Contingent Transfers as an Incentive for Cooperation in Noncooperative Games

Pradeep Dubey*and Siddhartha Sahi†

18 March 2024

Dedicated to the memory of K.R.Parthasarathy[1]

Abstract

Consider a noncooperative game at whose outcomes commodities accrue to players, which are both valued by them and susceptible to transfer between them. In this situation, outcome-contingent transfers of commodities form a natural schema for incentivizing cooperation, in the following sense. Transfers agreed upon by the players give rise to a new noncooperative game whose Nash Equilibria (NE) may engender a Pareto-improvement over some designated *status quo* NE of the original game.

However, as in the folk theorem for repeated games, an embarrassingly large set of NE may be sustainable via transfers. The main source for this pathology is the possibility of "threats", which can be understood as "off-shell" transfers, *i.e.* transfers at outcomes that are not actually being reached with positive probability at the NE under consideration. Our endeavor is to restore the discriminatory nature of NE by means of two simple ideas. We say that an NE of the post-transfers game is *transparent* if there are no off-shell transfers. This can also be viewed via the lens of "credibility", of something being seen in order to be believed. The

---

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

†Department of Mathematics, Rutgers University, New Brunswick, New Jersey; and Center for Game Theory, Department of Economics, Stony Brook University

[1] The authors have vivid memories of their meetings with Professor K.R.Parthasarathy, especially his succinct and insightful observations on a wide range of topics, including Game Theory.

other condition is *budget balance,* i.e., at the NE, there should be no need for a *net* injection of commodities from the outside in order to sustain transfers. An NE is considered *eligible* by us if it satisfies these two conditions, in addition to engendering Pareto-improvement. If a social welfare function is specified, then one may seek to determine eligible NE that maximize welfare on the domain of feasible transfers. We carry out this analysis for three classical noncooperative games. These are the Centipede Game, Contest, and the Prisoners' Dilemma; the last of which we discuss in some detail.[2]

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

Keywords: Games of Income**,** Nash Equilibrium, transfers, transparency, budget balance, gains to transfers, Prisoners' Dilemma.

1 Introduction

The notions of strategy and payoff are fundamental to the theory of noncooperative games, and suffice to define Nash Equilibrium ([9]). However in many classes of games in the economic literature there is an intermediate notion of "income" — often of interest for its own sake — from which the payoff is derived through a utility function. Income typically consists of commodities that are valued by players and susceptible to transfer between them. The canonical example is of course Cournot oligopoly ([1]). Here there is a single commodity called "money" and each firm's payoff is linear in the money that it earns.

In consonance with this literature, we consider a game with the following features. There is commodity income that accrues to players, contingent upon the state of the world, or "outcome", realised in the game. The outcomes could be random, but their probability distribution depends on the strategies selected by the players. Moreover, we assume that while outcomes — and the commodity transfers made there — are publicly observable, the strategies themselves are private knowledge to the players.[3]

---

[2] This paper updates, extends, and supersedes an earlier version [2].

[3] Consider, for example, a stochastic generalization of Cournot oligopoly, where each of n firms produces random output conditional on its deterministic investment of money; and where its sales revenue is also random, since the price at which the output sells is determined by the total output of all firms and an exogenously fixed market demand function. The investment is private information to the firm, but its output is on market and publicly known, as is the market demand. The outcomes here correspond to the n-tuples of firms' outputs, the probability distribution on outputs is determined by the n-tuple of firms' strategies (i.e., investments), and finally the outcome-contingent income of each firm is the money it makes from sales. More abstractly, our model

In this scenario, and assuming that players are in an NE of the original game which serves as their *status quo*, what would be the means available to them if they wanted to cooperate for mutual benefit? Let us suppose that not only is preplay communication possible between players, but that they have access to a "court" under whose auspices they may sign a contract as to what they will do. It is evident that they cannot contract to play specific strategy selections, since these are unobservable, and hence unverifiable, by the court. Think of the case where many different strategy selections, each characterized by its own distinct set of "spoilsport" players, give rise to the same bad outcome. Then how can the court infer from the outcome who the culprits are that need to be taken to task?

Any contract must be limited to what is observable by the court, i.e., outcome-contingent commodity transfers. In our model, all commodity incomes are fixed, with no scope for enhancing them. Thus the only course of action open to the players — provided no prohibition is placed upon it — is to contract *transfers* of commodities at the court[4], thereby *transforming the game they have to play*. If this maneuver leads to utility gains for all, via the emegence of "better" NE in the transformed game, then players will be incentivized — indeed, often impelled — to innovate these contracts, especially if the gains are significant. (Note that there is *no* transfer of utiles being contemplated here, only the transfer of commodities, which of course does affect utilities.)

A typical instance we have in mind occurs when certain strategy selections lead to a huge amount of commodities in total, but with a skewed distribution, so that some players get very little. Such players will tend to play safe strategies that preclude the huge total, but guarantee them better payoffs. In order to get them to fall in line for the huge total, it becomes needful for the big winners to transfer sufficient commodities to them so that none has any desire to deviate. Of course, as we shall see in the examples below, there can be subtler reasons for benefits from transfers, based upon the full chain effect of the strategic interaction among the players, *i.e.*, upon NE.

It is worth emphasizing that, despite the element of cooperation introduced by transfers, the noncooperative character of the game is left intact. Strategies chosen privately by the players, upon which outcomes depend, must perforce arise in an endogenous manner, with each player acting in his own self-interest, in the

---

includes extensive form games ([7]) in which payoffs are derived fom utility of income at the terminal nodes. To the extent that games of incomplete information may be viewed as extensive games of imperfect information ([4]), these are also included.

[4] The court — tasked to implement transfers — is of course necessary, since transfers cannot possibly be "self-enforcing", for who would ever voluntarily part with valuable commodities?

strategic competition that ensues once transfers are put into place. In the parlance of game theory, players' strategies must constitute an NE of the transformed game which results from the transfers. In general any such game may have multiple NE. Since we wish to remain agnostic about which NE the players will settle upon, the objects of study in this paper are *pairs* (τ, σ), where τ denotes outcome-contingent transfers and σ is an arbitrary NE of the post-transfers game induced by τ.

A crucial condition that we place, upon the pair (τ, σ), is as follows: *transfers τ must be restricted to outcomes that occur with positive probability at the NE σ;* in short, they must occur in the "play"[5] of the game generated by σ. This is a joint condition on τ and σ. The idea behind it is to render transfers "*transparent*"(or "factual", disregarding completely all counterfactual transfers at outcomes that do not occur in the play of σ). Transparency of τ in (τ, σ) lends more credence to τ. Think of an NE σ as the prevalent and persistent mode of behavior in the game generated by τ (all else being rendered ephemeral by players' strategic deviations). Then transfers in τ that are "on shell", *i.e.*, witnessed within the play of σ, acquire the imprimatur of "legal precedence" (in the court tasked to implement τ) and thus become more credible to the players. To put it bluntly, transfers must be "seen to be believed" and those that are "out of sight are out of mind". Also note that, were the transparency condition to be lifted, the discriminatory power of NE would be quite lost. Indeed, as in the "folk theorem", a plethora of outcomes, accompanied by the original payoffs of the pre-transfers game, could then be supported as NE by means of drastic conjectural transfers that are not subjected to a reality check via the NE play. (See Remark 1 in section 2.4.)

Transfers can be *intra*-outcome, with commodities exchanged between players at every outcome. These are "simple" transfers and, for ease of presentation, form the focus of all our examples. However, in the general model we make room for the much broader category of *inter*-outcome transfers, in which players may (collectively) give up some commodities at certain outcomes in order to get them at other outcomes[6]. To sustain such transfers, we need to postulate — again in the background, like the court — a risk-neutral agency, e.g., a "bank"[7], which is willing to lend select commodities (such as money) to players at zero interest rate provided it recoups the loan. In the light of risk-neutrality, this is tantamount to

---

[5] i.e., the set of outcomes that occur with positive probability under σ.

[6] Inter-outcome transfers are the center of attention in our forthcoming companion paper. For a preliminary version, see [2].

[7] Or, perhaps, even one of the players within the game who is endowed with sufficient wealth as to be risk-neutral.

the *budget balance* condition that the expected profit of the bank be zero for every commodity-loan.[8] Notice that, like transparency, this too is a *joint* condition on the pair (τ, σ): the strategy selection σ must be an NE of the game determined by transfers τ, and simultaneously τ must be budget-balanced based upon the probability distribution of outcomes generated by σ.

Thus our domain consists of pairs (τ, σ) which satisfy the twin conditions of transparency and budget balance, along with the "individual rationality" constraint that σ yield to each player no less than his *status quo* payoff. Our focus is on the set of pairs that are Pareto-optimal with respect to some exogenous set of feasible transfers. This set is clearly invariant under independent affine transformations of players' utilities, i.e., it is "cardinal" in character. Now if a "social welfare function" is specified, then one may select pairs which maximize welfare on the Pareto set . In this event, we define the *gains to transfers* as the gain in welfare achieved by the selection. One could think of it as the *incentive to cooperate* inherent in the noncooperative game. This notion might be useful both normatively, from the point-of-view of a policy-maker trying to steer the game towards welfare gains; and also descriptively, as an explanatory variable that correlates with players' deviation from NE behavior observed in empirical data. (Needless to say, it is *not* a cardinal notion, since welfare willy-nilly entails interpersonal comparison of utilities.)

In conclusion, let us note that transfers have long been at the center of attention in the analysis of cooperative games. Indeed the Core and the Shapley Value, as well as several other solution concepts, are all concerned with how the fruit of full cooperation is to be transferred to the players. It seems to us both natural and worthwhile to also consider transfers in a noncooperative setting. We are not aware of too much research in this direction. The references [5], [10], [12], [15] are in the context of transferable-utility models of bargaining with strategic threats, and do not have much overlap with the model here.

The paper is organized as follows. After presenting our abstract model of transfers, we bring it to bear on three examples: the Centipede Game, Contest and (most elaborately) the Prisoners' Dilemma. In all of these — for simplicity, and in the footsteps of Cournot — players get a single commodity called money at every outcome and all players are presumed risk-neutral[9]; furthermore, social welfare

---

[8] In Section 7, we consider the case where the bank may charge positive interest or even seek to maximize expected profit.

[9] These assumptions may give the illusion of "transferable utility"(TU). Indeed, with a single commodity money and risk-neutral utilities, our model is *mathematically* isomorphic to the TU model. However, *conceptually* it stands apart: transfers are postulated for money alone, and

is taken to be “egalitarian” (sum of players’ utilities). Even in this simplest of frameworks, the computation of gains to transfers can be quite subtle. In particular, since we view payoffs as cardinal and not ordinal (thereby admitting mixed strategies in our analysis), there is a large domain of Prisoners’ Dilemma games. The gains to transfers vary widely across this domain, enabling us to demarcate regions where they are pronouncedly high (or, low). In the process, we show that *reducing money in the game can often be of benefit to both players when transfers are allowed* (even when that reduction is detrimental to them when transfers are forbidden).

For the most part, we restrict to deterministic transfers, although it is straightforward to accomodate *random transfers* in our model. However, a special kind of randomization is briefly discussed in our analysis of the Prisoners’ Dilemma: nature picks, and fully reveals to the players, different games for them to play, each of which is induced by deterministic transfers. This has the advantage of restoring symmetry to the solution of a symmetric problem, by randomizing over deterministic solutions, each of which necessarily entails symmetry-breaking.

2 The General Model

2.1 Games of Income

We formalize the class of noncooperative games discussed in the introduction. For simplicity we make a number of finiteness assumptions which can be relaxed in the usual manner.

We consider the following setup. Let N and K be finite sets, of *players* and *commodities*, respectively. For each player $i \in N$ let Σ^i be a set, of *strategies*, and let $\Sigma = \prod_{i \in N} \Sigma^i$ be the Cartesian product. Also let Ω be a finite set, of *outcomes*, and let $\Delta(\Omega)$ be the collection of probability distributions on Ω.

A *noncooperative* game of income $\Gamma = \Gamma(\varphi, \mu, u)$ is specified by a triple of maps

$$\varphi : \Sigma \longrightarrow \Delta(\Omega), \quad \mu : N \times \Omega \longrightarrow \mathbb{R}^K, \quad u : N \times \Omega \times \mathbb{R}^K \longrightarrow \mathbb{R} \tag{1}$$

The vector $\mu(i, \omega) \in \mathbb{R}^K$ is the *income* of commodities that accrues to player i at ω, and $u^i_\omega(z) = u(i, \omega, z)$ is the *payoff,* or *utility,* that i derives from consumption of z at ω. The game is defined as follows. A *strategy selection* $\sigma = \left(\sigma^i\right)_{i \in N} \in \Sigma$

---

changes in utiles are pure byproducts. (That they seem identical is an artifact of risk-neutrality).

of the players leads to the probability distribution φ_σ on Ω. The *expected payoff* of i at σ is given by

$$\pi^i(\sigma) = \sum_{\omega\in\Omega} \varphi_\sigma(\omega)\, u^i_\omega(\mu(i,\omega)). \tag{2}$$

Given $\sigma \in \Sigma$ and $s \in \Sigma^i$, we denote by $(\sigma \mid_i s) \in \Sigma$ the strategy selection which is the same as σ except that σ^i is replaced by s. We say (see [9]) that $\sigma_\#$ is a *Nash Equilibrium* (NE) of Γ if

$$\pi^i(\sigma) \geq \pi^i((\sigma \mid_i s)) \text{ for all } i \in N \text{ and } s \in \Sigma^i.$$

2.2 Transfers and Eligible Nash Equilibria

A *transfer* (of commodities) in Γ is a map

$$\tau : N \times \Omega \longrightarrow \mathbb{R}^K$$

Such a τ defines a new income map $\mu + \tau$; we write $\Gamma_\tau = \Gamma(\varphi, \mu + \tau, u)$ for the modified game, and denote by π^i_τ its payoff functions as in (2).

A *pair* (σ, τ), consisting of strategy selection $\sigma \in \Sigma$ and a transfer τ, is called *transparent* if, for all $\omega \in \Omega$,

$$\tau(i,\omega) \neq 0 \text{ for some } i \implies \varphi_\sigma(\omega) > 0;$$

and *balanced* if it satisfies commodity conservation in the sense of expectation:

$$\sum_{\omega\in\Omega} \left[\varphi(\omega,\sigma) \sum_{i\in N} \tau(i,\omega) \right] = 0 \tag{3}$$

A transfer is said to be *simple* if commodity conservation holds at each outcome:

$$\sum_{i\in N} \tau(i,\omega) = 0 \text{ for all } \omega \in \Omega$$

It is clear that if τ is simple then (σ, τ) is balanced for all $\sigma \in \Sigma$.

Let σ^* be an NE of Γ – the *status quo*. We say that (σ, τ) is σ^**-dominant* if

$$\pi^i_\tau(\sigma) \geq \pi^i(\sigma^*) \text{ for all } i \in N.$$

If (σ, τ) is transparent, balanced and σ^*-dominant we say that it is σ^**-eligible.*.

2.3 Optimal Transfers

Let Γ and σ^* be as above, and let T be an exogeneously given set of "permissible" transfers. Define

$$\mathscr{E}(\sigma^*, T) = \{(\sigma, \tau) \in \Sigma \times T : (\sigma, \tau) \text{ is } \sigma^*\text{-eligible and } \sigma \text{ is an NE of } \Gamma_\tau\},$$

and the corresponding set of payoff vectors

$$Z(\sigma^*, T) = \left\{\pi_\tau(\sigma) \in \mathbb{R}^N : (\sigma, \tau) \in \mathscr{E}(\sigma^*, T)\right\}.$$

We shall assume throughout the set T satisfies two conditions

$$0 \in T \tag{4}$$

$$Z(\sigma^*, T) \text{ is bounded} \tag{5}$$

We note that (4) implies that $(\sigma^*, 0) \in \mathscr{E}(\sigma^*, T)$ and hence $\mathscr{E}(\sigma^*, T)$ and $Z(\sigma^*, T)$ are nonempty. As we shall see shortly (5) implies the existence of NE that are "approximately optimal", either in the Pareto sense or with respect to a social welfare function.[10]

2.3.1 Pareto Optimal Transfers

We say a vector u dominates a vector v in $\mathbb{R}^N$ if $u_i \geq v_i$ for all i, and that it ε-dominates v for some $\varepsilon > 0$ if $u_i \geq v_i + \varepsilon$ for all i. For a set $V \subset \mathbb{R}^N$ we define its *Pareto frontier* PV and *ε-Pareto frontier* $P_\varepsilon V$ to be those vectors that are not dominated and not ε-dominated, respectively, by any other vector in V.

Lemma 1 *If $V \subset \mathbb{R}^N$ is nonempty and bounded then $P_\varepsilon V$ is nonempty for any $\varepsilon > 0$. If V is compact then PV is nonempty, although it need not be compact.*

Proof. First suppose V is bounded and $\varepsilon > 0$. Then the function $\Sigma(u) = u_1 + \cdots + u_N$ has a supremum s on V and we can find v in V with $\Sigma(v)$ arbitrarily close to s. In particular we can find v such that

$$\Sigma(v) > s - |N|\varepsilon$$

---

[10] The boundedness of $Z(\sigma^*, T)$ can be deduced from explicit conditions on the exogenous data T, u^i of the model, e.g., each u^i has finite supremum on its entire domain; or, alternatively, T is bounded and each u^i has finite supremum on any bounded subset of its domain.

If some u were to ε-dominate v then we would have $\Sigma(u) \geq \Sigma(v) + |N|\varepsilon > s$, which is a contradiction. Thus v belongs to $P_\varepsilon V$ and thus $P_\varepsilon V$ is nonempty.

If V is compact then any v at which Σ attains its supremum belongs to PV. Finally, the following example shows that PV need not be compact.

$$V = \left\{(x,y) \in \mathbb{R}^2_+ : x+y=1\right\} \cup \{(0,2)\} \implies PV = V \setminus \{(0,1)\}.$$

■

Since $Z(\sigma^*, T)$ is nonempty and bounded, Lemma 1 immediately implies

Proposition 2 *$P_\varepsilon Z(\sigma^*, T)$ is nonempty for any $\varepsilon > 0$.*

Note that, since transparency is not a closed condition, $\mathscr{E}(\sigma^*, T)$ and $Z(\sigma^*, T)$ need not be closed, and therefore the Pareto frontier $PZ(\sigma^*, T)$ may be empty, which is what led us to consider the ε-Pareto frontier $P_\varepsilon Z(\sigma^*, T)$.

2.3.2 Gains to Transfers

A (social) *welfare function* is a map $w : \mathbb{R}^N \longrightarrow \mathbb{R}$ which is strictly increasing in each coordinate. Given such a w, we define *gains to transfers* at $\sigma \in \mathscr{E}_{\sigma^*}(\tau)$ to be

$$G_w(\sigma^*, \sigma, \tau) = w(\pi_\tau(\sigma)) - w(\pi(\sigma^*)).$$

and the gains to transfers on T by

$$G_w(\sigma^*, T) = \sup\{G_w(\sigma^*, \sigma, \tau) : (\sigma, \tau) \in \mathscr{E}(\sigma^*, T)\}.$$

Proposition 3 *If w is bounded on $Z(\sigma^*, T)$ then $G_w(\sigma^*, T)$ is finite; moreover for any $\varepsilon > 0$ there exists $(\sigma, \tau) \in \mathscr{E}(\sigma^*, T)$ such that*

$$G_w(\sigma^*, \sigma, \tau) \geq G_w(\sigma^*, T) - \varepsilon.$$

Proof. Since w is bounded, $G_w(\sigma^*, T)$ can be approximated arbitrarily closely in $\mathscr{E}(\sigma^*, T)$. ■

In the subsequent sections we will study G_w in greater detail for certain classical games and the "egalitarian" welfare w.

2.4 Remarks on the Model

(1) (**Transparency**) Consider games of income in normal form as in [9], where outcomes are identified with pure strategy selections. Without the transparency condition, *any* pure strategy selection $\sigma_{\#} = \left(\sigma_{\#}^{i}\right)_{i \in N} \in \Sigma$ can be sustained as an NE via transfers. Indeed at every $\left(\sigma_{\#} \mid_i \sigma^i\right)$, for $i \in N$ and $\sigma^i \in \Sigma^i \diagdown \left\{\sigma_{\#}^{i}\right\}$, let a *lot* of commodities[11] be transferred from i to other players. Then no i will have incentive to unilaterally deviate from $\sigma_{\#}$, with the upshot that $\sigma_{\#}$ will be an NE. Indeed, by such punishments for deviations, a large number of NE can be generated in any given game via a single outcome-contingent transfer. For instance, in an $n \times n$ bimatrix game, any n entries such that no two lie in the same row or column, may be sustained as NE. In general, in a multimatrix game with k players, the number of NE pure-strategy selections grows exponentially with k. For instance, consider binary games where each of the k players has two pure strategies. Then to sustain any strategy selection as NE, we need to punish k unilateral deviations from that NE, one for each player. Thus we may create at least n NE where $n = 2^k - nk$, *i.e*, $n = 2^k / (1+k)$. In general, we have $n = (l_1 \times \ldots \times l_k) / (1 + l_1 + \ldots l_k - k)$ where l_i is the number of pure strategies of i.

Our notion of transparency is reminiscent of "rational expectations equilibrium" (transfers anticipated conform to transfers realized). and even more of "self-confirming equilibrium" (SCE) in extensive form games (see [3]). In SCE players are permitted to hold arbitrary, idiosyncratic conjectures regarding each others' strategies outside of the NE play, but within the NE play their conjectures must concur with the "objective reality" of what is going on there. (Needless to say SCE further requires that no player must want to to unilaterally deviate, in the light of his conjectures.) However, note also a difference. Our transparency condition says that *no* transfers may be conjectured at outcomes outside of the NE play, i.e., transfers that are "out of sight are out of mind".

Transparency can be of wide-ranging application, going well beyond the games of income considered here. Think of abstract, outcome-contingent "consequences" and suppose there exist (verifiable and contractable) "alterations" in the consequences that affect utiles and thereby the game. Then transparency can be defined for alterations in general, exactly as it is for transfers in our model.[12]. Even in

---

[11] We assume in this argument that T permits"suitably" large transfers, e.g., suppose players get non-negative commodity bundles at every outcome which they strictly prefer to the zero bundle and that it is permitted in T to take away the entire bundle of any player i and give it to others, thus punishing i adequately whenever necessary.

[12] Transparency may be an especially relevant consideration when the alterations are to be made

our context of commodity income, one could think, for instance, of destruction — rather than transfers — of commodities as such an alteration.

(2) (**Mechanism Design)** There is burgeoning literature in which transfers are considered from a mechanism design perspective. The idea is to “design” an extensive form game (a “bargaining protocol”) in which players propose and accept/reject transfers amongst themselves, possibly over several rounds, at the end of which certain transfers are agreed upon (condional on the realized play of the game); and then finally the original one-shot game is played, with payoffs amended according to those transfers. This gives rise to an overall enhanced game whose (typically, “sequential”) NE are analysed. The protocol is meant to be *universal* in that it enables efficiency (welfare-maximization) to be achieved as *an* NE outcome for every one-shot game (chosen from an *a priori* given class of such games). It is worth emphasizing that efficiency is *enabled but not ensured* in this framework. Indeed many other inefficient NE also occur, and often there is a “folk theorem effect” with a plethora of NE outcomes. (See [15], and the references therein, for more details.)

Our approach is complementary to this literature. We take it for granted that players are eager to register transfers in a court, in a spirit of cooperation. (They need not be persuaded to participate in an elaborately designed mechanism, where cooperation could arise out of pure self-interested NE behavior.) Nevertheless, for any transfer that players unanimously agree upon, there remains a residual noncooperative game. In a forthcoming paper, we discuss a scheme for endogenizing transfers, based on appending markets for commodity trade to the original game; but the resultant enhanced game is still *one-shot*, and complementary to the mechanism-design approach.

3 Games of Money

In the rest of the paper, we shall illustrate our concepts through three examples:

- Centipede Game,
- Contest,
- Prisoners’ Dilemma.

---

voluntarily by individual players at *ex post* personal cost, albeit for *ex ante* personal gain.

In all these examples, incomes are denominated purely in terms of a single commodity which we call *money;* and players are risk-neutral. Thus the set K is a singleton; $u^i_\omega(x) = x$ for all i, ω, x; and the games are specified by

$$\varphi : \Sigma \longrightarrow \Delta(\Omega), \quad \mu : N \times \Omega \longrightarrow \mathbb{R}, \quad \pi^i(\sigma) = \sum_{\omega \in \Omega} \varphi_\sigma(\omega)\, \mu(i, \omega). \tag{6}$$

In addition: (i) each of the games has a *unique* NE σ^* prior to transfers; (ii) we invoke the *egalitarian* welfare function $w(\pi) = \sum_{i \in N} \pi^i$ (so that welfare is simply the *total money* in the hands of the players). On account of (i) and (ii), we may suppress σ^* and w, and write $\mathscr{E}(T)$, $G(T)$ etc. instead of $\mathscr{E}(\sigma^*, T)$, $G_w(\sigma^*, T)$ etc.

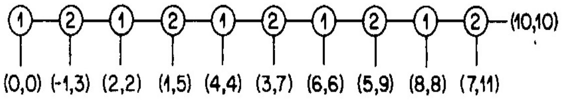
It further turns out, though this is not obvious, that the gains to transfers $G(T)$ are actually achieved via eligible NE in our examples, not just approximated.

For Contest, the post-transfer games that achieve $G(T)$ also have unique NE. This is no longer true for the Centipede Game or Prisoners' Dilemma. However, uniqueness is restored for "almost all" transfers in the Centipede Game once we restrict to *subgame perfect* NE. And it is "essentially" restored for Prisoners' Dilemma, for a minuscule price: given an arbitrarily small $\varepsilon > 0$, there exist transfers such that the post-transfers game has a unique NE that is both eligible and achieves gains to transfers of at least $G(T) - \varepsilon$. (See Remark 11 for details.)

The uniqueness of NE makes our examples all the more compelling, since we steer clear of the (generally unavoidable) arbitrariness of NE-selection in the course of determining either the status quo σ^* or the gains to transfers $G(T)$.

4 Centipede Game

The original "Centipede Game", due to Rosenthal (see [11]), is given by the extensive form



To recast it as a game of money, the numbers at the terminal nodes (outcomes) must be viewed as both the payoffs and the dollar rewards to players 1,2. The unique NE outcome entails player 1 choosing "down" at the left-most node of the game tree, resulting in rewards $(0, 0)$ to the players. Now consider the transfer τ^* in which player 1 transfers 1 dollar to player 2 at the right-most terminal node, changing rewards there from $(10,10)$ to $(9,11)$, while all other rewards remain untouched. Then $(9,11)$ becomes an eligible NE payoff in the post-transfers game, though $(0,0)$ also persists as another (albeit ineligible) NE payoff. Both the NE are, moreover, subgame perfect. (However note that, for any larger transfer $1 < \tau^* < 2$, the right-most node becomes the *unique* subgame-perfect NE outcome, which is also eligible and has maximum welfare 20, thus achieving the gains to transfers in the Centipede Game.

Proposition 4 *Let* T *be* any *bounded set of transfers at the rightmost node that contains* $T^* = \{\tau : 1 \leq \tau^* \leq 2\}$. *Then* $G(T) = 20$ *and is achieved by any* $\tau \in T^*$.

Proof. This is obvious in light of our earlier discussion, since $10 + 10 = 20$ is the maximum total money among all the terminal nodes; and since the welfare at any NE (eligible or not) obtained via any transfer, being an expectation of the total money across terminal nodes, cannot exceed 20. ■

5 Contest

There is a vast economic literature on Contest, starting with [14] and [8] (for an extensive survey, see [6]). Here we consider a simple example of an asymmetric binary contest. Suppose Rowena (the row player) and Colin (the column player) can each exert either low or high effort, denoted L and H respectively, which incur disutility d and produce output t, as follows:

Rowena	d	t
L	0	80
H	30	120

Colin	d	t
L	0	0
H	30	82

(Note: Rowena is uniformly stronger than Colin.). An outcome is the pair of outputs produced by the two, and is observable and owned by Emily who is their employer. However Emily has announced that she will keep only half the total output to herself and hand out the other half as bonus to the one who produces more (over and above a fixed wage, which w.l.o.g we take to be 0). This induces the following game between Rowena and Colin

	L	H
L	$(40,0)$	$(0,51)$
H	$(30,0)$	$(71,-30)$

Claim 5 *Let T be any bounded set of simple transfers at (H,H) that contains transfers of* $30+\Delta$ *from Rowena to Colin at* (H,H), *for* $0<\Delta<5.938$. *Any such* Δ *achieves the gains to transfers* $G(T)$.

Proof. It is easy to check that the pre-transfer game has a unique NE, where Rowena plays the (mixed) strategy $(30/81, 51/81)$ and Colin plays $(71/81, 10/81)$, with expected payoffs of $2480/81 = 35.062$ and 0 respectively. However a transfer of $30+\Delta$ units (out of her bonus of 101) from Rowena to Colin at the entry (H,H), for $0<\Delta<41$, will ensure that (H,H) is the unique NE outcome with payoffs $41-\Delta$ to Rowena, and Δ to Colin. For $0<\Delta<5.938$, both will be better off compared to the original NE, though Rowena only moderately so and Colin by a considerable amount. The NE is clearly eligible. The Claim now follows. ■

Note that the gift of money by the strong Rowena to the weak Colin is of the benefit to both contestants.

A "converse" claim may be formulated when T is the set of all simple transfers at (L,H). Indeed, a transfer of $40+\Delta$ (out of his bonus of 81 from the weak Colin to the strong Rowena), at the entry (L,H), will ensure that (L,H) is the unique eligible NE outcome with both better off.

At first glance, transfers from Colin to Rowena at (L,H) appear to be more efficient because (L,H) is the "social optimum" in the game, in the sense that the sum of the two contestants' payoffs is maximum there. However, if we include Emily in the picture, the scenario changes completely because the (expected total) output at (H,H) is 202, which is considerably better (even after compensating contestants for their effort) than the output of 162 at (L,H) or of 127.5 at the NE of the original pre-transfers game. There is every incentive for Emily to announce a higher bonus at (H,H) to lure both players to choose effort H. (In section 7, we introduce a *game designer,* such as Emily, injecting money at judiciously chosen outcomes of the game — thus going beyond transfers — in order to influence players' strategic behavior.)

6 Prisoners' Dilemma

We consider a class of symmetric Prisoners' Dilemma games (of money), where each player has two pure strategies: 0 (defect) and 1 (cooperate). We write $[p,q]$ for the mixed strategy pair where the two players choose to cooperate with probabilities p and q, respectively. The games we consider have payoffs given by the following bi-matrices[13], whose entries may be viewed as dollars or utiles:

$$\Gamma^{\alpha,\beta} = \begin{array}{|c|c|c|} \hline & 0 & 1 \\ \hline 0 & (\beta,\beta) & (1,0) \\ \hline 1 & (0,1) & (\alpha,\alpha) \\ \hline \end{array} \qquad 1 > \alpha > \beta > 0 \tag{7}$$

The cells of the matrix correspond to pure strategy pairs; thus $[0,1]$ is the top right cell etc. It is easy to see that $\Gamma^{\alpha,\beta}$ has a unique NE $\sigma^* = [0,0]$ with welfare 2β. The "dilemma" is, of course, that the non-equilibrium pair $[1,1]$ yields higher payoff $\alpha > \beta$ to each player. For simplicity we shall restrict attention to simple transfers in the cell $[0,1]$, giving rise to the games

$$\Gamma^{\alpha,\beta}_{\tau} = \begin{array}{|c|c|} \hline (\beta,\beta) & (1-\tau,\tau) \\ \hline (0,1) & (\alpha,\alpha) \\ \hline \end{array} \qquad 1 \geq \tau \geq 0. \tag{8}$$

This set T of transfers shall be held fixed throughout, so we suppress it in our notation and denote the gains to transfers in $\Gamma^{\alpha,\beta}$ by

$$G(\alpha,\beta) = \max_{\tau,\sigma} w(\pi_\tau(\sigma)) - w(\pi(\sigma^*)) = \max_{\tau,\sigma} \left(\pi^1_\tau(\sigma) + \pi^2_\tau(\sigma) - 2\beta\right),$$

where (recall) $\pi^i_\tau(\sigma)$ is the payoff to player i at σ in the game $\Gamma^{\alpha,\beta}_\tau$.

Note that although $\Gamma^{\alpha,\beta}$ is symmetric, $\Gamma^{\alpha,\beta}_\tau$ is not, and thus the eligible NE that achieve the gains to transfers $G(\alpha,\beta)$ will in general be *symmetry-breaking.* To the reader who finds this troubling, we note that one may we view $\Gamma^{\alpha,\beta}_\tau$ as *half* the story. Consider the transposed game with the simple transfer τ from Column to Row at the cell $[1,0]$. Let nature choose between the two games with equal probability, after which both players are fully informed of the transformed game they are in. This leads to an extensive form whose subgame perfect equilibrium achieves the same gains to transfers as $\Gamma^{\alpha,\beta}_\tau$, with equal payoffs to the players, and symmetry restored. With this explicit understanding, we proceed with our asymmetric analysis.

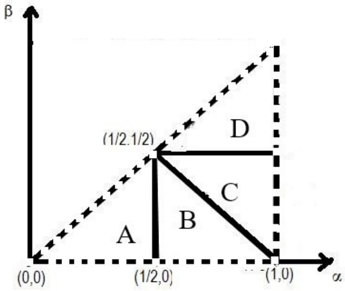
---

[13] The original example reported by Tucker ([13]) in the early 1950's, was as follows:

$$\begin{array}{|c|c|} \hline (-1,-1) & (1,-2) \\ \hline (-2,1) & (0,0) \\ \hline \end{array}$$

By a change of scale for measuring money (*i.e,* a common affine transformation applied to all the entries of the matrix, namely adding the constant 2 and scaling by $1/3$) the example corresponds to our case with $\alpha = 2/3$ and $\beta = 1/3$. Indeed any Prisoners' Dilemma may be put into our format $\Gamma^{\alpha,\beta}$ by such a change of scale.

Consider the triangular region $\mathscr{R} = \{(\alpha, \beta) : 1 > \alpha > \beta > 0\}$. It will be convenient to divide $\mathscr{R}$ into four (non-disjoint) smaller triangles A, B, C, D shown in the Figure below.



Note that the perimeter of $\mathscr{R}$ is understood to be missing. Also note that the intersection of any two consecutive triangles in the list A, B, C, D is the straight line (minus its endpoints) that is the common boundary between them. We shall denote the interiors of these regions by A', B', C', D'.

The following theorem identifies all pairs (σ, τ) such that

$$G(\alpha, \beta) = \pi_\tau^1(\sigma) + \pi_\tau^2(\sigma) - 2\beta$$

showing that in particular that $G(\alpha, \beta)$ is in fact achieved.

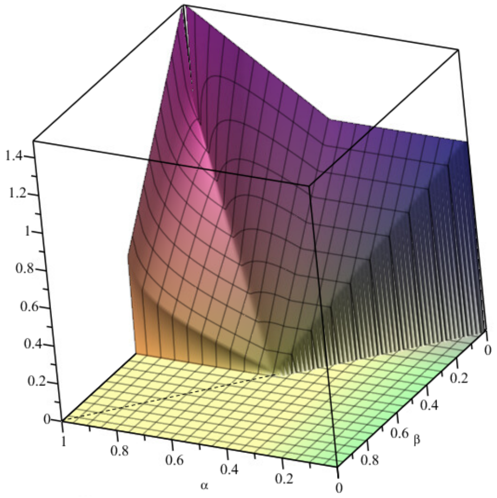
Theorem 6 *For $(\alpha, \beta) \in \mathscr{R}$ let us denote*

$$\alpha' = 1 - \alpha, \beta' = 1 - \beta \text{ and } \zeta = \sqrt{\alpha\beta\alpha'\beta'}.$$

The table below shows all pairs $(\tau, \sigma) \in \mathscr{E}\left(\Gamma^{\alpha,\beta}\right)$ at which $G(\alpha, \beta)$ is achieved

Region	τ	σ	$G(\alpha, \beta)$
A'	$\beta \le \tau \le \alpha'$	$[0, 1]$	$1 - 2\beta$
$A \cap B$	$\alpha' = 1/2$	$[p, 1] : 0 \le p \le \frac{1-2\beta}{3-2\beta}$	1
B'	α'	$\left[\frac{\tau-\beta}{\tau-\beta+\alpha'}, 1\right]$	$\frac{1+\alpha-2\alpha^2-2\alpha\beta}{2-2\alpha-\beta} - 2\beta$
C	$\alpha' - \beta + 2(\alpha\beta + \zeta)$	$\left[\frac{\tau-\beta}{\tau-\beta+\alpha'}, \frac{\beta}{\tau-\alpha'+\beta}\right]$	$\frac{1}{2} + \frac{\alpha\beta-\zeta}{\alpha+\beta-1} - 2\beta$
D'	0	$[0, 0]$	0

Moreover for $(\alpha, \beta, \tau, \sigma)$ in the table, if $\tau \ne \alpha'$ then σ is the unique *NE of* $\Gamma_\tau^{\alpha,\beta}$.



The figure shows the graph[14] of the function G as a surface[15] above $\mathscr{R}$.

First note that the surface is continuous on $\mathscr{R}$ except for $C \cap D$, which is a line of discontinuity, since G drops abruptly to 0 as we leave C and enter D'. Next

$$\lim_{\beta \to 0} G(1-k\beta,\beta) = \frac{3k-2}{2k-1} = \frac{3}{2} - \frac{1}{2(2k-1)}$$

i.e., as we approach $(1,0)$ along different lines $\alpha + k\beta = 1$ from within B', the limit of G rises from 1 to 1.5 as k rises from 1 to ∞, showing that G cannot be continuously extended to $(1,0)$. The supremum of G on $\mathscr{R}$ is 1.5 and is not achieved on $\mathscr{R}$. Finally, by Lemma 8 below, $G(\alpha,\beta) > 1-2\beta$ in $B' \cup C'$. (Intuitively speaking this is so because, in the completely mixed eligible NE that achieves $G(\alpha,\beta)$, the small welfare $2\beta < 1$ at the cell $(0,0)$ is more than compensated for by the probability of visiting the cell $(1,1)$ with the large welfare $2\alpha > 1$.).

In light of this picture, consider points (α,β) close to $(1,1/2)$ in D'. The gains to transfers are 0, so that players' payoffs are approximately 0.5 each after making optimal transfers. Now *reduce* β to β', so that (α,β') is in B', near $(1,0)$ and on the line $x + ky = 1$ for large enough k. Then the gains to transfers will be near 1.5, and (after symmetrization, as discussed above) each player is better off with payoff just under 0.75.

---

[14] We are grateful to Yusra Naqvi for generating this beautiful picture.

[15] The line segment, connecting $(0,0)$ to $(1,1)$ is not in $\mathscr{R}$. Its second half, shown as a dashed line in the figure, corresponds to the missing hypotenuse of D; and the dark cliff above its first half, which is meant to be vertical, is not part of the surface.

6.1 Proof of Theorem 6

For the proof we will compute all eligible NE $\sigma = [p,q]$ of $\Gamma_\tau^{\alpha,\beta}$ and the payoffs c,r to the column and row player. There are three kinds of NE.

1. *Completely mixed*, where each player uses a mixed strategy
2. *Pure*, where each player chooses a pure strategy, i.e., either 0 or 1.
3. *Half-mixed*, where one player chooses a pure strategy and the other, mixed.

The next two lemmas deal with completely mixed NE. For convenience we work with the slightly larger class of *equalizing* NE, where each player is indifferent between choosing 0 and 1.

Lemma 7 *The game $\Gamma_\tau^{\alpha,\beta}$ has an eligible equalizing NE if and only if*

$$(\alpha,\beta) \in A \cup B \cup C \text{ and } \beta' \geq \tau \geq \max\left\{\alpha',\beta\right\}.$$

In this case there is a unique *such equilibrium, given by*

p	q	c	r
$\frac{\tau-\beta}{\tau-\theta}$	$\frac{\beta}{\tau+\theta}$	$\frac{\tau-\pi}{\tau-\theta}$	$\frac{\pi}{\tau+\theta}$

where $\pi = \alpha\beta$ and $\theta = \alpha+\beta-1$.

Proof. Suppose $[p,q]$ is an equalizing NE of the game

$$\Gamma_\tau^{\alpha,\beta} = \begin{array}{|c|c|} \hline (\beta,\beta) & (1-\tau,\tau) \\ \hline (0,1) & (\alpha,\alpha) \\ \hline \end{array}, \quad 1 \geq \tau \geq 0,$$

with payoffs c and r to the row and column players. Then by definition of "equalizing" we get

$$\begin{aligned} c &= (1-p)\beta + p = (1-p)\tau + p\alpha \\ r &= (1-q)\beta + q(1-\tau) = q\alpha \end{aligned}$$

Solving these equations yields

$$p = \frac{\tau-\beta}{\tau-\theta},\ c = \frac{\tau-\pi}{\tau-\theta},\ q = \frac{\beta}{\tau+\theta},\ r = \frac{\pi}{\tau+\theta} \tag{9}$$

as in the table.

The numbers p,q,r,c are required to satisfy the following inequalities

$$1 > p \geq 0, \quad 1 \geq q > 0, \quad c \geq \beta, \quad r \geq \beta,$$

where the inequalities on r,c and the strict inequalities on p,q comprise the *eligibility* of σ. Now (9) gives

$$1 - \frac{1}{1-p} = \frac{\tau - \beta}{\alpha'}, \quad 1 - \frac{1}{q} = \frac{\tau - \alpha'}{\beta}$$

and since $\alpha', \beta > 0$ on $\mathscr{R}$, the conditions on p,q imply

$$\tau \geq \beta \text{ and } \tau \geq \alpha'. \tag{10}$$

Another simple calculation using (9) gives

$$c - \beta = \frac{\beta'}{\tau - \theta}\left(\tau - \beta\right), \quad r - \beta = \frac{\beta}{\tau + \theta}\left(\beta' - \tau\right). \tag{11}$$

Combining (9) and (10) we get $\tau + \theta > 0$ and $\tau - \theta > 0$. Therefore (11) gives

$$\tau \geq \beta \text{ and } \tau \leq \beta'. \tag{12}$$

and by (10) and (12) we get

$$\beta' \geq \tau \geq \max\left\{\alpha', \beta\right\}.$$

as needed. Since $\alpha' < \beta'$ in the region $\mathscr{R}$, such τ exist if and only if

$$\beta' \geq \beta \iff \beta \leq \frac{1}{2} \iff (\alpha, \beta) \in A \cup B \cup C.$$

■

In the view of the above lemma, for $(\alpha, \beta) \in A \cup B \cup C$, we define

$$\begin{aligned} I(\alpha,\beta) &= \left\{\tau : \beta' \geq \tau \geq \max\left\{\alpha', \beta\right\}\right\} \\ W_{eq}(\alpha,\beta,\tau) &= r + c = \frac{\tau - \pi}{\tau - \theta} + \frac{\pi}{\tau + \theta} \\ MW_{eq}(\alpha,\beta) &= \max\left\{W_{eq}(\alpha,\beta,\tau) : \tau \in I(\alpha,\beta)\right\} \end{aligned}$$

Lemma 8 *For (α,β) in A' we have $MW_{eq}(\alpha,\beta) < 1$. In regions B and C the maximum MW_{eq} is attained as follows*

	MW_{eq}	τ
B	$\frac{1+\alpha-2\alpha^2-2\alpha\beta}{2-2\alpha-\beta}$	α'
C	$\frac{1}{2}+\frac{\sqrt{\alpha\beta}}{\sqrt{\alpha\beta}+\sqrt{\alpha'\beta'}}$	$\left(\sqrt{\alpha\beta}+\sqrt{\alpha'\beta'}\right)^2$

On B' and on $C \setminus B$, we have $MW_{eq}(\alpha,\beta) > 1$. On $A \cap B$ and $B \cap C$, $MW_{eq}(\alpha,\beta) = 1$ and is achieved by the limits of the equalizing, eligible NE from the interior (of either side).

Proof. Let $\theta = \alpha+\beta-1, \pi = \alpha\beta$ as before and put $\pi' = \alpha'\beta' = (1-\alpha)(1-\beta)$, then we have

$$\pi - \pi' = \theta$$

and it follows by a direct calculation that

$$W_{eq}(\alpha,\beta,\tau) - 1 = \frac{\pi}{\tau+\theta} + \frac{\tau-\pi}{\tau-\theta} - 1 = \frac{\pi}{\tau+\theta} - \frac{\pi'}{\tau-\theta} \tag{13}$$

$$= \frac{\theta\left[\tau - (\pi+\pi')\right]}{\tau^2-\theta^2}. \tag{14}$$

For $\tau \in I(\alpha,\beta)$ we have $\tau \geq \alpha'$ and also

$$\tau^2 > \theta^2 \tag{15}$$

For (α,β) in A' we further have $\theta < 0$ and $\alpha < 1/2$ which implies $\alpha' > \alpha$ and hence

$$\pi + \pi' = \alpha\beta + \alpha'(1-\beta) \leq \alpha' - \left(\alpha' - \alpha\right)\beta < \alpha'. \tag{16}$$

Thus for (α,β) in A' and $\tau \in I(\alpha,\beta)$ the expression (14) is negative. It follows that $MW_{eq}(\alpha,\beta) < 1$ for (α,β) in A'.

For regions B and C we examine more closely the extrema of $W_{eq}(\alpha,\beta,\tau)$ as τ varies in $I(\alpha,\beta)$. Differentiating (13) we get

$$\frac{d}{d\tau}W_{eq}(\alpha,\beta,\tau) = \frac{\pi'}{(\tau-\theta)^2} - \frac{\pi}{(\tau+\theta)^2} \tag{17}$$

$$= \frac{-\theta}{\left(\tau^2-\theta^2\right)^2}\left(\tau^2 - 2(2\pi-\theta)\tau + \theta^2\right)$$

$$= \frac{-\theta}{\left(\tau^2-\theta^2\right)^2}(\tau-\tau_+)(\tau-\tau_-) \tag{18}$$

where $\tau_{\pm}$ are given by the quadratic formula

$$\tau_{\pm} = (2\pi-\theta) \pm \sqrt{(2\pi-\theta)^2 - \theta^2} = 2\pi - \theta \pm 2\sqrt{\pi}\sqrt{\pi-\theta} \tag{19}$$

$$= \left(\sqrt{\pi} \pm \sqrt{\pi-\theta}\right)^2 = \left(\sqrt{\pi} \pm \sqrt{\pi'}\right)^2 \tag{20}$$

Thus the maximum of $W_{eq}(\alpha,\beta,\tau)$ can only occur at the endpoints of $I(\alpha,\beta)$, or at $\tau_{\pm}$ if these happen to lie in the interior.

We note that $\tau_+ \geq \tau_- \geq 0$ and moreover

$$\tau_+\tau_- = \left(\sqrt{\pi}+\sqrt{\pi'}\right)^2 \left(\sqrt{\pi}-\sqrt{\pi'}\right)^2 = \left(\pi - \pi'\right)^2 = \theta^2.$$

It follows that $\tau_- \leq |\theta|$ and by (15) we conclude that τ_- does *not* belong to the interior of $I(\alpha,\beta)$.

For (α,β) in B', we have $\alpha' > \beta$ and so $I(\alpha,\beta)$ is the closed interval $[\alpha',\beta]$. Also $\theta < 0$ thus in formula (18) the derivative changes from negative to positive near τ_+, and so τ_+ is a local *minimum*. We claim that the maximum value of $W_{eq}(\tau)$ is obtained at the left endpoint α'. In fact we will prove the sharper result

$$W_{eq}\left(\alpha'\right) > 1 > W_{eq}(\beta).$$

Since $\theta < 0$ by formula (14) it suffices to show that

$$\alpha' < \alpha\beta + \alpha'\beta' < \beta,$$

which is proved as in (16). Thus we get

$$MW_{eq}(\alpha,\beta) = W_{eq}(\alpha') = \frac{\pi}{\theta+\alpha'} + \frac{\pi-\alpha'}{\theta-\alpha'}$$
$$= \frac{\alpha\beta}{\beta} - \frac{\alpha\beta - 1 + \alpha}{2-2\alpha-\beta} = \frac{1+\alpha-2\alpha^2-2\alpha\beta}{2-2\alpha-\beta}.$$

In region C' we have $\theta > 0$, thus the above argument shows τ_+ is a maximum and we get

$$MW_{eq}(\alpha,\beta) = W_{eq}(\tau_+) = \frac{\pi}{\theta+\tau_+} + \frac{\pi-\tau_+}{\theta-\tau_+}. \tag{21}$$

To simplify (21) we set

$$a = \sqrt{\alpha\beta}, \quad b = \sqrt{\alpha'\beta'}.$$

Note that in C'

$$a > b \tag{22}$$

Now we can write $\pi = \alpha\beta = a^2$ and we get

$$\begin{aligned}\theta &= \alpha\beta - \alpha'\beta' = a^2 - b^2 \\ \tau_+ &= (a+b)^2 = a^2 + 2ab + b^2\end{aligned}$$

This implies

$$\begin{aligned}\tau_+ + \theta &= 2a^2 + 2ab = 2a(a+b) \\ \tau_+ - \theta &= 2b^2 + 2ab = 2b(a+b) \\ \tau_+ - \pi &= b^2 + 2ab = b(2a+b)\end{aligned}$$

Thus the maximal payoff (21) in region C' becomes

$$\begin{aligned}MW_{eq}(\alpha,\beta) &= \frac{a^2}{2a(a+b)} + \frac{b(2a+b)}{2b(a+b)} = \frac{1}{2(a+b)}[a+2a+b] \\ &= \frac{1}{2} + \frac{a}{a+b} = \frac{1}{2} + \frac{\sqrt{\alpha\beta}}{\sqrt{\alpha\beta}+\sqrt{\alpha'\beta'}} \end{aligned} \tag{23}$$

as claimed. Moreover by (22) we get $MW_{eq}(\alpha,\beta) > 1$.

Finally we consider the boundary lines. On $B \cap C$ we have $\theta = 0$ hence we get

$$W_{eq}(\tau) = \frac{\pi}{\tau} + \frac{\tau - \pi}{\tau} = 1$$

for all $\tau \in I(\alpha,\beta)$. On $A \cap B$ we have $I(\alpha,\beta) = \{\tau : \beta' \geq \tau \geq 1/2\}$ and

$$\alpha = \alpha' = 1/2, \quad \pi + \pi' = 1/2, \quad \theta < 0$$

By (14), we get

$$W(\tau) - 1 = \frac{\theta[\tau - 1/2]}{\tau^2 - \theta^2}$$

Thus $W(1/2) = 1$ and $W(\tau) < 1$ for $\beta' \geq \tau > 1/2$ and hence $MW_{eq}(\alpha,\beta) = 1$. ■

Lemma 9 *The eligible pure NE of* $\Gamma_\tau^{\alpha,\beta}$ *are*

1. $[0,0]$ *for* $\tau = 0$ *and any* (α,β)*, with* $r = c = \beta$ *and* $r + c = 2\beta$

2. $[0,1]$ *for* $\beta \leq \tau \leq \alpha'$ *and* $(\alpha,\beta) \in A \cup B$*, with* $r = 1-\tau, c = \tau$ *and* $r + c = 1$.

Proof. For $\tau = 0$ and any $\Gamma^{\alpha,\beta}$, clearly 0 is a strictly dominant strategy for each player, which proves (1). If $\tau > 0$, then transparency requires that $[0,1]$ be the NE, with payoffs $1-\tau$ and τ to the row and column players respectively, and therefore welfare 1. Now for 1 to be a best reply of the column player (to 0), we must have $\tau \geq \beta$; and for 0 to be a best reply of the row player (to 1) we must have $1-\tau \geq \alpha$. The last two inequalities imply $\beta \leq \tau \leq 1-\alpha$; and in this case each player's payoff is also at least β. Such τ exist if, and only if, $\alpha + \beta \leq 1$, i.e., $(\alpha,\beta) \in A \cup B$. This proves (2). ■

Finally we examine the half-mixed NE. Transparency requires that the cell $[0,1]$ occur with positive probability, thus the half-mixed NE can be one of two types: $[0,q]$ or $[p,1]$, with $0 < p, q < 1$.

Lemma 10 1. *Type* $[0,q]$ *exists if and only if* $\tau = \beta$ *and* $(\alpha,\beta) \in C' \cup D$.

2. *Type* $[p,1]$ *exists if and only if* $\tau = \alpha'$ *and* $(\alpha,\beta) \in A \cup B'$*; and in this case there is a continuum of such NE given by*

$$\mathscr{C}(\alpha,\beta) = \{[p,1] : 0 \leq p \leq p(\alpha,\beta)\}; \quad p(\alpha,\beta) := \frac{1-\alpha-\beta}{1-\alpha-\beta+1}$$

Proof. For type $[0,q]$, indifference of the column player (between his two pure strategies) implies $\tau = \beta$; and the choice of the first row implies $(1-q)\beta + q(1-\beta) \geq q\alpha$, i.e., $q(\alpha+\beta+\beta-1) \leq \beta$, which has solutions $0 < q < 1$ if and only if $\alpha + \beta > 1$, i.e., $(\alpha,\beta) \in C' \cup D$.

For type $[p,1]$. indifference of the row player implies $\tau = \alpha'$, and the choice of the second column implies

$$(1-p)(1-\alpha) + p\alpha \geq (1-p)\beta + p \tag{24}$$

Since $(1-\alpha) \leq \beta$ in $C \cup D$ and $\alpha < 1$ always, we conclude that (24) cannot hold in $C \cup D$, hence such NE do not exist in $C \cup D$.

On the other hand, $1-\alpha-\beta > 0$ on $A \cup B'$, hence $0 < p(\alpha,\beta) < 1$. Moreover (24) is equivalent to $p \leq p(\alpha,\beta)$. These two inequalities imply that $\mathscr{C}(\alpha,\beta)$ is the set of eligible half-mixed NE as claimed. ■

Proof. (Completion). Denote by $W_m(q)$ the welfare at the half-mixed NE $[0,q]$. Then $W_m(q) = (1-q)(2\beta) + q$. Now if $(\alpha,\beta) \in C'$ then $2\beta < 1$, so $W_m(q) < 1$; if $(\alpha,\beta) \in D'$ then $2\beta > 1$, so $W_m(q) < 2\beta$; and if $(\alpha,\beta) \in C \cap D$, then $2\beta = 1$, so $W_m(q) = 1$. Then, by Lemma 8 and Lemma 9, W_m is surpassed in all three scenarios by the welfare of an eligible, equalizing NE or an eligible, pure NE.

Next, the welfare at any $[p, 1] \in \mathscr{C}(\alpha, \beta)$ is

$$W_m(p) = (1 - p) + p(2\alpha) \tag{25}$$

In A' we have $2\alpha < 1$, hence (25) implies that $W_m < 1$, which is surpassed (see Lemma 9) by the welfare of the pure NE $[0, 1]$. In B', clearly $[p(\alpha, \beta), 1] \in \mathscr{C}(\alpha, \beta)$ is the eligible, equalizing NE of Lemma 7; moreover by (25) and the fact that $2\alpha < 1$, all the other NE in $\mathscr{C}(\alpha, \beta)$ have strictly lower welfare than $[p(\alpha, \beta), 1]$. Finally, on $A \cap B$, where we have $2\alpha = 1$ and $p^* = (1 - 2\beta) / (3 - 2\beta)$, it is evident that all the NE have welfare 1. ■

Theorem 6 is now immediate from Lemmas 7, 8, 9 and 10, after substituting $1 - \alpha, 1 - \beta$ for α', β' and doing the algebraic simplifications so that all expressions are in terms of α, β, ζ.

Remark 11 ***(Uniqueness and Eligibility of NE in the Post-Transfers Games $\Gamma_\tau^{\alpha,\beta}$).*** *Theorem 6 displays the optimal pairs (τ, σ) for any given game $\Gamma^{\alpha,\beta}$. One may ask the question: once the players contract upon an optimal τ, and generate the post-transfers game $\Gamma_\tau^{\alpha,\beta}$, how likely is it that they will play the concommitant, "correct" σ? It is already a strong behavioral assumption that players will get to some NE of $\Gamma_\tau^{\alpha,\beta}$. The correct σ calls upon them to do more: they must get to an NE that is not just eligible, but the best eligible . For instance (see Lemma 10) the optimal τ is $\alpha' = 1 - \alpha$ for $(\alpha, \beta) \in B'$ and gives rise to a continuum $\mathscr{C}(\alpha, \beta)$ of eligible NE in $\Gamma_{1-\alpha}^{\alpha,\beta}$, only one of which maximizes welfare and is"best" or "correct". Fortunately we may bypass this vexing problem of NE-selection for a miniscual price. A careful reading of our proof shows: for any $(\alpha, \beta) \in \mathscr{R}$ and any $\varepsilon > 0$, there exists τ such that $\Gamma_\tau^{\alpha,\beta}$ has a unique NE, which is both eligible and achieves gains to transfers above $G(\alpha, \beta) - \varepsilon$.*

7 Variation on the Theme: Beyond Transfers to Alterations

Consider a *principal* (a "government", or "game designer") with the ability to credibly offer *rewards* of outcome-contingent commodities to players (e,g. via legally binding escrow accounts). The principal might also propose what players should occasionally give up, in exchange for the rewards. However, while players are ever ready to *get* rewards, it is quite another matter to persuade them to *give up* anything. They will need to be convinced that there are significant gains for them in the process. (See the "participation constraints" below). Moreover, no player will be gullible enough to take on faith other players' promises to give up commodities: this will have to be seen to be believed! (See the "semi-transparency" condition below.)

For ease of presentation, we make two simplifying assumptions. First, as usual, there is only one commodity which we call "money", i.e., K is a singleton set. Second, the objective of the principal is to maximize his expected profit.[16]

Let τ be a transfer as in section 2.2, except that we shall now call it an "*alteration*", in view of the fact that the budget balance condition will no longer be imposed.

Definition 12 *The pair* (τ,σ), *consisting of an alteration* τ *and a strategy choice* σ, *is called a* Nash Equilibrium with Transparent Alterations (NETA) *in the game* (φ,μ) *if*

(a) for every $\omega\in\Omega$

$$\tau(\omega)\ngeqq 0 \Longrightarrow \varphi(\omega,\sigma)>0$$

(b) σ *is a Nash Equilibrium of* $(\varphi,\mu+\tau)$;

The symbol $\ngeqq$ above means that at least one of the components of the vector $\tau(\omega)$ is strictly negative. Thus the "semi-transparency"condition (a), only requires transparency when any player gives up money.

The profit maximization problem of the principal may now be formulated as follows (where T is a given set of feasible alterations):.

$$\text{Max} \sum_{\omega\in\Omega}\left[\varphi(\omega,\sigma)\sum_{i\in N}-\tau^{i}(\omega)\right]$$

subject to:

$$\tau\in T \text{ and } (\tau,\sigma) \text{ is NETA in } (\varphi,\mu)$$
$$\tau\neq 0 \implies u_i(\sigma,\varphi,\mu+\tau)\geqslant\kappa_i \text{ for all } i\in N$$

---

[16] The reader can easily check that the model is well defined for multiple commodities, and an objective that is an arbitrary function of players' outcome-contingent consumptions as well as the principal's own outflows and inflows of commodities across outcomes. (Moreover, the variantion where each player is required to repay his loan at some stipulated *positive* interest rate, can be incorporated in our model in the obvious manner and is also left to the reader.)

The last display constitutes the *participation constraints,* where the are κ_i arbitrary constants. Thus, as before, κ_i could be the payoff of i at some status quo NE of the pre-transfers game; or, going further, κ_i could be the maximum expected money that i can get across all NE of the pre-transfers game, creating more incentive for player i will accept the alteration τ. (Note that by choosing κ_i sufficiently negative, the participation constraints may also be rendered irrelevant.).

The game designer is here like a *lender* of money who (via τ) gives each player i money $\tau^i(\omega)$ (at outcomes ω where $\tau^i(\omega) > 0$) in exchange for getting money $-\tau^i(\omega')$ from i (at outcomes ω where $\tau^i(\omega) < 0$). Since $0 \in T$ the maximum in the above optimization problem is non-negative, i.e., the game-designer does not make a loss.

7.1 Promises in Extensive Form Games: an Opportunity for Arbitrage

We present a special example of NETA, which may be of some interest in its own right. Our domain consists of finite extensive form games with perfect information and without chance moves; and, throughout this subsection, by "NE" we shall mean a "*subgame-perfect* NE".

In terms of our general model, the specifications are as follows: Σ^i is the set of *pure* strategies of player $i \in N$; Ω is the set $\mathbb{T}$ of terminal nodes of the game tree; φ maps $\sigma \in \Sigma$ to the terminal node reached under σ; K is a singleton set (money); and $u_t^i(x) = x$ for all $i \in N$ and $t \in \mathbb{T}$. This leaves $\mu : N \times \mathbb{T} \longrightarrow \mathbb{R}$ to finish defining the game. We shall vary μ, keeping the rest of the data fixed; thus, without confusion, we may refer to the "game μ" from now on. We constrain μ to be in the "generic" set[17]

$$\Pi = \{\mu : \text{all the values } \mu(i,t) \text{ are distinct across } N \times \mathbb{T}\}.$$

For $\mu \in \Pi$, there is a unique NE which we denote σ_μ. Also we denote by t_μ the terminal node reached under σ_μ; and by $\mu(t_\mu) = \left(\mu(i,t_\mu)\right)_{i\in N} \in \mathbb{R}^N$ players' payoffs at the NE σ_μ.

Next define

$$\Pi(\mu) = \left\{\widetilde{\mu} \in \Pi : \widetilde{\mu} \geq \mu, \widetilde{\mu}\left(t_{\widetilde{\mu}}\right) = \mu\left(t_{\widetilde{\mu}}\right), \widetilde{\mu}\left(t_{\widetilde{\mu}}\right) \geq \mu\left(t_\mu\right)\right\}$$

(Note that $\mu \in \Pi(\mu)$, so $\Pi(\mu)$ is nonempty.) Given the original game μ, the set of rewards (transfers) available to the principal is $T(\mu) = \{\widetilde{\mu} - \mu : \widetilde{\mu} \in \Pi(\mu)\}$.

---

[17] In fact, the complement of Π in $\mathbb{R}^{N\times T}$ is a finite union of lower-dimensional planes.

The reward $\widetilde{\mu}-\mu \geq 0$ transforms the game μ to $\widetilde{\mu}$. The NE $\sigma_{\widetilde{\mu}}$ is clearly NETA since players not giving up any money; moreover, our definition of $\Pi(\mu)$ guarantees that the *ensuing play* $t_{\widetilde{\mu}}$ *of* $\sigma_{\widetilde{\mu}}$ *does not call upon the principal to redeem his promises and actually hand out any rewards at all.* Thus the principal can manuever the play to any terminal node in $\mathbb{T}(\mu)=\left\{t_{\widetilde{\mu}} \in \mathbb{T} : \widetilde{\mu} \in \Pi(\mu)\right\}$ *at no cost whatsoever to himself* and with no loss to any player relative to his status quo payoff at t_μ. The principal's promises $\widetilde{\mu}$ lurk in the background, and players think about them while calculating their moves, but the play $t_{\widetilde{\mu}}$ of the game that is realized in equilibrium steers clear of his promises.

We may view the situation in the form of a maximization problem, as in the preceding section. Given the game μ, let $\kappa_i=\mu\left(i,t_\mu\right)$. And let the principal choose rewards $\widetilde{\mu}-\mu$ in $T(\mu)$ so as to maximize the total money $\sum_{i\in N}\mu\left(i,t_{\widetilde{\mu}}\right)$ in the hands of the players (equivalently, to achieve the gains to transfers under the egalitarian welfare[18]). This maximum $M(\mu)$ is a well-defined number for every extensive form game of money μ (with perfect information). Were we to suppose that, by way of "commission", the principal can collect a fixed fraction K of every player's monetary gain (in the transition from the old NE σ_μ to the new NE $\sigma_{\widetilde{\mu}}$), then $K\left(M(\mu)-\sum_{i\in N}\kappa_i\right)$ is the money he can take away from the game μ, after engineering $(1-K)\left(M(\mu)-\sum_{i\in N}\kappa_i\right)$ welfare gains for the players. In effect, the game is like a "money pump", giving him the opportunity of arbitrage!

By way of a numerical example of NETA, consider the following extensive form. Player 1 moves first, followed by 2, followed again by 1. All moves are L,R. Drawing L to the left and R to the right, we have terminal nodes going from left to right: LLL, LLR, LRL etc. with payoffs $(3,3),(1,1),(1.1,2),(1.5,4),(0,1.5),(1.3,1.1),(0.5,1.2),(2,2)$. The equilibrium play is RRR leading to payoffs $(2,2)$. Now let the designer boost $(1.1,2)$ to $(2.9,2)$ and $(0.5,1.2)$ to $(3.1,1.2)$. The Nash equilibrium play switches to LLL with payoffs $(3,3)$ and clearly constiutes a NETA. The principal does not have to keep his promises. He engenders the increase in welfare, and makes his commision, at no cost.

References

[1] Cournot A.A.(1838). Recherches sur les Principles de la Theorie des Richesses, Translated by Bacon (1897), *MacMillan & Co. , Ltd. (London)*

---

[18] i.e., the sum of players' utilities.

[2] P.Dubey & S.Sahi (2016). Optimal Transfers in Noncooperative Games, *Working Paper, Stony Brook Department of Economics*

[3] D. Fudenberg. & D.K.Levine (1993). Self-Confirming Equilibrium, *Econometrica, Vol. 61, No. 3, pp 523-545.*

[4] J.Harsanyi (1967). Games with Incomplete Information played by "Bayesian" players, I-III, Part I. The Basic Model, *Management Science, Vol. 14, No. 3, pp. 159-182.*

[5] A.Kalai & E.Kalai (2013). Cooperation in Strategic Games Revisited, *Quarterly Journal of Economics, 128, pp 917-966.*

[6] K.A.Konrad (2009). Strategy and Dynamics in Contest, *New York, NY: Oxford University Press*

[7] H.J. Kuhn (2003). Lectures on the Theory of Games, *Princeton University Press ISBN 978-0-691-02772-2* (from photocopies of notes from 1952)

[8] E.P.Lazear & S.Rosen (1981) Rank-Order Tournaments as Optimum Labor Contracts, *Journal of Political Economy89(5), pp.841-864*

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

[10] A.Neyman. & E.Kohlberg (2017). Cooperative Strategic Games, *The Federmann Center for the Study of Rationality, DP 706*

[11] R.W.Rosenthal (1981). Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox, *Journal of Economic Theory 25, pp. 92-100*

[12] L.S.Shapley (1984). Mathematics 147 Game Theory, Lecture Notes, *UCLA Department of Mathematics (1984, 1987, 1988, 1990)*

[13] A.W.Tucker (1983) The Mathematics of Tucker: A Sampler, *The Two-Year College Mathematics Journal, Vol 14, No.3, pp 228-232*

[14] G.Tullock (1980). Efficient Rent Seeking, in *J.M.Buchanan, R.D.Tollison, G.Tullock (Eds), Toward a Theory of Rent-Seeking Society, Texas A&M University Press, pp 97-112*

[15] J.Watson (2018). Contractual Chains, *Working Paper, UCSD, Dept. of Economics*