that $C(z, g(z)) = 0 \in \mathbb{R}^{n-1}$ for all $z \in U$. Here $g(z^0) = y^0$ and moreover g is differentiable on U. Let exist a compact set $D \subseteq int(U)$ such that $g(d) \in G_d \subseteq int(V), \forall d \in D$ and $D \times G_d \cap S \neq \emptyset$; G_d is compact since it is the image of a continuous function on a compact set. By Tychonoff's theorem $D \times G_d = R^z$ is also a compact set.

The payoff functions are continuous and derivable on R^z, hence by the extreme value theorem they must attain their maximum and minimum value, each at least once.

If $\bar{z} = \arg\max_z \phi_i(z, g(z))$, $i = 1, \dots, n$ then $(\bar{z}, g(\bar{z})) \in R^z$ is a *temporised* equilibrium point satisfying 17. For suppose that it were not. Then ϕ_l attains its own maximum in $(\bar{z}, g(\bar{z}))$ but $\phi_{l'}$ does not, for some $l = 1, \dots, n-1$; there would be a point $(\tilde{z}, g(\tilde{z})) \in R^z$ such that $\phi_{l'}(\tilde{z}, g(\tilde{z})) > \phi_{l'}(\bar{z}, g(\bar{z}))$, but this condition violates constraint C_l, therefore $(\bar{z}, g(\bar{z})) \notin R^z$. □

We need to focus our attention on the behaviour of g in a limited part of the definition space. Theorem 1 proofs that C defines an implicit function that can be made locally explicit in one of its vector-variables. When such a function exists it is continuous in some open set near a solution point. Let us call SC the set of all C's solution points that fulfil Dini's Theorem requirements; extending the notation from Theorem 1, we indicate with R^s any maximal subset of SC. Therefore for every $s = \langle s_1, \ldots s_{m-n+1}, u_1, \ldots u_{n-1} \rangle \in SC$ then $(s, g(s)) \in R^s$ for at least one R^s. R^s is called the *local graph* of s. Moreover if $e, t \in SC$ and $(e, g(e)) \in R^e$, $(t, g(t)) \in R^t$ and $e \neq t$ then $R^e \cap R^t = \emptyset$; finally the *global graph* is defined as $R = \bigcup_s R^s$. Following this formalism we can give an alternate definition of a n-person k-game: a n-person game is a k-game if $R \cap S \neq \emptyset$.

6.1.2. Uniqueness of the equilibrium point

In order to discuss the uniqueness of an equilibrium point we must describe the regularity condition of the constraints C more explicitly and discuss the properties of R. Let J^y be the square matrix of the partial derivatives in the y's; in Theorem 1's proof it has been shown that the gradients of the constraints C are linearly independent since $det(J^y) \neq 0$. This is a sufficient condition for the satisfaction of the Kuhn-Tucker constraint qualification. Moreover, if c_l and r_l represent J^y lth's column and row, respectively, from $det(J^y) \neq 0$ it follows that for every $l = 1, \ldots, n-1$, if $c_l = \bar{0}$ then $r_l \neq \bar{0}$ and vice-versa. This means that at least one element from each row must be non-zero. We can now state the following:

Theorem 2. *If $(\bar{z}, g(\bar{z}))$ is an equilibrium point satisfying 17 then it is unique in its local graph.*

Proof. Let us suppose that $\frac{1}{M_h} \partial \phi_h(z, y) / \partial y_p - \frac{k_h}{M_{h'}} \partial \phi_{h'}(z, y) / \partial y_p \neq 0$ is the J^y's element at the hth row and pth column. Obviously $\partial \phi_h / \partial y_p$ and $\partial \phi_{h'} / \partial y_p$ cannot be both zero, hence let us assume that $\partial \phi_h / \partial y_p \neq 0$.

It follows that $\nabla \phi_j \neq \bar{0}$ for some y_j and hence at least $(n-1) \mod 2$ payoff functions are strictly monotonic in some R^s for $j = 1, \ldots, n-1$. A function which is quasi-concave and monotonic is strictly quasi-concave. Suppose $(w, g(w))$, $(\bar{z}, g(\bar{z})) \in R^s$ are both maximizers of ϕ_j. Then $w \neq \bar{z}$ implies $\phi_j \left(\frac{1}{2} w + \frac{1}{2} \bar{z}, g \left(\frac{1}{2} w + \frac{1}{2} \bar{z} \right) \right) > min\{\phi_j(w, g(w)), \phi_j(\bar{z}, g(\bar{z}))\} = \phi_j(w, g(w))$ (definition of striclty quasi-concavity) which means that w is not a maximizer of ϕ_j. Consequently each stricly monotonic payoff function attains its own maximum in one unique point $(\bar{z}, g(\bar{z})) \in R^s$, while the other functions are in indifference points. □

We are interested in finding conditions for global uniqueness that translates into uniqueness of the temporised equilibrium point. In order to ensure global uniqueness the following must hold:

Theorem 3. *If $(\bar{z}, g(\bar{z}))$ is an equilibrium point satisfying 17 and SC is a closed set then the equilibrium point is globally unique.*

Proof. Let us consider the local graph R^s at some point $s \in SC$. If SC is closed the maximal closed subset of SC is SC itself, hence there exists a local graph R^s which coincides with SC and $R^s = R$. Applying Theorem 2 to this case, we conclude that there is an equilibrium point $(\bar{z}, g(\bar{z})) \in R$ satisfying 17. The equilibrium point is globally unique for R is the global graph of the game. □

From these conclusions the succeeding holds:

Lemma 1. *The equilibrium point $(\bar{z}, g(\bar{z}))$ is an element of $bd(R^s)$.*

Proof. The proof comes directly from the proof of Theorem 2. Considering every smooth and strictly monotonic ϕ_i in the compact set R^s, if there is a maximum $(\bar{z}, g(\bar{z})) \in int(R^s)$ then there exists an arbitrary small quantity ϵ such that $(\bar{z}+\epsilon, g(\bar{z}+\epsilon)) \in int(R^s)$ and $\phi_i(\bar{z}+\epsilon, g(\bar{z}+\epsilon)) > \phi_i(\bar{z}, g(\bar{z}))$.

But this contradicts the hypothesis for which $(\bar{z}, g(\bar{z})) \in int(R^s)$ is a maximum. Hence $(\bar{z}, g(\bar{z})) \in bd(R^s)$. □

Obviously if the hypotheses of Theorem 3 hold then the equilibrium point is an element of $bd(R)$.

6.1.3. 2 × 2 *games*

In addition to the properties exposed in the previous sections there are some peculiar characteristics that holds for this specific class of games. Figure 7 represent a general 2 × 2 game.

(a_1, b_1)	(a_2, b_2)
(a_3, b_3)	(a_4, b_4)

Table 7. General 2 × 2 game

Proposition 1. *The class of* 2 × 2 *k-games is closed with respect to translation when* $k = \frac{M2}{M1}$.

Proof. C's generic expression for a 2 × 2 k-game is:

$$zy\left(\frac{a_1 - a_2 - a_3 + a_4}{M_1} - k\frac{b_1 - b_2 - b_3 + b_4}{M_2}\right) + z\left(\frac{a_2 - a_4}{M_1} - k\frac{b_2 - b_4}{M_2}\right) + \\ + y\left(\frac{a_3 - a_4}{M_1} - k\frac{b_3 - b_4}{M_2}\right) + \frac{a_4}{M_1} - k\frac{b_4}{M_2} = 0 \tag{20}$$

For $k = \frac{M2}{M1}$ we can multiply equation 20 by M_1 obtaining:

$$zy\left(a_1 - a_2 - a_3 + a_4 - b_1 - b_2 - b_3 + b_4\right) + z\left(a_2 - a_4 - b_2 + b_4\right) + \\ + y\left(a_3 - a_4 - b_3 + b_4\right) + a_4 - b_4 = 0 \tag{21}$$

It is easy to notice that if we add or subtract a positive quantity T to 21 the equation remains unchanged since in every pair the same quantity is added and subtracted. □

In [1] it has been demonstrated that every symmetric 2 × 2 game is a k-game and, by definition, $k = 1$. In a symmetric game $M_1 = M_2$ hence *the class of symmetric* 2 × 2 *games is closed with respect to linear transformation* is a special case of Proposition 1.

We have to make a correction to Theorem 1 in [1] for it is not always true that every 2 × 2 game has one unique temporised equilibrium.

The following theorem contains the revised version of Theorem 1 in [1].

Theorem 4. *Every* 2×2 *k-game has at most two equilibrium points satisfying 17.*

Proof. The explicit form of the constraint C is in general a homographic function. Let us suppose that C can be made explicit in y, so that $y = g(z, k)$ according to the expression:

$$y = \frac{q(k)z + w(k)}{e(k)z + r(k)} \tag{22}$$

with $q(k) = -\frac{a_2 - a_4}{M_1} + k\frac{b_2 - b_4}{M_2}$, $w(k) = -\frac{a_4}{M_1} + k\frac{b_4}{M_2}$, $e(k) = \frac{a_1 - a_2 - a_3 + a_4}{M_1} - k\frac{b_1 - b_2 - b_3 + b_4}{M_2}$ and $r(k) = \frac{a_3 - a_4}{M_1} - k\frac{b_3 - b_4}{M_2}$. From Theorem 1 there exists at least one value for k such that $y \in [0, 1]$ when $z \in [0, 1]$. The study of homographic functions reduces to the discussion of three cases:

- $e(k) = 0$: 22 is the equation of a straight line hence $SC = \{z, g(z, k)\}$ is closed and R is the global graph;
- $q(k)r(k) = w(k)e(k)$: 22 is a straight line and it is parallel to the x axis, $SC = \{z, g(z, k)\}$ is closed and R is again the global graph;
- $e(k) \neq 0$ and $q(k)r(k) \neq w(k)e(k)$: 22 is the equation of an equilateral hyperbola with asymptotes parallel to the coordinate axes. In general if the vertical or horizontal asymptote falls in $(0, 1)$ then 22 can be discontinuous in one point inside $(0, 1)^2$. To avoid such situation it is sufficient to find a value of k such that only one branch of 22 intersects $[0, 1]^2$; however this is not always possible (i.e. if $b_1 = b_2 = b_3 = b_4 = 0$ and $a_1 = 5$, $a_2 = -1$, $a_3 = -1$, $a_4 = 1$ then for any value of k both vertices of 22 are in $[0, 1]^2$). Obviously each branch of the hyperbola is continuous therefore if only one branch intersects $[0, 1]^2$ then $SC = \{z, g(z, k)\}$ is a closed set and R is the global graph, if both branches intersect $[0, 1]^2$ then there exist two local graphs R^1 and R^2.

From Theorem 3. when SC is closed there exists one unique point $(\bar{z}, g(\bar{z}))$ satisfying 17, otherwise there are two equilibrium points satisfying 17, one for each local graph. □

6.2. Further considerations

At this point, we study the solution proposed by the equilibrium concept that we have just discussed to the Prisoner's Dilemma. Table 8 represent the numeric version of the game depicted by Table 1. Player A's strategies are indicated with a and player B's strategies are indicated with b.

$(4, 4)$	$(0, 8)$
$(8, 0)$	$(1, 1)$

Table 8. 2-players Prisoner's Dilemma

By definition, the deterrent in a symmetric game is $k = 1$, hence, the parametric function C takes the form:

$$C(a, b) : \frac{4ab + 8(1-a)b + (1-a)(1-b) - 4ab - 8a(1-b) - (1-a)(1-b)}{10} \tag{23}$$

The two explicit forms of C coincide and are $a = b$; hence when deterred, players' utility functions are:

$$f_1(a) = -3a^2 + 6a + 1, \tag{24}$$

and

$$f_2(b) = -3b^2 + 6b + 1 \tag{25}$$

It is evident from Figure 4, that if A maximises is own utility function on a and B maximises his own utility function on b, then both player will agree to play the cooperative strategy. If this is the case, the solution point is the temporised equilibrium and it is stable and unique.

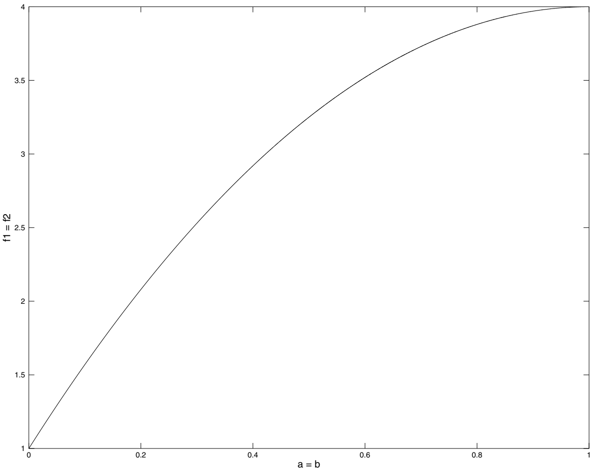

Figure 4. Utility functions in the temporised Prisoner's Dilemma

The key result of our research is that, in most cases, i.e. 8, rationality alone is not sufficient to ensure that each party obtains an optimal outcome - a proper combination of rationality and deterrence can instead implement a fair optimal solution.

6.3. Applications to network analysis

One of the practical field that has mostly benefited from game theory, is the analysis of traffic in telecommunication networks. In particular, our concern is to show the solutions proposed by the temporised framework to classic network problems that have already been thoroughly studied by means of canonical solution concepts. As we noted in [1], the power of the temporised equilibrium lies is that it does not require to include any incentive mechanism to obtain efficient outcomes, for such a feature is already part of the notion of equilibrium.

Thus, for the sake of completeness, we cite four examples of 2 × 2 games, taken from [1] and for each of them we compare the Nash solution with our solution.

6.3.1. Forwarder's dilemma

Suppose we have a simple network made of four nodes, two senders (S_1, S_2) and two receivers (R_1, R_2). Each sender has an assigned receiver and for a packet to go from a sender to its receiver it has to pass through the other sender which then will forward it to its destination.

Again, each sender, when operates as a forwarder, has two options: to forward the packet or to drop it. The cost of forwarding a packet is expressed by the value $0 << c << 1$, the reward for having a packet correctly sent to destination is 1 and the gain for a lost packet is 0. Such a configuration can be formalised as a Prisoner's Dilemma game as depicted by Table 9; S_1 plays rows and S_2 plays columns.

(1-c, 1-c)	(-c, 1)
(1, -c)	(0, 0)

Table 9. Forwarder's Dilemma

It is well known that the Nash equilibrium for this game prescribes each node to drop other's packet.

As we have shown before in this section the temporised solution is to play cooperatively that, eventually, leads to the outcome $(1 - c, 1 - c)$, which is Pareto efficient.

6.3.2. Joint packet forwarder game

In this case, we have a network made of one sender, one receiver and two step nodes S_1, S_2. The sender needs to send a packet to the receiver but, due to the network topology, the packet must pass through S_1 and S_2. Each step node has two choices: to forward the packet or to drop it. Again the cost of forwarding a packet is $0 << c << 1$ for each step node. If both S_1 and S_2 successfully forward the packet, then they gain 1. Table 10 represents the bimatrix associated with the game: again S_1 plays rows and S_2 plays columns.

(1-c, 1-c)	(-c, 0)
(0, 0)	(0, 0)

Table 10. Joint packet forwarder

There exist two Nash equilibria that provide the outcomes $(1 - c, 1 - c)$ and $(0, 0)$.

Let us suppose that the value of the deterrent information is $k = 1$; in this case the temporised solution would be as advantageous as one of the Nash equilibria, prescribing the outcome $(1 - c, 1 - c)$.

6.3.3. Multiple access game

We have two senders S_1 and S_2 that want to access a shared communication channel to deliver some packets to their receivers. Each sender has to option: to access the channel to send the packet or to wait. Let us assume that senders and receivers are in the same power range, thus their transmissions create mutual interference. The cost of accessing the channel and send the packet is $0 << c << 1$ for each sender. A packet is successfully delivered, and it is rewarded 1, when no collisions occur.

Table 11 is the bimatrix associated with the game: as usual S_1 plays rows and S_2 plays columns.

One may notice that this is the game of Chicken. This game has three Nash equilibria which give the outcomes: two in pure strategies $(0,\ 1 - c)$, $(1 - c,\ 0)$, and one in mixed strategies $(1/4 - 1/2c, 1/4 - 1/2c)$. Since the game of Chicken is symmetric, by definition $k = 1$, thus

(0, 0)	(0, 1-c)
(1-c, 0)	(-c, -c)

Table 11. Multiple access

the temporised equilibrium provides the payoff $1/4c^2 - 1/2c + 1/4$. Such payoff is clearly more efficient than the Nash mixed strategy equilibrium as predicted by Corollary 1 in [1].

6.3.4. Jamming game

In the last example we have one sender S_1 and one receiver. Let us assume that the wireless medium is split into two channels and that S_1 is able to select which channel to use at each time step of transmission. Suppose there exist another sender S_2 that is also able to send packet via one of the channels and that S_2's aim is to jam S_1's transmission. Eventually, S_1 receives 1 if its signal is not jammed and -1 otherwise; S_2 receives 1 if it jams S_1's signal and -1 otherwise.

The bimatrix of the game is represented by Table 12. S_1 plays rows and S_2 plays columns.

(-1, 1)	(1, -1)
(1, -1)	(-1, 1)

Table 12. Jamming game

In this last example, the Nash and the temporised solutions coincide.

7. Social choice theory

As many economists have stressed, there exist some basic requirements that a social choice function should possess in order to provide a sustainable structure for its participants. The social choice function must be Pareto optimal in the sense that if every individual prefers any alternative x to another alternative y, then society must prefer x to y. It should also guarantee a mild form of individual liberty, which means that if there are at least two individuals, then each individual has at least a pair of alternatives over which he is decisive. And, every set of individual orderings must be included in its domain set. The Paretian liberal paradox, formulated by Sen in [18], establishes that there is no social choice function that can simultaneously be Pareto optimal and liberal. Such a result has been thoroughly studied by many authors, with different solutions in both economics and game theory.

One fundamental assumption made by Sen is that the Pareto optimality and the liberal conditions must be simultaneously combined to restrict the domain of the social choice function. If this is the case, there is no way out of the paradox, as explained in [18] and [6], for, with at least two individuals, if the domain set of the social choice function contains at least two alternatives, then some form of circular ordering of preferences may always arise.

In our view, Nozick's way out of the paradox with the inclusion of individual rights into states selection, is the most conclusive. In fact when an individual exercises his own rights over possible states of the world, he is actually putting constraints on the set of alternatives open to the social choice. A right is the possibility for a group of individuals to restrain the set of social states to a subset of the original set of states.

Rights do not establish any ordering over the states of the world, but divide them into classes [6]. Once a class is defined, the exercise of a right might exclude it from any further consideration in terms of collective choice. Following the Kantian argument, for which rights must precedes welfare, Nozick depicts a two step procedure for implementing an effective social choice:

1. each individual or group of individuals exercises some right, thus restraining the set of available alternatives;
2. some selection mechanism over the set of remaining alternatives is employed to determine the final choice.

Given this procedure, it is possible to redefine the Pareto optimality condition, given in [18]: if every individual prefers any alternative x to another alternative y, *among the alternatives that are still available after the rights have been exercised*, then society must prefer x to y.

In [6], Gardenfors pushes the analysis of Nozick's ideas even further by creating a rights system formal model. However, to the scope of our argumentation, we underline one aspect of the rights-based approach: (a) *the introduction of a procedure to bound the set of relevant social alternatives to a subset of non-conflicting states*.

A problem, that is strictly related to Sen's paradox, has been proposed by Schelling, as reported by Fischer in [5]. He notices how irrational collective outcomes, in the sense of an inefficient allocation of resources, might arise from individual rational behaviour. His analysis is centred on what structural features a liberal society should possess in order to prevent irrational social outcomes. If societies must guarantee, at least, a mild form of freedom and an efficient distribution of resources, then the relationship between Schelling's problem and Sen's paradox becomes evident.

From the examination of how indeed society are structured, Schelling observes that, the most part of a society consists of institutional arrangements, or practices using Rawls' terminology. The purpose of these arrangements is to overcome individual irrationalities in the interest of a common goal. Such arrangements are not only institutions in the economic sense, but they include forms of popular wisdom, some traditions taken to the level of moral principles or the shrewd and undisputed use of violence. As remarked by Nozick in [21], violence has profound connections with deterrence and fairness when the authority which uses it is confined to a night-watchman state. Finally, in the interest of our discussion, we would like to point out an aspect that emerges from these ideas: (b) *an effective allocation of resources can be reached when rational individuals are forced to follow certain behavioural patterns.*

One might notice that condition (b) might include condition (a); in fact, when some individuals' behaviour is constrained, the set of alternative social state they can reach could be restrained. This is the key concept of temporised equilibria.

When the judge provides the deterrent information, that is by definition symmetric and incomplete [1], it coerces the players to calculate the explicit forms of Equation 19 in order to have a clear understanding of the deterrent threat. A temporised equilibrium is then a point in the coupled constrained set of strategy (a), where individual rationality coincides with collective rationality, for each player is forced, by the judge reputation, to not deviate from the designated pattern (b).

This analogy is also supported by the examples proposed in the preceding section, where it is evident how the effective restraining of the set of alternatives leads to a more convenient overall outcome.

From the analysis in this paragraph, we shall, eventually, draw the conclusions that: the temporised solution realises a social outcome that is Pareto efficient and liberal, and, Nozick's system of rights is somehow equivalent to the application of an accurate utilisation of deterrence as a mean of fairness.

8. Conclusions

In this chapter we have discussed some of the fundamental aspects which link game theory to the broader issues of justice, fairness and social choice. Our argumentation moved from the philosophical foundation of deterrence and justice to the operative definition of game models inspired by such theoretic frameworks. Our own contribution is then the mathematical synthesis of the work from many authors on the subject of deterrence and fairness.

Moreover, we have drawn a correspondence between the temporised equilibrium theory and Nozick's proposal to overcome Sen's paradox in the context of the social choice theory. From such a correspondence we have been able to utilise a game theoretic model to analyse the inefficiencies of rational behaviour when measured on a social scale.

Though most of the significance of the temporised solution is theoretical, we have identified two fields where practical applications could be prolific: artificial intelligence and opportunistic network analysis.

In our view, the analogy with Freud's theory of mind, that inspired the concept of temporised equilibria, could be further exploited: in fact, the mathematics involved could be implemented into an artificial system which can mimic the dynamics of the mind when it struggles to balance the requests from its internal elements. Together with an automated and effective method to assign preferences over a set of states of the world, i.e. hedonistic calculus, and an intelligent system of sensors to collect the stimuli from the environment, it could be possible to realise an artificial conscience, in which a conscious decision is represented by the temporised equilibrium. Such a solution will be optimal, in the sense that it is the best possible allocation taking into consideration the preferences of the building blocks of the mind.

Another interesting area of application is the analysis of opportunistic network [9]. Opportunistic networks (Oppnets) differ from traditional networks in which the nodes are all deployed together and where the size of the network and locations of all its nodes pre-designed (at least the initial locations for mobile networks). In oppnets, we first deploy a seed oppnet, which may be viewed as a pretty typical ad hoc network. The seed then self-configures itself, and then works to detect "foreign" devices or systems using all kinds of communication media-including Bluetooth, wired Internet, WiFi, ham radio, RFID, satellite, etc. Detected systems are identified and evaluated for their usefulness and dependability as candidate helpers for joining the oppnet. Best candidates are invited into the expanded oppnet. A candidate can accept or reject the invitation (but in life-or-death situation it might be ordered to join). Upon accepting the invitation, a helper is admitted into the oppnet. How to select the efficient helper nodes is a vital research field in Opportunistic Network. The resources of the admitted helper are integrated with the oppnet, and tasks can be offloaded to or distributed amongst this and all other helpers. A decentralised command centre presides over the operations of the oppnet throughout its life.

One may notice how the architecture of a typical opportunistic network resembles the theoretic model of temporised equilibria. Though in this field of application some work has already been done by the authors, i.e. the definition of the helper selection protocol and the formalisation of the set of rules to manage the assignment of shared resources, there is still a long way to go to design an efficient and self-sufficient system for finding helper nodes in a more articulated scenario.

Author details

Riccardo Alberti and Atulya K. Nagar
Centre for Applicable Mathematics and System Science (CAMSS), Department of Mathematics and Computer Science, Liverpool Hope University, United Kingdom

9. References

[1] Alberti, R. (2010). Temporised equilibria: a rational concept of fairness into game theory. *International Journal of Computing Science and Mathematics*, Vol. 3, No. 3, (December 2010)

[2] Ambrus-Lakatos, L. (2002). On Preferences for Fairness in Non-Cooperative Game Theory. (June 2002)

[3] Bestougeff, H., Rudnianski, M. (1998). Games of Deterrence and Satsficing Models Applied to Business Process Modelling. (1998)

[4] Downs, G. W. (1989). The Rational Deterrence Debate. *World Politics*, Vol. 41, No. 2, (January, 1989) page numbers (225-237)

[5] Fischer, C. S. (1981). Review: Solving Collective Irrationality. *American Journal of Sociology*, Vol. 87, No. 2, (September, 1981) page numbers (438-444)

[6] Gardenfors, P. (1981). Rights, Games and Social Choice, *Noûs*, Vol. 15, No. 3 (September 1981), page numbers (341-356)

[7] Langlois, J. P. (2002). Applicable Game Theory, (2002) Chapter 3

[8] Levine D. K., Levine R. A. (2005). Deterrence in the Cold War and the "War on terror". (September, 2005)

[9] Lilien L.; Bhuse V. & Gupta A., (2006). Opportunistic Networks: The Concept and Research Challenges in Privacy and Security. *International Workshop on Research Challenges in Security and Privacy for Mobile and Wireless Networks*, (March 2006)

[10] Lukasova, A. (1995). Nozick's Libertarianism: A Qualified Defence, (1995), 2pp

[11] Myerson, R. B. (2006). Force and restraint in strategic deterrence: a game-theorist's perspective. *based on a talk presented at the Chicago Humanities Festival on Peace and War*, (November, 2006)

[12] Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, Vol. 83, (1993) page numbers (1281-1302)

[13] Rawls, J. (1955). Two Concepts of Rules. *The Philosophical Review*, Vol. 64, (1955) page numbers (3-32)

[14] Rawls, J. (1957). Justice as fairness. *The Journal of Philosophy*, Vol. 54, No. 22, (October, 1957) page numbers (653-662)

[15] Roemer, J., E. (2010). Kantian Equilirium. *Scandinavian Journal of Economics*, Vol. 112, No. 1, (2010) page numbers (1-24)

[16] Risse, M. (2004). Does Left-Libertarianism Have Coherent Foundations?. (April, 2004)

[17] Rosen, J. B. (1965). Existence and Uniqueness of Equilibrium Points for Concave N-Person Games, *Econometrica*, Vol. 33, No. 3 (1965), page numbers (520-534)

[18] Sen, A. (1970). The Impossibility of a Paretian Liberal, *Journal of Political Economy*, Vol. 78, (1979), page numbers (152-157)

[19] Schelling, T. C. (1962). The Role of Deterrence in Total Disarmament. *Foreign Affairs*, Vol. 40, No. 3, (April, 1962) page numbers (392-406)

[20] Schelling, T. C. (1968). Game theory and the study of ethical systems. *The Journal of Conflict Resolution*, Vol. 12, No. 1, (March, 1968) page numbers (34-44)

[21] Vallentyne, P. (2006). Robert Nozick, Anarchy, State and Utopia. *Central Works of Philosophy*, Vol. 5, (2006) page numbers (86-103)

[22] Zagare, F. C. (1985). Toward a Reformulation of the Theory of Mutual Deterrence. *International Studies Quarterly*, Vol. 29, (1985) page numbers (June 155-169)