Economic Theory
https://doi.org/10.1007/s00199-020-01246-4

RESEARCH ARTICLE

Does moral play equilibrate?

Immanuel Bomze[1] · **Werner Schachinger**[2] · **Jörgen Weibull**[3]

Received: 17 July 2019 / Accepted: 10 January 2020

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 players of arbitrary degrees of morality. We provide necessary and sufficient conditions for the existence of equilibrium and illustrate the results with examples and counter examples.

Keywords Nash equilibrium · Morality · Homo moralis · Social preferences · Incomplete information

JEL classification C72 · D01 · D64 · D82 · D91

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 participant in a public goods game may, for example, contribute the amount that would maximize the group's welfare if everybody would do likewise, in line with 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."

✉ Jörgen Weibull
jorgen.weibull@hhs.se

1 Department of Statistics and Operations Research (ISOR) Vienna Center of Operations Research (VCOR) Research Platform Data Science (ds:UniVie), University of Vienna, Vienna, Austria

2 Department of Statistics and Operations Research (ISOR), University of Vienna, Vienna, Austria

3 Department of Economics, Stockholm School of Economics, Stockholm, Sweden

Published online: 22 January 2020

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 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/she 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. Theorem 1 and Proposition 6 together 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 $S = \{1, \ldots, m\}$ 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 (the 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

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

[1] The idea that moral values may have been formed by evolutionary forces can be traced back to at least Darwin (1871). More recent treatments include Alexander (1987), de Waal (2006), and Bergstrom (1995, 2009).

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 $\theta = 1$ 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. The game being symmetric, $B = A^T$ is the matrix of material payoffs 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 (with respect to y and θ). 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 $\theta = 0$. Likewise, for $\theta = 1$ there is always a symmetric Nash equilibrium, namely (x, x) for any x in the non-empty set $\arg\max_{x \in \Delta} x^T A x$.

Remark 1 For any given material payoff matrix A, and any sequence $\langle \theta_t \rangle \to 0$, for each $t \in \mathbb{N}$ suppose that $\left(x^{(t)}, x^{(t)}\right)$ is a Nash equilibrium between two Homines morales of degree of morality $\theta_t > 0$. The set Δ^2 being compact, this equilibrium sequence contains a convergent subsequence. By upper hemi-continuity of β_θ, the limit point, (x^*, x^*) is a Nash equilibrium when $\theta = 0$. Hence, the requirement of robustness with respect to a small degree of morality constitutes a refinement of Nash equilibrium in standard game theory, see Example 1.

3 Games between equally moral players

The analysis in this section generalizes results for symmetric 2×2 games in Sect. 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 $m \in \mathbb{N}$. If players are only interested in their own material payoff, $\theta = 0$, then the best-reply correspondence is convex-valued, and a symmetric Nash equilibrium exists by standard arguments. However, as is illustrated in the following 2×2 example, the correspondence β_θ need not be convex-valued for positive degrees of morality.

Example 1 Consider the coordination game

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

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 $x^T Ax = ax_1^2 + bx_2^2$ for all $x \in \mathbb{R}^2$. Hence, this term is strictly convex in x, and so is also $u_\theta(x, y)$, for any given $\theta > 0$ and $y \in \Delta$. Therefore, $\beta_\theta(y) \subseteq \{e_1, e_2\}$. It is immediate that $e_1 \in \beta_\theta(e_1)$ for all $\theta \in [0, 1]$, and $e_2 \in \beta_\theta(e_2)$ iff $\theta \leq b/a$. So both e_1 and e_2 are fixed points for all $0 \leq \theta \leq b/a$, and there is only one fixed point for every $\theta > b/a$. For $\theta = b/a$, $\beta_\theta(e_2)$ is a binary set. For all other values of θ, both $\beta_\theta(e_1)$ and $\beta_\theta(e_2)$ are singletons. In sum, x^* is a fixed point only when $\theta = 0$, e_1 is always a fixed point, and e_2 is a fixed point iff $\theta \leq b/a$. In particular, each of the pure equilibria is robust to a small degree of morality, but the mixed equilibrium is not.

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 the 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* $\theta \in \{0, 1\}$. *The same is true for every* $\theta > 0$ *if* W *is concave.*

Clearly, a sufficient condition for W to be concave is that the symmetric matrix $A + A^T$ is negative semidefinite. See Proposition 4 for a more general result. In the following example, moral players, irrespective of how weak their morality is (as long as θ is positive), have strict preferences among mixed strategies, something that is never the case when players lack morality ($\theta = 0$).

Example 2 Consider the Hawk–Dove game

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

for $a, b > 0$. Then $A + A^T$ is indefinite: $x^T \left(A + A^T\right) x = 2(a+b)x_1x_2$ 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

$$e_1 \in \beta_\theta(y) \iff \frac{\mathrm{d}}{\mathrm{d}x_1} u_\theta(x, y)|_{x=e_1} \geq 0 \iff y \in \Delta_1,$$

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

$$e_2 \in \beta_\theta(y) \iff \frac{\mathrm{d}}{\mathrm{d}x_1} u_\theta(x, y)|_{x=e_2} \leq 0 \iff y \in \Delta_2,$$

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

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

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, $x_1 = a/(a+b)$ defines the unique fixed point when $\theta = 0$, and $x_1 = 1/2$ the unique fixed point when $\theta = 1$. In this example, the unique symmetric equilibrium is robust to a small degree of morality.

A game-theoretically important class of games in which W is concave on Δ are all constant-sum games (then $A + A^T$ is 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 (as in Example 1). We begin with an example showing that existence of a symmetric Nash equilibrium cannot be taken for granted.

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 exists no fixed point for $0 < \theta < 1$ in generalized RSP games with values of a in this interval. In sum, in these games, when $-1 < a < 0$, there is no symmetric Nash equilibrium at any degree of morality θ between pure self-interest ($\theta = 0$) and pure Kantian morality ($\theta = 1$).

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 $\beta_\theta(y) \subseteq \{e_1, \ldots, e_m\}$ for all $y \in \Delta$ and $\theta > 0$, and e_i is a fixed point under β_θ if and only if*

$$a_{ii} \geq \theta a_{kk} + (1-\theta)\, a_{ki} \quad \forall k \in S. \tag{3}$$

In turn, (3) is satisfied for all small θ if (e_i, e_i) is a strict Nash equilibrium.

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 third claim follows by continuity. □

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

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*

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

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

is negative semidefinite (positive definite).

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

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

with $v = x - y$ being orthogonal to 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 $2v^T Av = u^T Du$ by definition of D, 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 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 welfare 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,

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

implying $x^T Ay \leq y^T Ay$, 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,

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

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

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

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, see, e.g., Bomze (2002). □

In case of symmetric A there may indeed be fixed points $x \in \beta_\theta(x)$ that are local, but not global, maximizers of $x^T Ax$ subject to $x \in \Delta$. This happens in Example 1 for small $\theta \geq 0$. In other words, any fixed point of β_θ is neutrally stable in partnership games

(again, see Bomze 2002). If A is not symmetric, neither (a) nor (b) need 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 2 Alger and Weibull (2013) establish that every symmetric 2×2-game admits at least one symmetric Nash equilibrium between equally moral players, irrespective of their common degree of morality. This can be easily confirmed by the above analysis, since a univariate quadratic function is either concave or strictly convex.

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.

A *strategy* is a Borel-measurable function $\xi : \Theta \to \Delta$, assigning to each type $\theta \in \Theta$ a strategy $\xi(\theta) \in \Delta$. 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 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, equivalently, (ξ, ξ) constitutes a symmetric Nash equilibrium under incomplete information, if the following condition holds for all $\theta \in \Theta$:

$$\xi(\theta) \in \arg\max_{x \in \Delta} \int_\Theta u_\theta(x, \xi(\tau))\, d\mu(\tau). \tag{4}$$

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 $\theta \in \Theta$ 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

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

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, type $\theta \in \Theta$ and pure strategy $i \in S$, we define the following *polyhedral cone:*

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

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 condition that significantly differs from the usual global convexity/concavity requirements on $H(\theta)$ and W. The reason why this particular second-order condition is sufficient is that all types' payoff functions are linear-quadratic in their own strategy choice (in the underlying game). We split the statement of the result into two parts and provide a joint proof.

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 $\alpha_i : \Theta \to \mathbb{R}$ for $i = 0, 1, \ldots, m$ such that, for all $i \in S$ and $\theta \in \Theta$:*

$$[H(\theta)\xi(\theta)]_i + (1-\theta)[A\bar{\xi}]_i + \alpha_0(\theta) + \alpha_i(\theta) = 0, \tag{5}$$

$$\alpha_i(\theta)\xi_i(\theta) = 0, \tag{6}$$

$$v^T H_i(\theta)v \geq 0 \quad \forall v \in \Gamma_i(\theta) \ \text{ if } \xi_i(\theta) > 0. \tag{7}$$

To state the second part of the result, we call a strategy $\eta : \Theta \to \Delta$ a *better reply* than $\xi : \Theta \to \Delta$ against ξ for type θ if $u_\theta(\eta(\theta), \bar{\xi}) > u_\theta(\xi(\theta), \bar{\xi})$.

Proposition 6 *If (7) is violated for some pure strategy i and type θ with $\xi_i(\theta) > 0$ and some $v \in \Gamma_i(\theta)$, 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 Bomze (1997a, b), where also the arguments for Proposition 6 can be found. □

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. We proceed to establish a sufficient condition for the existence of symmetric Nash equilibrium under incomplete information when the type distribution μ has no atoms. It follows from this result that the non-existence of symmetric equilibrium under complete information and equally moral players, in Example 3, is non-robust to arbitrarily small degrees of incomplete information about morality, as measured, e.g., in the L^1-norm.

To state the result, we write $\tilde{\beta}_\theta(x) \subseteq S$ for the set $\arg\max_{i \in S} u_\theta(e_i, x)$ of *pure* best replies to a mixed strategy $x \in \Delta$ and $supp(x) \subseteq S$ for the support of x.

Proposition 7 *Suppose that μ can be represented by a probability density function $f : \Theta \to \mathbb{R}_+$. If the welfare function $W : \Delta \to \mathbb{R}$ is convex and there exists a strategy $x^* \in \Delta$ such that*

$$\tilde{\beta}_\theta(x^*) = \tilde{\beta}_\tau(x^*) = supp(x^*) \tag{8}$$

for all types θ and τ in the support of f, then there exists a Nash equilibrium (ξ, ξ) under incomplete information, and $\bar{\xi} = x^*$.

Proof Let $x^* \in \Delta$ be as stated. Partition the type space Θ into m cells B_i such that $\mu(B_i) = x_i^*$ for each $i \in S$. This is possible since μ has no atoms. Define $\xi : \Theta \to \Delta$ by setting $\xi(\theta) = e_i$ for all $\theta \in B_i$ and $i \in S$. Then $\bar{\xi} = x^*$. Since W is convex, $u_\theta(x, x^*)$ is convex in $x \in \Delta$, and thus $i \in \tilde{\beta}_\theta(x^*) \Rightarrow e_i \in \beta_\theta(x^*)$ for all $i \in S$ and $\theta \in \Theta$. Hence, $\xi(\theta) \in \arg\max_{x \in \Delta} u_\theta\left(x, \bar{\xi}\right)$ for all $\theta \in \Theta$, so (ξ, ξ) is a Nash equilibrium. □

We note that since, for each $x \in \Delta$ and $\theta \in \Theta$, $\tilde{\beta}_\theta(x)$ belongs to the finite power set of S, and $u_\theta(e_i, x)$ is continuous (in fact, linear) in θ, $\tilde{\beta}_\theta(x)$ is piecewise constant on Θ, for any given $x \in \Delta$. Thus, the equality in (8) holds generically for all θ and τ in some open set $U \subset \Theta$.

Applying Proposition 7 to Example 3, let x^* be the barycenter x^o of the mixed-strategy simplex Δ. For any atom free type distribution μ, partition the type space Θ into three cells, each with probability measure $1/3$. Allocate all individuals in each type cell to one and the same pure strategy. For example, the 1/3 least moral individuals in the support of f may play strategy 1, the morally intermediate 1/3 of the population may play strategy 2, and the most moral 1/3 of the population may play strategy 3. For $a < 0$ the welfare function W is convex and symmetric, so all three pure strategies are optimal against x^*, for any $\theta \in (0, 1)$. The hypothesis of Proposition 7 is thus met, so there exists a symmetric Nash equilibrium and it has x^o as its outcome.

Acknowledgements Open access funding provided by Stockholm School of Economics. We are grateful for helpful comments from the editorial team, an anonymous referee, Erik Mohlin, Jonathan Newton, Ron Peretz, and the participants of the "Oberseminar Optimization" at the University of Augsburg. Weibull gratefully acknowledges financial support from the Knut and Alice Wallenberg Research Foundation (Sweden), and from the National Research Agency ANR, Chaire IDEX ANR-11-IDEX-0002-02 (France).

References

Alexander, R.D.: The Biology of Moral Systems. Aldine De Gruyter, New York (1987)
Alger, I., Weibull, J.: Homo moralis–preference evolution under incomplete information and assortative matching. Econometrica **81**, 2269–2302 (2013)
Aliprantis, C.D., Border, K.C.: Infinite Dimensional Analysis: A Hitchhiker's Guide, 3rd edn. Springer, Berlin (2006)
Benaïm, M., Hofbauer, J., Hopkins, E.: Learning in games with unstable equilibria. J. Econ. Theory **144**, 1694–1709 (2009). https://doi.org/10.1016/j.jet.2008.09.003
Bergstrom, T.: On the evolution of altruistic ethical rules for siblings. Am. Econ. Rev. **85**, 58–81 (1995)
Bergstrom, T.: Ethics, Evolution, and Games among Neighbors. University of California, Santa Barbara (2009)
Bomze, I.M.: Evolution towards the maximum clique. J. Glob. Optim. **10**, 143–164 (1997a)
Bomze, I.M.: Global escape strategies for maximizing quadratic forms over a simplex. J. Glob. Optim. **11**, 325–338 (1997b)
Bomze, I.M.: Regularity versus degeneracy in dynamics, games, and optimization: a unified approach to different aspects. SIAM Rev. **44**, 394–414 (2002)
Bomze, I.M.: Copositivity for second-order optimality conditions in general smooth optimization problems. Optimization **65**, 779–795 (2016)
Darwin, C.: The Descent of Man, and Selection in Relation to Sex. John Murray, London (1871)
de Waal, F.B.M.: Primates and Philosophers. How Morality Evolved. Princeton University Press, Princeton (2006)
Kant, I.: Grundlegung zur Metaphysik der Sitten. Hartknoch, Riga (1785). [In English: Groundwork of the Metaphysics of Morals. Harper Torch Books, New York (1964).]
Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (2009). 3rd printing
Weibull, J.: Evolutionary Game Theory. MIT Press, Cambridge (1995)