Optimal Transfers in Noncooperative Games

Pradeep Dubey* and Siddhartha Sahi†

10 July 2016

Abstract

We consider "social contracts" which alter the payoffs of players in a noncooperative game, generating new Nash Equilibria (NE). In the domain of contracts which — in conjunction with their concomitant NE — are "self-financing", our focus is on those that are (Pareto) optimal. By way of a key example, we examine optimal levels of crime and punishment in a population equilibrium.

JEL Classification: C70, C72, C79, D44, D63, D82.

1 Introduction

We consider the possibility of a "social contract" among the players of a noncooperative game, which specifies alterations in their payoffs (or, more generally, in the outcomes on which payoffs depend), contingent on their choice of pure strategies. The purpose of the contract is to make everyone better off. However, there are certain constraints which it would be natural for the contract to satisfy.

To begin with, in the scenario that we consider, the contract itself can be specified exogenously, and is capable of being monitored, but such is not the case with players' behavior[1]. That perforce arises in an endogenous manner, with each

*Stony Brook Center for Game Theory; and Cowles Foundation for Research in Economics, Yale University

†Department of Mathematics, Rutgers University, New Brunswick, New Jersey

[1] In other words, the terms of the contract are observable, namely the pure strategies used by the players and the alterations made in their resultant payoffs. Thus any violation of the contract can be discerned (and presumably corrective action taken, though that lies outside our model). What

player acting in his own self-interest in the competition that ensues once the contract is in place. In the parlance of game theory, players' behavior must constitute a (mixed-strategy) Nash Equilibrium (NE) of the game with the altered payoffs.

We rule out the presence of an external agency that can add to, or subtract from, players' payoffs. This leads us to restrict attention to contracts that are "*self-financing*" (or, more generally, "self-sustaining"). It turns out that most contracts are *not* self-financing; those that are, typically form a "thin" submanifold in the space of all contracts, within which the search for socially optimal contracts must be carried out. For concreteness, we work out an example in which one can pinpoint "optimal levels" of crime and punishment in a population equilibrium. This may be of some interest in its own right, as it brings to light a new, game-theoretic rationale for crime in a society, complementing prior inquiries (see, especially, [1]) into the topic.

We also sketch a scheme for extending our analysis to general noncooperative games.

The paper is organized as follows. In section 2 we describe noncooperative games of money and the notion of optimal transfers in them. As a special case, we examine two-person symmetric games whose NE correspond to population equilibria (see section 3). Our lead example is a 3×3 matrix game with crime and punishment, where the focus is on self-financing contracts that lie on a curve of degree 2, so that becomes tractable to compute socially optimal points. Section 4 outlines a scheme for extending our analysis to general noncooperative.

2 Noncoperative Games of Money

To crystallize matters, let us first consider a noncoperative game in which all players are paid in the same coin[2], which we shall simply call *money*. Thus the outcome to each player is a specific amount of money that depends on everyone's choice of pure strategies. Further suppose, to begin with, that each player $i \in N = \{1,\ldots,n\}$ is *risk neutral*; so that w.l.o.g. i's cardinal utility function[3] may be taken to be $u^i(x) = x$.

The game is denoted $\Gamma = \left(S^1,\ldots,S^n;\pi^1,\ldots,\pi^n\right)$ where S^i is the (finite) set of

cannot be done is to dictate the strategies that the players will choose.

[2]e.g., gold, wheat, or more generally a basket of goods that are convertible (in either direction) to money, at fixed prices.

[3]By an affine transformation of the utility function, which does not affect behavior in mixed strategies, we may assume that utility is of this form.

pure strategies of player $i \in N$ and, letting $S = S^1 \times \ldots \times S^n$, the payoff of i is given by a real-valued function

$$\pi^i : S \longrightarrow \mathbb{R}$$

where

$$\pi^i(s) = \text{the money that accrues to } i = \text{the payoff of } i$$

for all $i \in N$ and $s \in S$. Let X^i denote the set of mixed strategies of i and let $X = X^1 \times \ldots \times X^n$. For any $x \in X$ and $s \in S$, we shall denote by $\Pr(s,x)$ the probability that s occurs if x is played.

2.1 Social Contracts

A *social contract* τ specifies transfers of money to players, contingent on the n-tuple $s \in S$ of pure strategies played by them:

$$\tau : N \times S \longrightarrow \mathbb{R} \tag{1}$$

Here $\tau(i,s) > 0$ (resp. $\tau(i,s) < 0$) denotes the money given to i (resp. taken from i) when s is played. The imposition of contract τ on the game Γ gives rise to the game $\Gamma_\tau = \left(S^1, \ldots, S^n; \pi_\tau^1, \ldots, \pi_\tau^n\right)$ where

$$\pi_\tau^i(s) = \pi^i(s) + \tau(i,s)$$

for $s \in S$ and $i \in N$. (Thus $\Gamma_0 = \Gamma$.)

Without confusion, let $\Gamma_\tau = \left(X^1, \ldots, X^n; \pi_\tau^1, \ldots, \pi_\tau^n\right)$ continue to denote the *mixed extension* of the above game, with payoffs

$$\pi_\tau^i(x) = \sum_{s \in S} \Pr(s,x) \pi_\tau^i(s) \tag{2}$$

for $x \in X$ and $i \in N$.

Definition 1 *The contract τ is* self-financing *at $x \in X$ in the game Γ if*

$$\sum_{i \in N} \sum_{s \in S} \Pr(s,x) \tau(i,s) = 0$$

This simply says that transfers net to zero; i.e., the expected outgo to players, given x and τ, is exactly covered by the income expected from them.

There may be several additional *a priori* constraints on transfers, to reflect concerns that we shall not explicitly model (such as upper and lower bounds on transfers, or on their ratios, etc.). To accomodate these, we introduce

Notation 2 *Let Ω denote the collection of all maps representing transfers (see (1)); and let $\Omega(\Gamma) \subset \Omega$ be a subcollection, with $0 \in \Omega(\Gamma)$ (of maps that are deemed a priori "available".)*

Definition 3 *The pair (τ, x) is* viable *in the game Γ if: (i) $\tau \in \Omega(\Gamma)$; (ii) x is an NE of Γ_τ; (iii) τ is self-financing at x in Γ.*

Under the heroic, though standard, hypothesis that whenever there are multiple NE, it is possible to select any one desired (by the consent of "society", or the dictum of some "social planner"), we shall consider all viable pairs (τ, x) and introduce a preorder $\succeq$ on them.

Notation 4 *Let $\Lambda(\Gamma) = \{(\tau, x) : (\tau, x) \text{ is viable in } \Gamma\}$*

By Nash's theorem (see [3]), there exists $x^* \in X$ which is an NE of Γ. Then clearly $(0, x^*)$ is viable in Γ, which implies that $\Lambda(\Gamma)$ is not empty.

Definition 5 *Given (τ, x) and (τ', x') in $\Lambda(\Gamma)$, we write $(\tau, x) \succeq (\tau', x')$ if*

$$\pi^i_\tau(x) \geq \pi^i_{\tau'}(x') \text{ for all } i \in N$$

It is evident that $\succeq$ is transitive and reflexive, and thus constitutes a preorder on $\Lambda(\Gamma)$.

Definition 6 *We shall say that the contract τ is* optimal *in the game Γ if there exists x such that (τ, x) is a $\succeq$-maximal element of $\Lambda(\Gamma)$.*

Remark 7 *In other words, τ is optimal in the game Γ if there exists x such that $(\tau, x) \in \Lambda(\Gamma)$ and there does not exist any $(\tau', x') \in \Lambda(\Gamma)$ such that*

$$\pi^i_{\tau'}(x') \geq \pi^i_\tau(x) \text{ for all } i \in N, \text{ with } > \text{ for some } i$$

In view of the remark, a closely allied notion, that will be useful to us, is that of ε-optimal contracts. These are contracts that, intuitively speaking, are optimal up to (small) error ε.

Definition 8 *We shall say that the contract τ is ε-*optimal *in the game Γ if there exists $x \in X$ such that $(\tau, x) \in \Lambda(\Gamma)$, and there does not exist any (τ', x') in $\Lambda(\Gamma)$ such that*

$$\pi^i_{\tau'}(x') - \varepsilon \geq \pi^i_\tau(x)$$

for all $i \in N$.

Proposition 9 *Suppose* $\Omega = \{\tau : -\beta < \tau(i,s) < \beta\}$ *for some* $\beta \geq 0$ *(i.e.,* Ω *admits all transfers with bound* β*). Then, for any* $\varepsilon > 0$*, there exists an* ε*-optimal contract in the game* Γ.

Proof. This is obvious. ■

Proposition 10 *Suppose that, in addition to the boundedness hypothesis of Proposition 9, the set of available contracts* $\Omega(\Gamma)$ *is closed. Then there exists an optimal contract in* Γ.

Proof. This is also obvious. ■

2.2 Beyond Risk Neutrality

In this case, $\pi^i(s)$ is the money that accrues to i when $s \in S$ is played, and his payoff is $u^i\left(\pi^i(s)\right)$. Thus the above analysis can be replicated *mutatis mutandis* with one amendment: equation (2) must be rewritten

$$\pi^i_\tau(x) = \sum_{s \in S} \Pr(s,x) u^i\left(\pi^i_\tau(s)\right)$$

One could, of course, go beyond games of money, to abstract noncoperative games where pure strategies $s \in S$ map into an outcome space on which players' payoffs are defined. In section 4, we shall outline a scheme for extending our analysis. to that setting.

But let us first turn to the special class of "population games" for which sharp results can be derived, starting with a 3×3 example which may be of some interest in its own right.

3 A Special Example: Crime and Punishment in a Population Equilibrium

Consider a society made up of a large number of individuals who meet each other in random pairwise interactions. Each individual can adopt the role of a worker (W), or a criminal (C) or a policeman (P). The payoffs arising from the interaction between any two individuals depend on their roles and are given by the following 3×3 matrix, whose rows and columns are indexed by W, C, P :

	W	C	P
W	b	d	i
C	c	e	g
P	a	f	h

The $\alpha\beta$-entry of the matrix is the payoff to (an individual of) role α in an encounter with role β. The curious choice of letters for the matrix entries is dictated by the following consideration. For $\alpha, \beta \in \{W, C, P\}$, say that α *dominates* β, and denote this by $\alpha \succ \beta$, if α always obtains a higher payoff than β in the subgame given by the 2×2 submatrix indexed by α and β (i.e., α gets more than β regardless of whether the adversary is α or β; thus $W \succ P$ if $b > a$ and $i > h$). We assume that the matrix has the following "cyclic domination" property

$$W \succ P \succ C \succ W \tag{3}$$

It is readily verified that (3) is equivalent to the following three sets of "alphabetical" inequalities, one for each column:

$$a < b < c, \quad d < e < f, \quad g < h < i \tag{4}$$

The justification of these inequalities is as follows. W represents a hard and honest worker, the "pillar of society", who produces significant wealth. In contrast neither C nor P are very productive. However, wealth changes hands (post-production) in the pairwise encounters among individuals. C appropriates much of the wealth produced by W, without incurring W's cost of production, in a WC encounter; indeed, this is what makes C a criminal. The role of the policeman P is to punish C in a CP encounter, by doing unto C precisely what C does to W. In all the other encounters CC, WW, PP, PW the individuals do not interfere much with each other. The three inequalities are now immediate[4].

[4] To further elaborate on the game, think that there is one unit of time in which individuals produce perishable "wealth", followed by a transfer from W to C (or, C to P) during a random WC (or, CP) encounter. Both the production and the transfer are automatically determined by individuals' roles, but an individual has the strategic freedom to choose which role he wants to play. Population equilibria are pure-strategy Nash equilibria (NE) of this one-shot game Γ with a continuum of players.

Note that the payoff to any player in Γ depends (because of the continuum) on his own strategy and the *population distribution* of others' strategies. Thus no unilateral deviation of a player is observable by his rivals. The upshot is (see, e.g., [2]) that the only NE *play* of a T-fold repetition of Γ consists in playing an NE of Γ (possibly varying over time) in each of the T periods . But, as

The inequalities imply that there is a unique "*population equilibrium*", which yields the *same* payoff to W, C, P (see Propositions 13 and 14 below). This common payoff unambiguously measures "social welfare" (obviating the need to consider weighted averages of the three payoffs, with weights for W, C, P that would perforce have to be *ad hoc*).

A natural question arises: can wealth transfers enhance welfare?

To make this precise, consider a *social contract* in which some individuals voluntarily undertake to alter their wealth levels, from the levels stipulated in the matrix, in the course of their interactions. For example, suppose W and P form a "coaltion of the willing", in which W undertakes to alter his wealth in the amounts $\varepsilon_{WW} < 0$ and $\varepsilon_{WP} < 0$ during the WW and WP encounters. (The negative numbers signify that W is giving up wealth in both cases; and, since he cannot give up more than he has, we have the natural constraints that $-\varepsilon_{WW} \leq b$ and $-\varepsilon_{WP} \leq i$.) The wealth so collected from W is earmarked to increase the reward to P in the amount $\varepsilon_{PC} > 0$ whenever P punishes a criminal (*i.e.*, during any PC encounter), an arrangement with which P has no quarrel. We shall make the implicit assumption that these transfers are not so drastic as to disturb the alphabetical inequalities of the last display[5]; and that therefore a unique NE obtains in the game with transfers.

The hope is that, with a judicious choice of a contract $\varepsilon = (\varepsilon_{WW}, \varepsilon_{WP}, \varepsilon_{PC})$, the policeman P can be so incentivized that society is rendered less crime-prone and more productive, benefiting *everyone* in the process, not just the recipients P, but also the donors W, and *even* — though that might have been far from the intent — the criminals C who were not party to the contract in the first place! Since all roles are bound to get the same payoff at equilibrium, all fortunes must rise or fall together; and thus the benefit to C is a price to be tolerated for the sake of enhancing social welfare. Indeed, were such a contract put up for vote in the population, it would receive unanimous approval as it stands to the benefit of all.

However, a moment's reflection reveals that not all contracts $\varepsilon \equiv (\varepsilon_{WW}, \varepsilon_{WP}, \varepsilon_{PC})$ *will be viable*. The announcement of a contract ε changes the matrix and causes a new population equilibrium $x(\varepsilon) = (x_W(\varepsilon), x_C(\varepsilon), x_P(\varepsilon))$ to emerge, where $x_\alpha(\varepsilon)$ denotes the fraction of the population that has adopted role α. Since pairwise encounters are random, the total amount of wealth collected from workers, by ex-

---

we shall see, Γ has a *unique* NE. Hence this NE will be played constantly not only in Γ,but also in any repeated game based upon Γ, making it very prominent indeed. (It is the main object of study in this section.)

[5]These limitations on transfers correspond to the set $\Omega(\Gamma)$ of available transfers that was described in general in section 2.

ecuting the social contract ε, is $-(x_W(\varepsilon)x_W(\varepsilon)\varepsilon_{WW}+x_W(\varepsilon)x_P(\varepsilon)\varepsilon_{WP})$ while what must be handed out is $x_C(\varepsilon)x_P(\varepsilon)\varepsilon_{PC}$. But wealth is neither created nor destroyed, only transferred, via the social contract ε. Hence the equation

$$x_C(\varepsilon)x_W(\varepsilon)\varepsilon_{WW}+x_W(\varepsilon)x_P(\varepsilon)\varepsilon_{WP}+x_C(\varepsilon)x_P(\varepsilon)\varepsilon_{PC}=0 \tag{5}$$

must hold for any contract ε to be viable (note that $x(\varepsilon)$ is a function of ε). Our overall aim is to chart out the domain of all viable contracts and to examine those that are "optimal". In other words, we ask how a society can select a social contract in order the achieve *optimal levels of crime and punishment in population equilibrium*, *i.e.*, levels that generate the highest welfare for all individuals.

3.1 Population Equilibrium

We denote the entries of a matrix N by N_{ij}, and its transpose by N^T. We write $\mathbf{0}$ and $\mathbf{1}$ for the column vectors of all 0's and all 1's respectively. We say that a vector $\mathbf{x}$ is a probability vector if it has non-negative components whose sum is 1, where the latter condition can be written as $\mathbf{1}^T\mathbf{x}=1$. We say $\mathbf{x}$ is *strictly positive* if all its components are strictly positive, and in this case we write $\mathbf{x}\gg 0$.

Generalizing the discussion from the previous section, we may consider a population with many roles $R=\{1,2,\ldots,n\}$ and an $n\times n$ interaction matrix M. Let $\mathbf{x}=(x_\alpha)_{\alpha\in R}$ be a probability vector representing the fractions of the population in the various roles. Then the components of the vector $M\mathbf{x}$ denote the expected payoff to these roles. This leads us to define a *population equilibrium* (PE) of the matrix M to be a probability vector $\mathbf{x}$ at which there is no benefit to an individual who unilaterally deviates from one role to another. Our underlying assumption is that the population is sufficiently large so that we may approximate it by a continuum: if a single individual (or, an "infinitesimal" number of mutants) switches role, this does not affect $\mathbf{x}$.

Consider symmetric Nash Equilibria (SNE) of the two-person symmetric game also represented by M, *i.e.*, with the payoffs of the row player given by M and those of the column player given by symmetry, i.e., by the transpose of M. It is immediate that SNE x of M correspond to PE of M under the identification (see [4] for the original exposition):

probability x_α of playing pure strategy α
$\longleftrightarrow$ fraction x_α of the population that has adopted role α

While our focus is on PE, we shall often find it convenient to talk about them in terms of SNE (bringing the discussion in line with the general setting of section 2). In particular, we are interested in *completely mixed* symmetric Nash equilibria (CMSNE) of the game, in which each pure strategy is played with positive probability. Thus a CMSNE is a strict probability vector $\mathbf{x} \gg 0$ satisfying

$$M\mathbf{x} = v\mathbf{1},$$

where v is a scalar representing the common payoff to all pure strategies (roles) at $\mathbf{x}$.

3.2 S-matrices

Any symmetric game M has an SNE by Nash's theorem (see [3]), yielding a common payoff to both players. If there is only one SNE then, without confusion, we may refer to its payoff as "the value $\nu(M)$ of playing the game M". If furthermore the SNE is a CMSNE then all strategies also have the same value: each must be a best reply, yielding the payoff $\nu(M)$. Translating this to population equilibrium: $\nu(M)$ is the *common* payoff of *every* role in the population, and is thus an unambiguous measure of "social welfare". This motivates the following:

Definition 11 *A matrix M is an S-matrix if it has a unique SNE, which is moreover a CMSNE.*

Notation 12 *We write $\mathscr{S} \equiv \mathscr{S}_n$ for the set of all S-matrices of size $n \times n$.*

Going back to our example, let us at once note:

Proposition 13 *: Let M be a 3×3 matrix that satisfies the property (4),i.e., $W \succ P \succ C \succ W$. Then M is an S-matrix. Moreover, the unique SNE of M is in fact the unique NE of M.*

Proof. See the Appendix. ■

3.3 The Manifold $\mathscr{N}(M)$ of Viable Games

Let M be an S-matrix with value v. With a view to enhancing ν, we consider a perturbation $M + \varepsilon \in \mathscr{S}$. As was said before, we regard the matrix ε as a *social* *contract* according to which the amount $\varepsilon_{\alpha\beta}$ of wealth will be given to[6] α in an $\alpha\beta$ interaction (over and above what is stipulated in M). The first question is: which ε are viable for a society whose *a priori* interactions are given by the matrix M ?

Let $\mathbf{x} = \mathbf{x}_N$ denote the CMSNE of the matrix $N = M + \varepsilon$. At $\mathbf{x}$ the additional wealth needed to honor the social contract $\varepsilon = N - M$, in the $\alpha\beta$ interaction, is $x_\alpha \varepsilon_{\alpha\beta} x_\beta$. But since wealth is neither created nor destroyed by the social contract ε, but only redistributed, we must have[7] $\sum_{\alpha,\beta} x_\alpha \varepsilon_{\alpha\beta} x_\beta = 0$, *i.e.*

$$\mathbf{x}_N^T (N - M) \mathbf{x}_N = 0 \tag{6}$$

We therefore define the *manifold* $\mathscr{N}(M)$ *of viable games in* $\mathscr{S}$, that are anchored on M, as follows:

$$\mathscr{N}(M) = \left\{ N \in \mathscr{S} : \mathbf{x}_N^T (N - M) \mathbf{x}_N = 0 \right\}$$

If there are exogenous (“social” or “political”) constraints on ε, represented by a set Ξ, then we might need to restrict attention to $\mathscr{N}(M) \cap \mathscr{C}$ where

$$\mathscr{C} = \{ M + \varepsilon : \varepsilon \in \Xi \} .$$

3.4 Optimal Contracts for $\mathscr{N}(M)$

Optimal contracts seek to achieve the best welfare on $\mathscr{N}(M)$, *i.e.*, to achieve

$$\sup \{ \nu(N) : N \in \mathscr{N}(M) \}$$

As we shall see in the next section, ν is bounded on $\mathscr{N}(M)$ for any $M \in \mathscr{S}$. Without further ado, it follows that *approximately* optimal contracts always exist, to any desired level of accuracy.

Next observe that, by the definition of $\mathscr{S}$, SNE is unique on $\mathscr{S}$. But then the general upper-hemicontinuity of SNE implies that SNE, hence also ν, are in fact continuous on $\mathscr{S}$. The continuity is clearly inherited on $\mathscr{N}(M) \subset \mathscr{S}$. It follows that optimal contracts exist on $\mathscr{N}(M) \cap \mathscr{C}$, provided $\mathscr{N}(M) \cap \mathscr{C}$ is compact (as in the example below).

Our main goal is to develop algebraic techniques for computing optimal contracts in $\mathscr{N}(M)$ in canonical settings, including in particular our example.

[6] More precisely we should be saying: "to an individual of type α”. (When $\varepsilon_{\alpha\beta} < 0$, this as usual means that wealth is being taken away from α.)

[7] In population equilibria, it is clear that equation (6) means that expenditure equals income with certainty (not just in expectation).

3.5 Algebraic Description of $\mathcal{N}(M)$

For any square matrix N we let $\sigma(N)$ denote the sum of its entries, *i.e.*

$$\sigma(N) = \mathbf{1}^T N \mathbf{1}.$$

If N is invertible with $\det N = \delta \neq 0$ then we have the well-known formula

$$N^{-1} = \delta^{-1} C_N^T,$$

where C_N is the cofactor matrix of N, and $(\cdot)^T$ denotes the transpose. In this case the linear system

$$N\mathbf{x} = \mathbf{b}$$

admits the unique solution

$$\mathbf{x} = N^{-1}\mathbf{b} = \delta^{-1} C_N^T \mathbf{b}.$$

Alternatively, by Cramer's rule we have

$$\mathbf{x} = \delta^{-1} (\det N_1(\mathbf{b}), \dots, \det N_n(\mathbf{b}))$$

where $N_i(\mathbf{b})$ is obtained from N by replacing the i-th column of N by $\mathbf{b}$.

The equalization principle[8] of CMSNE leads to equations of the form

$$N\mathbf{x} = v\mathbf{1} \tag{7}$$

where $\mathbf{x}$ is some (unknown) strict probability vector (such that all $x_i > 0$ and $\sum x_i = 1$) and v is an (unknown) scalar. In this case we first solve the system

$$N\mathbf{y} = \mathbf{1}.$$

Then we have $\mathbf{y} = \delta^{-1}\mathbf{z}$ where $\mathbf{z}$ is given by either of two equivalent formulas

$$\mathbf{z} = C_N^T \mathbf{1}, \quad \mathbf{z} = (\det N_1, \dots, \det N_n)$$

where

$$N_i = N_i(\mathbf{1}).$$

[8] At any NE, a player must clearly be indifferent between the pure strategies to which he assigns positive probability. In our context of a CMSNE this implies that all pure strategies must yield the same payoff.

Now if the entries of $\mathbf{z}$ *have the same sign* (>0 or <0), we can get the desired probability vector $\mathbf{x}$ by renormalizing: $\mathbf{x} = \frac{1}{z}\mathbf{z}$, where $z = \mathbf{1}^T\mathbf{z} = \sum z_i$. This gives two equivalent formulas

$$z = \mathbf{1}^T C_N^T \mathbf{1} = \sigma\left(C_N^T\right) = \sigma\left(C_N\right), \quad z = \sum \det N_i$$

Finally to compute the scalar v, we rewrite (7) in the form $\mathbf{x} = vN^{-1}\mathbf{1}$, which gives

$$\mathbf{1}^T\mathbf{x} = v\mathbf{1}^T N^{-1}\mathbf{1} = v\sigma\left(N^{-1}\right)$$

Since $\mathbf{1}^T\mathbf{x} = 1$ we get

$$v = \sigma\left(N^{-1}\right)^{-1}$$

In view of this discussion, it makes sense to define, for any square matrix N

$$\mathbf{x}(N) = \frac{C_N^T\mathbf{1}}{\sigma\left(C_N\right)} = \frac{1}{\sum \det N_i}\left(\det N_1, \ldots, \det N_n\right), \quad v(N) = \sigma\left(N^{-1}\right)^{-1} \tag{8}$$

provided only that $\sigma\left(C_N\right) = \sum \det N_i \neq 0$ and $\sigma\left(N^{-1}\right) \neq 0$, respectively.

The above discussion establishes:

Proposition 14

1. *If N is an invertible matrix, then $\mathbf{x}(N)$ is the only possible CMSNE of N.*
2. *If all* $\det N_i$ *have the same sign then* $\mathbf{x}(N)$ is *a CMSNE, with common payoff* $v(N)$.

Proposition 14 allows us to obtain an algebraic description of $\mathcal{N}(M)$, as well as upper and lower bounds on $v(N)$.

Proposition 15 *We have*

$$\mathcal{N}(M) = \left\{N \in \mathcal{S} : \sigma(C_N(N-M)C_N^T) = 0\right\}$$

The value $v(N)$ is bounded by the minimum and maximum entries of M

$$\min_{i,j} M_{ij} \leq v(N) \leq \max_{i,j} M_{ij} \text{ for all } N \in \mathcal{N}(M). \tag{9}$$

Proof. See the Appendix. ■

For the sake of completeness, we note:

Lemma 16 *If N is not invertible, but is an S-matrix, then its value is* 0.

Proof. See the Appendix. ■

3.6 Semialgebraic Approximation of $\mathscr{S}$

To complete the description of $\mathscr{N}(M)$, it remains to determine $\mathscr{S}$. We shall instead describe below a subset $\mathscr{V}$ of $\mathscr{S}$ that is "very close" to $\mathscr{S}$ in that it differs from $\mathscr{S}$ only on lower-dimensional sets of measure zero and also encompasses all of $\mathscr{S}$ in its closure. Thus $\mathscr{V}$ will serve as a faithful "proxy" of $\mathscr{S}$.

If $I \subseteq \{1,\ldots,n\}$ and A is an $n\times n$ matrix, we write $A(I)$ for the matrix obtained by keeping only the rows and columns in I; we will call such an $A(I)$ the I *submatrix* of A, and we say it is *proper* if $I \neq \{1,\ldots,n\}$. Similarly for an n-component vector v, we write $v(I)$ for the I subvector obtained by keeping only the components in I.

Definition 17 *For a square matrix B, we write B_i for the matrix obtained when column i of B is replaced by all "ones". We write $\mathscr{V}$ for the set of all $n\times n$ matrices A, such that every submatrix $B = A(I)$ satisfies the following conditions:*

1. $\det B \neq 0$
2. $\det B_i \neq 0$ *for all $i \in I$.*
3. *If B is a* proper *submatrix, and all* $\det B_i$ *have the same sign, i.e.*

$$\frac{\det B_i}{\det B_j} > 0 \text{ for all } i, j \in I;$$

then for $\mathbf{x}(B)$, $v(B)$ as in (8), there is some $i' \notin I$ such that

$$r\left(i', I\right) \cdot \mathbf{x}(B) > v(B),$$

where $r(i', I)$ is the I-subvector of the i'-th row of A.

Let $\overline{\mathscr{V}}$ denote the closure of $\mathscr{V}$ in the Euclidean topology.

Proposition 18 *We have $\mathscr{V} \subset \mathscr{S} \subset \overline{\mathscr{V}}$.*

Proof. See the Appendix. ■

Proposition 18, as was said, justifies our view of $\mathscr{V}$ as a proxy for $\mathscr{S}$.

Remark 19 *Note that $\mathscr{V}$ is an open set, and both $\mathscr{V}$ and $\overline{\mathscr{V}}$ are* semialgebraic[9]. *It is natural to ask:* Is $\mathscr{S}$ also a semialgebraic set?

[9] A semialgebraic set is a subset of $\mathbb{R}^n$ defined by a finite collection of polynomial equations and inequalities, or a finite union of such sets.

3.7 A numerical example

Consider the matrix

$$A = \begin{bmatrix} 20-x & 3-x & 18-x \\ 24 & 9 & 3 \\ 6+y & 12+y & 8+y \end{bmatrix}$$

The contract (x,y) transfers wealth from the worker W to the policeman P but, for simplicity, the alterations in their payoffs are taken to be unconditional: x denotes a flat tax imposed on W and y a flat subsidy given to P. Suppose that more cannot be taken from W than he has, i.e., $x \leq 3$. Denote

$$v = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ and } h = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}.$$

and define

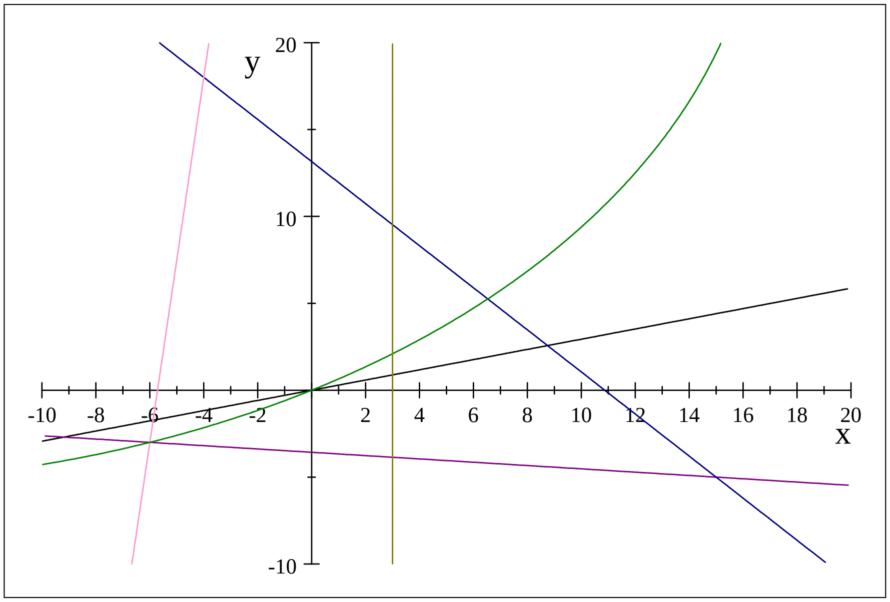
$$w = \left(hA^{-1}v\right)^{-1} = \frac{327}{445}y - \frac{96}{445}x + \frac{882}{89},$$

$$p = A^{-1}vw = \begin{bmatrix} \frac{2}{445}x + \frac{21}{445}y + \frac{15}{89} \\ \frac{50}{89} - \frac{19}{445}y - \frac{23}{445}x \\ \frac{21}{445}x - \frac{2}{445}y + \frac{24}{89} \end{bmatrix},$$

and

$$c = xp_1 - yp_3 = \frac{2}{445}x^2 + \frac{15}{89}x + \frac{2}{445}y^2 - \frac{24}{89}y$$

It is readily checked that, contingent on (x,y) being the contract, p is the NE, w is the welfare, and c is the budget constraint (see equation (6)). The following picture summarizes the entire situation.



The interior of the triangle bound by the pink, purple, and blue lines defines the set of contracts (x, y) which have (unique) completely mixed SNE. The green curve depicts the set of self-financing contracts, i.e., those for which the budget constraint equation (6) holds. Finally, the black line through the origin is a level set of w (for $w = 0$), whose value increases as we parallel shift it upwards; while the vertical brown line represents the non-negativity constraint $x \leq 3$ on interaction payoff.

Thus the maximum of w, in the feasible set of contracts, appears at the intersection of the green and brown lines, and gives the optimal levels of crime and punishment in the society. It is worth noting that if the constraint $x \leq 3$ were to be dropped, the exact optimal (at the intersection of the green curve with the blue side of the triangle) would become infeasible. However ε-optimal contracts would exist for all $\varepsilon \to 0$, and would converge to the exact optimal, with crime disappearing in the limit! (This is only so in the example we have constructed, and there is no reason to expect it to hold in general. If, for instance, P were to be paid only in PC encounters where P performs yeoman service, there would have to be positive levels of crime and punishment at the optimum.)

4 Towards a Scheme for General Noncooperative Games

We present a scheme for extending the analysis of section 2 to general noncooperative games, which we hope may be of use in formulating specific models in future applications.

Returning to the notation of section 2, let

$$\theta : S \to T$$

be a map from srategies to outcomes, and let

$$u^i : T \longrightarrow \mathbb{R}$$

denote the (cardinal) utility of player $i \in N$ on the outcome space T, so that i's payoff is $\pi^i(s) = u^i(\theta(s))$ for $s \in S$. Since we shall hold the utilities $\left(u^1, \ldots, u^n\right)$ fixed throughout and vary only the outcome function θ, the mixed-strategy extension of the above game will be denoted Γ_θ.

(Thus, in our example of games of money with risk neutrality, we have $T = \mathbb{R}^N$ and $u^i(z) = z_i$ for every $z = (z_1, \ldots, z_n) \in T$.)

A social contract specifies alterations in the outcomes from the original θ and is represented by a map

$$\tau : S \longrightarrow T$$

Thus, once the contract τ is in place, the outcome is altered from $\theta(s)$ to $\tau(s)$ for any $s \in S$, i.e.,the game changes from Γ_θ to Γ_τ.

Let Ω denote the collection of all maps from S to T , and $\Omega(\theta) \subset \Omega$ the alterations of outcome that are *a priori* available (when the original situation, or staus quo, is given by θ). No alteration leave us with the original game, thus we must have:

Axiom 20 $\theta \in \Omega(\theta)$ *(i.e. the status quo is always available)*

Since we do not have money, there is no notion of "transfers" between players, leave aside that of self-financing transfers. We now take, as abstractly given, not only the notion of an alteration τ (from the status quo θ) but also that of the "sustainability" of τ in conjunction with mixed-strategies $x \in X$.

Notation 21 *Let* $\Lambda(\theta) \subset \Omega \times X$ *be a collection of* self-sustaining *pairs* (τ, x), *where* τ *is an outcome function (altering the status quo* θ) *and* $x \in X$ *is a choice of mixed strategies.*

Since any $x \in X$ is admissible as a mixed strategy in the game Γ_θ, we must also require:

Axiom 22 *$(\theta, x) \in \Lambda(\theta)$ for all $x \in X$.*

Now we proceed, as in section 2, to develop the definition of viable pairs and of a preorder $\succeq$ on such pairs.

Definition 23 *The pair (τ, x) is* viable *if (i)$\tau \in \Omega(\theta)$; (ii)$(\tau, x) \in \Lambda(\theta)$; (iii) x is an NE of Γ_τ.*

Definition 24 *Given two viable pairs (τ, x) and (τ', x'), we write $(\tau, x) \succeq (\tau', x')$ if*

$$\sum_{s \in S} \Pr(s,x)\left[u^i(\tau(s))\right] \geq \sum_{s \in S} \Pr(s,x')\left[u^i\left(\tau'(s)\right)\right] \text{ for all } i \in N$$

By Nash's theorem, and by Axioms 20 and 22, the set of viable pairs is non-empty. Next we define

Definition 25 *A contract τ is* optimal *if there exists $x \in X$ such that (τ, x) is $\succeq$-maximal in the set of viable pairs.*

And, again as before, we may define

Definition 26 *A contract τ is ε-*optimal *if there exists $x \in X$ such that (τ, x) is viable, and there does not exist any viable pair (τ', x') such that*

$$\sum_{s \in S} \Pr(s,x')\left[u^i\left(\tau'(s)\right)\right] - \varepsilon \geq \sum_{s \in S} \Pr(s,x)\left[u^i(\tau(s))\right] \text{ for all } i \in N$$

Proposition 27 *Suppose the set $\left\{\left(u^1(t), \ldots u^n(t)\right) : t \in T\right\}$ is bounded. Then for any $\varepsilon > 0$, there exists an ε-optimal contract.*

Proof. This is obvious. ■

For the existence of an exact optimal, we shall need to assume that T is a metric space, and that u^i are continuous, and that — in addition to the boundedness hypothesis of Proposition 27 — the set $\Omega(\theta)$ is closed in the appropriate sense. We leave all details to the reader.

4.0.1 Competing Social Contracts?

It is tempting to think of different contracts that compete for adoption by a given society (or, alternately, of multiple societies, that have adopted different contracts and are competing amongst themselves). We do not have a "meta-model" of this kind; but if such were to be formulated, one might intuitively expect that contracts which are not self-financing, or which are Pareto-dominated, would fall by the wayside, leading to the survival of those that we have called "optimal."

5 Appendix

5.1 Proof of Proposition 15

Proof. Let $N \in \mathscr{S}$. Then, by definition, $N \in \mathscr{N}(M)$ if and only if

$$\mathbf{x}_N^T (N - M) \mathbf{x}_N = 0$$

where $\mathbf{x}$ denotes the unique CMSNE of N. But by Proposition 14 $\mathbf{x}$ is proportional to $C_N^T \mathbf{1}$. Thus we may substitute $C_N^T \mathbf{1}$ for $\mathbf{x}$ in the above equation to get the equivalent condition

$$\left(C_N^T \mathbf{1}\right)^T (N - M) \left(C_N^T \mathbf{1}\right) = 0$$

i.e.,

$$\mathbf{1}^T C_N (N - M) C_N^T \mathbf{1} = 0$$

i.e., recalling that $\mathbf{1}^T A \mathbf{1} = \sigma(A)$ for any square matrix A,

$$\sigma(C_N (N - M) C_N^T) = 0$$

Next, if $N \in \mathscr{N}(M)$, then by the budget balance condition (6), with $\mathbf{x} = \mathbf{x}_N$, we get $\mathbf{x}^T (N - M) \mathbf{x} = 0$ and hence

$$\mathbf{x}^T M \mathbf{x} = \mathbf{x}^T N \mathbf{x} = \mathbf{x}^T (v \mathbf{1}) = v \mathbf{x}^T \mathbf{1} = v$$

Since the left side is a weighted sum of the $M_{i,j}$, formula (9) follows. ■

5.2 Proof of Lemma 16

Proof. Suppose N is non-invertible but an S-matrix. Then N has a null vector $\mathbf{y} \neq \mathbf{0}$. If $\mathbf{x}$ (the unique CMSNE of N) is not proportional to $\mathbf{y}$ then adding a small multiple of $\mathbf{y}$ to $\mathbf{x}$ and renormalizing we get a different SNE, contradicting the uniqueness of $\mathbf{x}$. Thus $\mathbf{x}$ must be proportional to $\mathbf{y}$, hence $\mathbf{x}$ is a null vector as claimed, and now the equation $N\mathbf{x} = v\mathbf{1}$ implies $v = 0$. ■

5.3 Proof of Proposition 18

Proof. First let us show that $\mathscr{V} \subset \mathscr{S}$. Let $A \in \mathscr{V}$. By Nash's theorem (see [3]) there exists an SNE $\mathbf{x}$ for A. Suppose the support of $\mathbf{x}$ is a *proper* subset $I \subset \{1, \ldots, n\}$. Then by Conditions 1 and 2 of Definition 17, the proper submatrix $A(I)$, as well as the matrices $A_i(I)$ for $i \in I$, are all invertible. Hence, by Proposition 14, we have that $\mathbf{x} = \mathbf{x}(A(I))$ and that the payoff to all pure strategies of $A(I)$ at the SNE $\mathbf{x}(A(I))$ is $v(A(I))$. But by Condition 3 of Definition 17, there exists a row $i' \notin I$ such that

$$r\left(i', I\right) \cdot \mathbf{x}(A(I)) > v(A(I)).$$

This means that if a player were to unilaterally deviate to the pure strategy i' instead of playing the mixed strategy $\mathbf{x}(A(I))$, while his opponent stayed put at $\mathbf{x}(A(I))$, he would get a higher payoff in the matrix game A, contradicting that $\mathbf{x}(A(I))$ is an SNE of A. Thus the SNE $\mathbf{x}$ of A cannot have support on a proper subset of $\{1, \ldots, n\}$. We conclude that $\mathbf{x}$ is completely mixed, i.e., $\mathbf{x}$ is a CMSNE. Now, since A is invertible by Condition 1 of Definition 17, it follows again from Proposition 14 that there is no other CMSNE of A besides $\mathbf{x}$. This proves that $A \in \mathscr{S}$.

Next we need to show that $\mathscr{S} \subset \overline{\mathscr{V}}$. Suppose, to the contrary, that there exists $A \in \mathscr{S}$ and an open set $\mathscr{O}$ (in the Euclidean space of all $n \times n$ matrices) such that $\mathscr{O} \cap \mathscr{V}$ is empty. Since Conditions 1 and 2 of Definition 17 clearly hold for an open and dense set of matrices, there exists a sequence of matrices $A_l \rightarrow A$ as $l \rightarrow \infty$, such that each A_l satisfies both those conditions. Thus we may suppose (by going to a subsequence if necessary) that each A_l violates condition 3 of Definition 17 (otherwise each $A_l \in \mathscr{V}$ and hence $A \in \overline{\mathscr{V}}$, a contradiction). But then there exists a proper submatrix B_l of A_l such that a CMSNE $\mathbf{x}(l)$ of B_l is an SNE of A_l, for all l. By selecting further subsequences if necessary, we may also suppose that $\mathbf{x}(l) \longrightarrow \mathbf{x}$ and that $B_l \longrightarrow B$, where all the B_l and B are $I \times I$ matrices for some proper subset $I \subset \{1, \ldots, n\}$. But then by the upper-hemicontinuity of NE, $\mathbf{x}$ is a SNE of A that is not completely mixed (in fact $\mathbf{x}$ is a CMSNE of B). This contradicts that $A \in \mathscr{S}$. ∎

5.4 Proof of Proposition 13

We first establish

Lemma 28 *If a 3×3 nonnegative matrix M has the property (4), then M is invertible.*

Proof. Let M be singular. Then there exist α, β, γ, not all 0, such that

$$\alpha \begin{bmatrix} b \\ c \\ a \end{bmatrix} + \beta \begin{bmatrix} d \\ e \\ f \end{bmatrix} + \gamma \begin{bmatrix} i \\ g \\ h \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Suppose one of the α, β, γ is 0, say $\gamma = 0$. Then α, β must have opposite signs. Therefore:

$$\begin{bmatrix} b \\ c \\ a \end{bmatrix} = \kappa \begin{bmatrix} d \\ e \\ f \end{bmatrix}$$

for some positive constant κ. Since $d < f$, we would deduce $b < a$, which contradicts that M satisfies (4). So we may assume that all $\alpha, \beta, \gamma \neq 0$. But they cannot all have the same sign since $M \geq 0$. Therefore w.l.o.g. we may assume that one of them is negative, in fact -1, and the other two positive. Say $\alpha = -1$. Then we get

$$\begin{bmatrix} b \\ c \\ a \end{bmatrix} = \beta \begin{bmatrix} d \\ e \\ f \end{bmatrix} + \gamma \begin{bmatrix} i \\ g \\ h \end{bmatrix}$$

with $\beta, \gamma > 0$. This implies

$$c = \beta e + \gamma g \tag{10}$$

and

$$a = \beta f + \gamma h \tag{11}$$

By the property (4) of M, LHS of (10) $<$ LHS of (11), and also RHS of (10) $>$ LHS of (11), which is a contradiction. Similarly $\beta = -1$, or $\gamma = -1$, lead to contradictions. ■

Remark 29 *The conclusion of lemma (28) can be strenthened to: determinant $M > 0$. Indeed let Π_+ denote the set of all nonnegative matrices with the property (4). Clearly the determinant is continuous on Π_+ and, by Lemma (28), it is nonvanishing on Π_+. Since Π_+ is path-connected, indeed convex, it has the same sign throughout Π_+. Take any M in Π_+ and let $c = f = i \longrightarrow \infty$ keeping the rest of the entries of M fixed. We remain in Π_+ throughout and the determinant goes to infinity. This proves its positivity.*

Now we complete the proof of Proposition 13

Proof. Adding the same constant to every entry of M does not disturb the Nash Equilibria (NE), or the property (4), of M; so w.l.o.g. we may assume $M \geq 0$.

We first argue that any NE of M, symmetric or not, must be completely mixed. Let $x = (x_W, x_C, x_P)$ be the mixed strategy of a player at some NE. If x ascribes 0 probability to some pure strategy α then, in the remaining two pure strategies, one strictly dominates the other by property (4), so its probability must be 1. But then, again by property (4), it in turn is strictly dominated by α, therefore the probability of α must be 1, a contradiction. Thus $x = (x_W, x_C, x_P) >> 0$ and the NE is completely mixed.

Note that, by Lemma 28, M is invertible. Then, by Proposition 14, we see that the NE is unique and (perforce) symmetric. ∎

References

[1] G.S.Becker (1974). Crime and Punishment: an Economic Analysis, *Essays in the Economics of Crime and Punishment,ed. G.S.Becker & W.M.Landes, National Bureau of Economic Research, Vol. ISBN 0-87014-263-1*

[2] P. Dubey & M. Kaneko (1984). Information Patterns and Nash Equilibria in Extensive Games, *Mathematical Social Sciences, Vol. 8, No. 2, pp. 111-139.*

[3] J.F.Nash (1951). Non-Cooperative Games, *Annals of Mathematics, Vol. 54, No. 2, pp 286-295.*

[4] M. Smith & G.R. Price (1973). The Logic of Animal Conflict, *Nature, 246, pp. 15-18.*