Does moral play equilibrate?

Immanuel Bomze and Werner Schachinger and Jorgen Weibull

University of Vienna, University of Vienna, Stockholm School of Economics, and IAST

16 October 2018

DOES MORAL PLAY EQUILIBRATE?

Immanuel Bomze*, Werner Schachinger† and Jörgen Weibull‡

October 16, 2018

Abstract. Some finite and symmetric two-player games have no (pure or mixed) symmetric Nash equilibrium when played by partly morally motivated players. The reason is that the "right thing to do" may be not to randomize. We analyze this issue both under complete information between equally moral players and under incomplete information between arbitrarily moral players. We provide necessary and sufficient conditions for the existence of equilibrium and illustrate the results with examples and counter-examples.

JEL codes: C72, D01, D64, D82, D91.

Keywords: Nash equilibrium, morality, homo moralis, social preferences, incomplete information.

1. Introduction

In economics and non-cooperative game theory, economic agents and players are usually assumed to be pure consequentialists, that is, to evaluate their alternative courses of action (consumption or production plans, strategies) exclusively in terms of the consequence for themselves and perhaps also for others. However, people may to some extent also be driven by deontological motivations, such as a wish to "do the right thing" in the given situation. Such a partly morally motivated participant in a public goods game may, for example, contribute the amount that would maximize the group's welfare if everybody would do likewise. An individual who acts accordingly, even when expecting others not to follow suit, is not necessarily irrational, or prey to "magical thinking". Such a person may simply have a goal function that gives some weight to Immanuel Kant's (1785) categorical imperative, to "act only on the maxim that you would at the same time will to be a universal law".

In standard public goods games such partly morally motivated individuals may be behaviorally indistinguishable from altruists, individuals who are pure consequentialists but who attach a positive value to other's well-being. However, in other

*Department of Statistics & Operations Research (ISOR)/Vienna Center of Operations Research (VCOR) & Research Platform Data Science (ds:UniVie), University of Vienna.

†Department of Statistics & Operations Research (ISOR), University of Vienna.

‡Department of Economics, Stockholm School of Economics, and Institute for Advanced Study in Toulouse (IAST).

interactions, a Kantian moralist may behave quite differently from an altruist. Take a 2×2 coordination game, where both players obtain payoff 1 if both use their first pure strategy, 2 if both use their second pure strategy, and otherwise zero. An altruist who expects the opponent to play the first pure strategy will do likewise. By contrast, a Kantian moralist may instead use the second pure strategy. This will result in material payoff zero to both, but the moralist may obtain psychological utility from behaving in a way he wishes all would in such interactions. If two stern moralists would play the coordination game, they would do just fine. However, in some games moralists of intermediate degree, known by both, may not even have a Nash equilibrium, and this may also be the case when player's degree of morality is their private information.

We here explore exactly these questions, more precisely whether symmetric Nash equilibria exist in symmetric and finite games played by partly morally motivated players. As a formal representation of such players we use the *Homo moralis* preferences that Alger and Weibull (2013) showed are evolutionarily stable in populations under assortative random matching.[1] We establish the existence of symmetric Nash equilibria for certain game classes, when played by such players, and we also give examples of simple games with no such equilibria. Our main results, Theorems 1 and 2, establish necessary and sufficient conditions for the existence of symmetric Nash equilibrium between partly morally motivated players under incomplete information about others' degree of morality.

2. Definitions and preliminaries

In this note we consider finite and symmetric games. Let $S = \{1, ..., m\}$ be the set of pure strategies, and let Δ be the associated unit simplex of mixed strategies,

$$\Delta = \left\{ x \in \mathbb{R}^m_+ : e^T x = \sum_{i=1}^{m} x_i = 1 \right\}.$$

Here $e = \sum_{i=1}^{m} e_i$, where e_i is the i:th unity (column) vector, and the superscript T denotes transpose. We write o for the zero vector (origin).

Let A be an $m \times m$-matrix with "material" payoffs, let $\theta \in [0, 1]$ be a player type, and consider the associated payoff function $u_\theta : \Delta^2 \to \mathbb{R}$, defined by

$$u_\theta (x, y) = (1 - \theta)\, x^T A y + \theta \cdot x^T A x, \tag{1}$$

where x and y are (column) vectors in Δ. The parameter θ is the *degree of morality* of *Homo moralis*, with $\theta = 0$ representing pure self-interest, or *Homo oeconomicus*, and

[1] The idea that moral values may have been formed by evolutionary forces can be traced back to at least Darwin (1871). More recent informal treatments include, to mention a few, Alexander (1987) and de Waal (2006).

$\theta = 1$ representing pure (Kantian) morality, or *Homo kantiensis* (Alger and Weibull, 2013). Thus $u_\theta(x, y)$ is the payoff (or utility) to a player with degree of morality θ when using strategy x against an opponent using strategy y in a symmetric game with (material) payoff matrix A. Both player positions have the same set S of pure strategies, A is the matrix of material payoffs to the row player, and $B = A^T$ of those to the column player.

For a given matrix A and degree of morality $\theta \in [0, 1]$, let $\beta_\theta : \Delta \rightrightarrows \Delta$ be the best-reply correspondence of *Homo moralis* of degree θ:

$$\beta_\theta(y) = \arg\max_{x \in \Delta} u_\theta(x, y) \qquad \forall y \in \Delta .$$

Hence, a rational player with *Homo moralis* preferences of type θ will use some strategy x in the subset $\beta_\theta(y)$ if expecting the other player to use mixed strategy $y \in \Delta$. By Weierstrass' maximum theorem, $\beta_\theta(y)$ is a non-empty and compact set for every $\theta \in [0, 1]$ and $y \in \Delta$. However, as will be seen shortly, this set is not always convex. We will study the existence and nature of *fixed points* under β_θ, that is points $x \in \Delta$ such that $x \in \beta_\theta(x)$. These are then the *symmetric Nash equilibria* when two *Homines morales* of the same degree of morality meet.

By Berge's maximum theorem, β_θ is upper hemi-continuous. For $\theta = 0$ the correspondence β_0 is convex-valued. In fact, all its values are then sub-simplices, non-empty subsets of Δ spanned by finitely many vertices. This is the standard setting of non-cooperative game theory, and as is well known, there exists at least one fixed point whenever $\theta = 0$.

3. Games between equally moral players

The analysis in this section generalizes results for symmetric 2×2 games in Section 4 of Alger and Weibull (2013). We here consider strategic interactions under complete information between two equally moral players who play a symmetric $m \times m$ game in material payoffs, for any $m \in \mathbb{N}$. We begin by a 2×2 example that illustrates that the correspondence β_θ need not be convex-valued for positive degrees of morality.

Example 1. *Consider*

$$A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$$

for $a, b > 0$. For $\theta = 0$, there are three fixed points; the two unit vectors, e_1 and e_2, and the mixed strategy

$$x^* = \begin{pmatrix} b/(a+b) \\ a/(a+b) \end{pmatrix}.$$

Note that $x^T A x = ax_1^2 + bx_2^2$ for all $x \in \mathbb{R}^2$. Hence, this term is strictly convex in x and so is $u_\theta(x, y)$, for any given $\theta > 0$ and $y \in \Delta$. Therefore, $\beta_\theta(y) \subseteq \{e_1, e_2\}$. It is

immediate that $e_1 \in \beta_\theta(e_1)$ *iff* $a \geq \theta b$, *and* $e_2 \in \beta_\theta(e_2)$ *iff* $b \geq \theta a$. *So both* e_1 *and* e_2 *are fixed points for* $0 \leq \theta \leq \theta_0 := \min\{a/b, b/a\}$, *and there is only one fixed point for every* $\theta > \theta_0$. *For* $\theta = \theta_0$, *one of* $\beta_\theta(e_1)$ *and* $\beta_\theta(e_2)$ *is a binary set. For all other values of* θ, *both* $\beta_\theta(e_1)$ *and* $\beta_\theta(e_2)$ *are singletons. Next observe that for arbitrary* $\theta > 0$, $\beta_\theta(x^*) \subseteq \{e_1, e_2\}$ *is a singleton when* $a \neq b$, *but* $\beta_\theta(x^*) = \{e_1, e_2\}$, *a non-convex set, if* $a = b$.

Since $B = A^T$ is the payoff matrix of the column player,

$$W(x) = x^T \left(A + A^T\right) x = 2x^T Ax$$

is *welfare*, defined as the sum of the two players' material payoffs when both use strategy $x \in \Delta$. This defines the welfare function $W : \Delta \to \mathbb{R}$. Accordingly, the payoff function of a *Homo moralis* with degree of morality θ can be written in the form

$$u_\theta(x, y) = (1 - \theta)\, x^T Ay + \frac{\theta}{2} \cdot W(x) \ .$$

Hence, if W is concave, then $u_\theta(x, y)$ is concave in $x \in \Delta$, for every $y \in \Delta$, so existence of Nash equilibrium then follows immediately from Kakutani's fixed point theorem.

Proposition 1. *The set of fixed points is non-empty and compact if* $\theta = 0$. *The same is true for every* $\theta > 0$ *if* W *is concave.*

We note beforehand that a sufficient condition for W to be concave is that the symmetric matrix $A + A^T$ is negative-semidefinite. See Proposition 4 below for a more general result.

Example 2. *The payoff matrix*

$$A = \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix}$$

for $a, b > 0$ *makes* $A + A^T$ *indefinite:* $x^T \left(A + A^T\right) x = 2(a + b)x_1 x_2$ *for any* $x \in \mathbb{R}^2$. *However,* W *is concave on* Δ: *there* $W(x) = 2(a + b)\, x_1(1 - x_1)$. *Hence, there exists at least one fixed point. From strict concavity of* W *we know that the sets* $\beta_\theta(y)$ *are singletons for all* $y \in \Delta$ *and* $\theta > 0$. *Using first-order conditions, expressed in* x_1 *only (with* $x_2 = 1 - x_1$), *we conclude*

$$e_1 \in \beta_\theta(y) \iff \frac{d}{dx_1}\, u_\theta(x, y)\big|_{x=e_1} \geq 0 \iff y \in \Delta_1 \,,$$

where $\Delta_1 = \{y \in \Delta : (1-\theta)\,[a-(a+b)y_1] \geq \theta(a+b)\}$, *and likewise*

$$e_2 \in \beta_\theta(y) \iff \frac{d}{dx_1}\, u_\theta\,(x,y)|_{x=e_2} \leq 0 \iff y \in \Delta_2\,,$$

where $\Delta_2 = \{y \in \Delta : (1-\theta)\,[a-(a+b)y_1] \leq -\theta(a+b)\}$. *Finally, for all* $y \in \Delta \setminus (\Delta_1 \cup \Delta_2)$, *we have* $\beta_\theta(y) = \{x\}$, *where*

$$x_1 = \frac{1}{2} + \frac{1-\theta}{2\theta}\frac{a-(a+b)y_1}{a+b} \in (0,1)\,. \tag{2}$$

Since $e_1 \notin \Delta_1$ *and* $e_2 \notin \Delta_2$ *for all* $\theta \in [0,1]$, *neither* e_1 *nor* e_2 *can be fixed points for any* θ. *All fixed points (and we know at least one exists) are thus found by solving* $y_1 = x_1$ *with* x_1 *given by the necessary first-order condition* (2). *This leads to exactly one fixed point for every* $\theta \in (0,1]$, *namely*

$$x = \left(\frac{a+\theta b}{(1+\theta)(a+b)}, \frac{\theta a+b}{(1+\theta)(a+b)}\right)^T$$

In particular, $x_1 = a/(a+b)$ *and* $x_2 = b/(a+b)$ *define the unique fixed point when* $\theta = 0$.

A game-theoretically important class of games in which $A + A^T$ is negative-semidefinite are all constant-sum games (then $A + A^T$ is the a matrix with identical entries), with zero-sum games as the most prominent special case.

Proposition 2. *Let* A *be the payoff matrix of a symmetric constant-sum game. For any* $\theta < 1$, *the set of fixed points is identical with the non-empty set of fixed points when* $\theta = 0$, *while every* $x \in \Delta$ *is a fixed point when* $\theta = 1$.

In other words, all *Homines morales*, except *Homo kantiensis*, behave like *Homo oeconomicus* in all (finite and symmetric two-player) constant-sum games.

The remaining situation to investigate is thus when $\theta > 0$ and W is not concave. We begin with an example.

Example 3. *Consider the generalized Rock-Scissors-Paper (RSP) game matrix*

$$A = \begin{pmatrix} 1 & 2+a & 0 \\ 0 & 1 & 2+a \\ 2+a & 0 & 1 \end{pmatrix}$$

for any $a > -1$. *We note that this is a constant-sum game if and only if* $a = 0$. *For* $\theta = 0$, *the unique symmetric Nash equilibrium strategy is the barycenter* x^o. *As is well-known, this unique equilibrium is unstable in the replicator dynamic for all*

$a < 0$ and asymptotically stable for all $a > 0$.[2] *The function W is strictly concave if $a > 0$ and strictly convex if $a < 0$, because for any $x \in \Delta$:*

$$W(x) = 2 + a \cdot \left(1 - \|x\|^2\right) .$$

Henceforth, assume $a < 0$, fix $0 < \theta < 1$ and observe that $\varnothing \neq \beta_\theta(y) \subseteq \{e_1, e_2, e_3\}$ for all $y \in \Delta$. Moreover, $u_\theta(e_i, e_i) = 1$ for all $i \in S$ while

$$u_\theta(e_1, e_2) = u_\theta(e_2, e_3) = u_\theta(e_3, e_1) = (1 - \theta)(2 + a) + \theta .$$

Hence, $u_\theta(e_1, e_2) > u_\theta(e_i, e_i)$ iff $(1 - \theta)(1 + a) > 0$, so for $-1 < a < 0$, no vertex e_i is a fixed point for any $\theta \in (0, 1)$. Consequently, there exist no fixed point for $0 < \theta < 1$ in generalized RSP-games with values of a in this interval. In other words, if this game is played by two Homines morales of intermediate degree of morality, then there exists no pure or mixed strategy x that they could both play and thereby obtain a Nash equilibrium.

Proposition 1 ensures existence of at least one fixed point if the welfare function W is concave on Δ. If the welfare function instead is strictly convex, then fixed points may not exist. The next result provides necessary and sufficient conditions for existence in the latter case.

Proposition 3. *If W is strictly convex on Δ, then $\beta_\theta(y) \subseteq \{e_1, ..., e_m\}$ for all $y \in \Delta$ and $\theta > 0$, and e_i is a fixed point under β_θ if and only if*

$$a_{ii} \geq \theta a_{kk} + (1 - \theta) a_{ki} \quad \forall k \in S .$$

Proof. If W is strictly convex, so is $u_\theta(x, y)$ in x, and hence the first claim follows. The second claim is then obvious from $u_\theta(e_k, e_i) = \theta a_{kk} + (1 - \theta) a_{ki}$. ■

The usefulness of both Propositions 1 and 3 depends on how easy or hard it is to verify that the welfare function is either concave or strictly convex on the unit simplex. Here are necessary and sufficient conditions for each of these properties.

Proposition 4. *Let C be the expansion of the $(m-1) \times (m-1)$ identity matrix to an $(m-1) \times m$-matrix obtained by appending the column $(-1, \ldots, -1)^T \in \mathbb{R}^{m-1}$. Then W is concave (strictly convex) over Δ if and only if the symmetric $(m-1) \times (m-1)$ matrix*

$$D = C(A + A^T)C^T$$

is negative-semidefinite (positive-definite).

[2] See e.g. Section 3.1.5 in Weibull (1995), and references therein, and see also Benaïm, Hofbauer and Hopkins (2009) for a classification of finite symmetric games into "stable" and "unstable" games.

Proof. First observe that for any $0 < \lambda < 1$ and any two $\{x, y\} \subset \Delta$, we have

$$\lambda W(x) + (1-\lambda)W(y) - W(\lambda x + (1-\lambda)y) = 2\lambda(1-\lambda)v^T A v$$

with $v = x - y \perp e$. Writing $u = (v_1, \ldots, v_{m-1})^T \in \mathbb{R}^{m-1}$, we have for $v \in e^\perp \subset \mathbb{R}^m$ that $v = C^T u$, and $v \neq o$ if and only if $u \neq o$. Hence $2v^T A v = u^T D u$, and the result follows. ■

In some applications the payoff matrix A is symmetric; $A^T = A$. In such *potential* or *partnership* (or *doubly symmetric*) games, it is well-known that average payoff increases along all solution trajectories to the replicator dynamic (see e.g. Section 3.6 in Weibull, 1995). For such games and any positive degree of morality, any global welfare maximizer is a fixed point, and every fixed point is a local maximizer. Formally:

Proposition 5. *Suppose $A^T = A$, and let $\theta > 0$. Then*

(a) $x \in \arg\max_{z \in \Delta} W(z) \implies x \in \beta_\theta(x)$,

(b) $x \in \beta_\theta(x) \implies x \in \arg\max_{z \in \Delta \cap U} W(z)$ *for some neighborhood U of x.*

Proof. Define $h_{\theta,y} : \Delta \to \mathbb{R}$ by $h_{\theta,y}(x) = u_\theta(x, y)$. If $y \in \arg\max_{x \in \Delta} W(x)$ then $W(y) \geq W(x)$ for all $x \in \Delta$, and the directional derivative of W in the direction of $x - y$, evaluated at y, is not positive,

$$4(x-y)^T A y \leq 0 \quad \text{for all } x \in \Delta\,,$$

implying $x^T A y \leq y^T A y$, and therefore $u_\theta(x, y) \leq u_\theta(y, y)$ for all $x \in \Delta$, i.e., $y \in \beta_\theta(y)$.

Next assume $y \in \beta_\theta(y)$. Then y is a global maximizer of $h_{\theta,y}$ over Δ. In particular the directional derivative of $h_{\theta,y}$ in the direction of $x-y$, evaluated at y, is not positive,

$$(1+\theta)(x-y)^T A y \leq 0 \quad \text{for all } x \in \Delta\,.$$

In case that $(x-y)^T A y = 0$ for some x, also the second directional derivative of $h_{\theta,y}$ in the direction of $x - y$, evaluated at y, is not positive,

$$2\theta(x-y)^T A(x-y) \leq 0 \quad \text{for all } x \in \Delta \text{ such that } (x-y)^T A y = 0\,.$$

Now the two displayed inequalities are sufficient for y to be a local maximizer of W, as those inequalities are also statements about first and second directional derivatives of W. ■

In case of symmetric A there may indeed be fixed points $x \in \beta_\theta(x)$ that are local, but not global, maximizers of $x^T A x$ subject to $x \in \Delta$. This happens in Example 1 for small $\theta \geq 0$. If A is not symmetric, neither (a) nor (b) needs to hold. Example 3 shows that (a) can be violated, and violation of both (a) and (b) for $0 \leq \theta < 1$ is demonstrated by Example 2 when $a \neq b$.

Remark 1. *When applied to* 2×2*-games, the above analysis (in agreement with Alger and Weibull, 2013) establishes that at least one symmetric Nash equilibrium always exist between equally moral players.*

4. Incomplete information about others' morality

We now consider strategic interactions between two *Homines morales* who only know their own degree of morality, not that of the opponent. We will call an individual's degree of morality the individual's *type* and use the canonical notation $\Theta = [0,1]$ for the type space. We endow Θ with its Euclidean topology and let μ be a Borel probability measure on Θ, representing the type distribution in the population from which the players are drawn independently at random[3].

A *strategy* is a Borel-measurable function $\xi : \Theta \to \Delta$, assigning to each type $\theta \in \Theta$ a mixed strategy $\xi(\theta) \in \Delta$. A *Nash equilibrium under incomplete information* is a strategy ξ that is a best reply to itself. A strategy ξ is optimal against a mixed strategy $y \in \Delta$ if

$$\xi(\theta) \in \arg\max_{x \in \Delta} u_\theta(x, y) \qquad \forall \theta \in \Theta.$$

It follows from standard measurable-selection theory à la Kuratowski-Ryll-Nardzewski (see e.g. 18.3 and 18.4 in Aliprantis and Border, 2006, or 14.29 and 14.37 in Rockafellar and Wets, 2009) that such an optimal strategy $\xi : \Theta \to \Delta$ exists for each $y \in \Delta$. A strategy $\xi : \Theta \to \Delta$ is a best reply to itself, or constitutes a symmetric Nash equilibrium, if the following condition holds for all $\theta \in \Theta$:

$$\xi(\theta) \in \arg\max_{x \in \Delta} \int_\Theta u_\theta(x, \xi(\tau))\, d\mu(\tau). \tag{3}$$

By linearity of the payoff function with respect to y,

$$\int_\Theta u_\theta(x, \xi(\tau))\, d\mu(\tau) = u_\theta\left(x, \bar{\xi}\right)$$

where

$$\bar{\xi} = \mathbb{E}_\mu[\xi(\theta)] = \int_\Theta \xi(\theta) d\mu(\theta),$$

is the *representative agent*'s mixed strategy. In other words, in order to be a best reply to itself, a strategy $\xi : \Theta \to \Delta$ has to be optimal against its own representative agent's mixed strategy.

Existence is non-trivial. However one may characterize Nash equilibrium by way of first- and second-order optimality conditions. In order to state these, for each type

[3]The analysis in the preceding section thus concerns the special case when μ is a unit probability mass on one type θ; $\mu = \delta_\theta$, see Subsection 4.1 below.

$\theta \in \Theta$ let $H(\theta) = \theta \cdot (A + A^T)$, the Hessian matrix of $u_\theta(\cdot, y)$, for any $y \in \Delta$. For any strategy $\xi : \Theta \to \Delta$, let

$$g(\theta) = \theta \cdot (A + A^T)\xi(\theta) + (1 - \theta) \cdot A\bar{\xi}\,.$$

This is the gradient of the payoff $u_\theta(x, \bar{\xi})$ with respect to $x \in \Delta$, evaluated at $x = \xi(\theta)$. For each pure strategy $i \in S$, let

$$H_i(\theta) = e_i g^T(\theta) + g(\theta)e_i^T - \xi_i(\theta)H(\theta).$$

The matrix $H_i(\theta)$ is a symmetric rank-two update of the Hessian $H(\theta)$, using the gradient $g(\theta) \in \mathbb{R}^m$ and the i:th unit vector $e_i \in \Delta$. Finally, for any strategy ξ, each type $\theta \in \Theta$ and every pure strategy $i \in S$, we define the following *polyhedral cone*

$$\Gamma_i(\theta) = \{v \in e^\perp : \xi_i(\theta)v_j - \xi_j(\theta)v_i \geq 0 \quad \text{for all } j \in S\},$$

where $e^\perp \subset \mathbb{R}^m$ is the $(m-1)$-dimensional tangent space of the unit simplex Δ (that is, all vectors orthogonal to $e \in \mathbb{R}^m$).

The result to follow establishes that, given any type distribution μ, a strategy $\xi : \Theta \to \Delta$ constitutes a Nash equilibrium under incomplete information if and only if three conditions are met: a first-order (Lagrangian) condition, a complementary slackness condition, and a second-order (curvature) condition. The reason why a second-order condition is sufficient is that all types' payoff functions are linear-quadratic in their own strategy choice (in the underlying game). To ease reading, we split the result in two separate statements and provide a joint proof of them only after stating both.

Theorem 1. *For any Borel probability measure μ on Δ, a strategy $\xi : \Theta \to \Delta$ is a best reply to itself if and only if there are Borel-measurable functions $\alpha_0 : \Theta \to \mathbb{R}$ and $\alpha_i : \Theta \to \mathbb{R}_+$ for all $i \in S$ such that, for all pure strategies i and for all types θ:*

$$[H(\theta)\xi(\theta)]_i + (1 - \theta)[A\bar{\xi}]_i + \alpha_0(\theta) + \alpha_i(\theta) = 0\,, \tag{4}$$

$$\alpha_i(\theta)\xi_i(\theta) = 0\,, \tag{5}$$

$$v^T H_i(\theta)v \geq 0 \quad \textit{for all } v \in \Gamma_i(\theta)\,,\ \textit{if } \xi_i(\theta) > 0\,. \tag{6}$$

We say that a strategy $\eta : \Theta \to \Delta$ is a *better reply* than $\xi : \Theta \to \Delta$ for type θ with $\xi_i(\theta) > 0$ if $u_\theta(\eta(\theta), \bar{\xi}) > u_\theta(\xi(\theta), \bar{\xi})$.

Proposition 6. *If (6) is violated for some pure strategy i and type θ, then there exists a better reply for this type θ, namely, the strategy $\eta : \Theta \to \Delta$ that agrees with ξ for all types $\tau \neq \theta$ but has*

$$\eta(\theta) = \xi(\theta) - \frac{\xi_i(\theta)}{v_i} \cdot v\,.$$

Proof. The assertions in Theorem 1 follow from (Bomze 2016, Thm.2.3), formulated for minimizing the negative $-u_\theta(\cdot,\bar\xi)$ there; note that as Δ is compact, we can ignore the index $i=0$ dealing with unbounded feasible rays there. The case of Δ has been dealt already in the previous papers (Bomze 1997a,1997b) where also the arguments for Proposition 6 can be found. ■

We saw in Section 3 an example of a game that has no equilibrium for equally moral players, irrespective of their common degree of morality, as long as it is positive. In terms of the machinery in the present section, the observation can be recast as the statement that this game has no Nash equilibrium under incomplete information when the type distribution μ places unit probability mass on some type $\theta > 0$. The following subsection deals with this case by providing conditions which simplify those of Theorem 1.

4.1. A homogeneous population. Suppose now that μ places unit probability on some $\theta \in \Theta$. Applying the above general machinery, we search for a fixed point of the best-reply correspondence β_θ, i.e. a strategy $\bar x \in \Delta$ which coincides with both $\xi(\theta)$ and $\bar\xi$ by virtue of the special nature of $\mu = \delta_\theta$. We use the notation C_I for a principal submatrix of $C = [C_{ij}]_{(i,j)\in S\times S}$ and of a subvector w_i of $w \in \mathbb{R}^m$, both referring to a (non-empty) index set $I \subseteq S$:

$$C_I = [C_{ij}]_{(i,j)\in I\times I} \quad \text{and} \quad w_I = [w_i]_{i\in I}\,.$$

We consider the $m \times m$ matrix

$$C(\theta) := A + \theta A^T\,.$$

in order to state the characterization for a homogeneous population:

Theorem 2. *A point $\bar x \in \Delta$ with support $I = \{i : \bar x_i > 0\}$ is a fixed point of the correspondence β_θ if and only if conditions (a)–(c) are satisfied for some $\gamma \in \mathbb{R}$:*

(a) $(\bar x, \gamma) \in \mathbb{R}^{m+1}$ satisfies the linear equation system

$$\left\{\begin{array}{lcll} C_I(\theta)\bar x_I - \gamma e_I &=& o & \\ \bar x_i &=& 0 & \text{for all } i \in S\setminus I \\ e^T\bar x &=& 1\,. & \end{array}\right.$$

(b) $(\bar x, \gamma) \in \mathbb{R}^{m+1}$ satisfies the linear inequalities

$$\left\{\begin{array}{lcll} \gamma e - C(\theta)\bar x &\geq& o & \\ \bar x_i &\geq& 0 & \text{for all } i \in I\,. \end{array}\right.$$

(c) For all $i \in I$, the matrix $H_i(\theta)$ is $\Gamma_i(\theta)$-copositive, i.e.,

$$v^T H_i(\theta) v \geq 0 \quad \text{for all } v \in \Gamma_i(\theta)\,.$$

Proof. We apply Theorem 1. First observe that $\xi(\theta) = \bar{\xi} = \bar{x}$ implies $H(\theta)\xi(\theta) + (1-\theta)A\bar{\xi} = C(\theta)\bar{x}$. Further, (5) implies $\alpha_i(\theta) = 0$ for all $i \in I$, so that, with $\gamma = -\alpha_0(\theta)$, we get via (4)

$$[C_I(\theta)\bar{x}_I]_i = [C(\theta)\bar{x}]_i = [H(\theta)\xi(\theta) + (1-\theta)A\bar{\xi}]_i = \gamma \quad \text{for all } i \in I\,.$$

Finally, condition (c) is exactly condition (6) in Theorem 1. ■

Note that by construction $v^T g(\theta) = v^T C(\theta)\bar{x} \leq 0$ for all $v \in \Gamma_i(\theta)$ and all $i \in I$ as $v_j \geq 0$ holds for all $j \in S \setminus I$. Hence condition (c) is ensured if $(A + A^T)$ is negative-semidefinite, as $v \in \Gamma_i(\theta)$ implies also $v_i \leq 0$ and since

$$v^T H_i(\theta) v = 2(v_i\, v^T g(\theta) - \bar{x}_i v^T A v)\,.$$

On the other hand, already local optimality of $\bar{x}$ for the function $u_\theta(.,\bar{x})$ implies for $\theta > 0$ that $[A + A^T]_I$ is negative-semidefinite on $[e_I]^\perp$, a linear subspace of co-dimension one in $\mathbb{R}^I$. This necessary condition can be viewed as a localized version (relative to the face of the simplex that contains $\bar{x}$ in its relative interior) of the sufficient existence criterion in Proposition 1; see Proposition 4.

Remark 2. *In Example 3 we noted that no symmetric Nash equilibrium exists under complete information in a game between equally moral players when $-1 < a < 0$ and $0 < \theta < 1$. Formally, such a situation can be represented as incomplete information with a Dirac measure placed on that particular type θ. Consider instead any continuous type distribution μ on $\Theta = [0,1]$. We may then divide the type space into three disjoint intervals I_k with $\mu(I_k) = 1/3$, for $k = 1,2,3$. If all types in I_k play pure strategy k, then all types $\tau \in \Theta$ best respond to $\bar{\xi} = x^o$, the barycenter of the strategy simplex. Hence, the non-existence of symmetric equilibrium under complete information and equally moral players may be non-robust to arbitrarily small degrees of incomplete information about morality, as measured in the L^1-norm.*

Acknowledgments: We thank Erik Mohlin and Ron Peretz for helpful discussions.

References

[1] Alexander, R. D. (1987): *The Biology of Moral Systems*. New York: Aldine De Gruyter.

[2] Alger, I. and J. Weibull (2013): "Homo moralis—preference evolution under incomplete information and assortative matching," *Econometrica* 81, 2269-2302.

[3] Aliprantis, C.D. and K.C. Border (2006): *Infinite dimensional analysis: a hitchhiker's guide*, 3rd ed. Berlin: Springer.

[4] Benaïm, M., J. Hofbauer and E. Hopkins (2009): "Learning in games with unstable equilibria", *Journal of Economic Theory* 144, 1694-1709.

[5] Bomze, I.M. (1997a): "Evolution towards the maximum clique", *Journal of Global Optimization* 10, 143-164.

[6] Bomze, I.M. (1997b): "Global escape strategies for maximizing quadratic forms over a simplex", *Journal of Global Optimization* 11, 325-338.

[7] Bomze, I.M. (2016): "Copositivity for second-order optimality conditions in general smooth optimization problems", *Optimization* 65, 779-795.

[8] Darwin, C. (1871): *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.

[9] de Waal, F. B.M. (2006): *Primates and Philosophers. How Morality Evolved*. Princeton: Princeton University Press.

[10] Kant, I. (1785): *Grundlegung zur Metaphysik der Sitten*. [In English: *Groundwork of the Metaphysics of Morals*. 1964. New York: Harper Torch books.]

[11] Rockafellar, R.T., and R.J.-B. Wets (2009): *Variational analysis*, 3rd printing. Berlin: Springer.

[12] Weibull, J. (1995): *Evolutionary Game Theory*. MIT Press: Cambridge MA.