Manuscript submitted to
AIMS' Journals
Volume **X**, Number **0X**, XX **200X**

doi:10.3934/xx.xx.xx.xx

pp. **X–XX**

APPROACHABILITY IN POPULATION GAMES

Dario Bauso*

Jan C. Willems Center for Systems and Control, ENTEG, Fac. Science and Engineering, University of Groningen and Dip. di Ingengneria, Università di Palermo, IT

Thomas W.L. Norman

Magdalen College, Oxford, UK

(Communicated by the associate editor name)

Abstract. This paper reframes approachability theory within the context of population games. Thus, whilst a player still aims at driving her average payoff to a predefined set, her opponent is no longer malevolent but instead is extracted randomly at each instant of time from a population of individuals choosing actions in a similar manner. First, we define the notion of *1st-moment approachability*, a weakening of Blackwell's approachability. Second, since the endogenous evolution of the population's play is then important, we develop a model of two coupled partial differential equations (PDEs) in the spirit of mean-field game theory: one describing the best-response of every player given the population distribution, the other capturing the macroscopic evolution of average payoffs if every player plays her best response. Third, we provide a detailed analysis of existence, nonuniqueness, and stability of equilibria (fixed points of the two PDEs). Fourth, we apply the model to regret-based dynamics, and use it to establish convergence to Bayesian equilibrium under incomplete information.

1. Introduction.

We consider a game played by a large population of individuals in continuous time. At each instant, every individual engages in play with a random opponent extracted from the population and the resulting payoff, which depends on the actions of both players, is a vector. Such vector payoffs can be interpreted as deriving from a collection of noninterchangeable goods; formally, we can think of the completeness axiom being satisfied along each dimension of our vector but failing across them, giving a special case of Aumann's [1] framework.[1] Alternatively, each player may be representative of a group of individuals whose preferences may not be aggregated into a single ordering, so that the vector payoff has one component for each individual in the group. Notably, payoff vectors also arise naturally when analysing a player's regret from not having made each possible deviation.

In a standard repeated game, a set of payoffs is approachable [9] if a player can guarantee that for any strategies used by her opponent, her average payoff uniformly approaches the set with probability 1. The uniformity requirement ensures the approachable payoffs even in the worst-case scenario that the player faces a malevolent

2010 *Mathematics Subject Classification.* Primary: 58F15, 58F17; Secondary: 53C35.
Key words and phrases. Approachability, Population games, Mean-field games.
The first author is supported by NSF grant xx-xxxx.
* Corresponding author: Dario Bauso.

[1] Indeed, vector payoffs may also be appropriate when the continuity axiom fails (see [11]).

opponent who seeks to minimise her payoffs. However, here we are concerned with a player facing, not a single malevolent opponent, but a sequence of opponents drawn at random from our population. This being the case, we are interested in whether her *expected* payoffs against the population distribution of play can approach a particular set of payoffs, allowing us to weaken Blackwell's approachability condition. The distribution of play is then endogenous and evolving in response to the players' choices, and it is important that we allow for this evolution; under Blackwell's worst-case approach, by contrast, any opposing play is allowed and hence the details of its evolution can be ignored. We offer conditions guaranteeing that a player can approach a given set of payoffs on average in such a population-game setting.

For an application of this setting, consider the Google AdWords Auction (see [6] for background), whereby a pool of advertisers bid on keywords related to their products. When an Internet user runs a Google search related to one of these keywords, an auction is triggered amongst the relevant bidders, whose bids and website characteristics then determine their ranking in the displayed adverts. Edelman, Ostrovsky, and Schwarz [13] and Varian [40] solve for equilibrium of this auction under complete and incomplete information. But the auction is run billions of times each month, each advertiser facing an unknown set of opponents from the pool in each instance. Our population-game analysis incorporates these features, and offers conditions and a strategy under which a bidder could ensure himself payoffs from a particular approachable set. Moreover, if the players' payoffs are maximal regrets, our model provides a foundation for the above authors' Bayesian equilibrium.

Main results. First, we provide a new model that combines approachability and population games. Given that the opponent is randomly extracted from the population, the approach by Blackwell—which looks at the worst-case payoff—may appear conservative. Thus, we relax Blackwell's conditions, assuming that the opponent is not malevolent but instead is simply extracted from a population with given distribution; we call this 1*st-moment approachability*. Second, to model the endogenous evolution of play, we build upon the theory of mean-field games and adapt the concept of mean-field equilibrium to our evolutionary set-up; we call this *self-confirmed equilibrium*. The mean-field game constitutes a generalized version of the approachability problem as the value function of the mean-field game can be viewed as a measure of the distance of a point from the target. In addition, the mean-field game is a *consistent* model as the *advection or transport* equation (representing the macroscopic evolution of the density distribution of the players) provides an accurate description of the population evolution when all players implement approachability strategies. Third, we discuss existence and nonuniqueness of the equilibrium. Finally, we explore the regret interpretation of our model; whereas 1st-moment approachability of nonpositive regrets no longer implies Nash equilibrium (as in [20]), we show that nonpositive maximal regret does imply one-shot Bayesian equilibrium under incomplete information.

Related literature. The theory of "approachability" dates back to Blackwell [9] and culminates in the well known Blackwell's Theorem. Approachability arises in several areas of game theory, such as allocation processes in coalitional games [28], regret minimization [30, 20], adaptive learning [12, 16, 18, 19], excludability and bounded recall [31], and weak approachability [41], just to name a few. For instance, in coalitional games one asks whether the core is an approachable set, and which allocation processes can drive the complaint vector to that set. In regret minimization, one considers the nonpositive orthant in the space of regrets; a player tries to adjust her strategy based on the current regret so as to make that set approachable by the regret vector. Once all players have nonpositive regret, the resulting outcome is an equilibrium for the game. This idea of adapting the new action to the current state of the game is common to adaptive learning and evolutionary games as well, but in regret-based dynamics the state is in payoff (rather than strategy) space. The evolutionary literature on population games is large and developed (see [36]), but distinct from the current enterprise by virtue of our focus on rational, rather than myopic, players.

Despite its discrete-time nature in the original Blackwell formulation, approachability has been extended to continuous-time repeated games, thus showing common elements with Lyapunov theory [20]. Though first formalized in finite-dimensional spaces, a definition of approachability in infinite-dimensional space has been provided by Lehrer [29]. Approachability can be viewed as the extension (to a vector space) of the von Neumann minmax theorem [42] and can be reframed within differential games and as such can be studied using differential calculus and stability theory [10, 33, 38]. [2]

The approachability principle is also behind the notion of excludability; along this line, some authors investigate which sets are approachable and which ones excludable under imperfect information (bounded recall, delayed and/or stochastic monitoring) [31]. Connected to approachability as well is the concept of "attainability." Attainability is a new notion developed in [7, 32] in the context of 2-player continuous-time repeated games with vector payoffs. Attainability arises in many contexts such as transportation networks, distribution networks, production networks applications. The main question is: "Under what conditions does a strategy for the Row player exist such that the cumulative payoff converges (in the lim sup sense) to a preassigned set (in the space of vector payoffs) independently of the strategy used by the Column player?"

A second stream of literature we follow in the present study is the one on *mean-field games*. This theory originated in the work of M. Y. Huang, P. E. Caines and R. Malhamé [23, 21, 22], and independently in that of J. M. Lasry and P. L. Lions [25, 26, 27], where the now standard terminology of Mean-Field Games (MFG) was introduced. Explicit solutions in terms of mean-field equilibria are not common unless the problem has a linear-quadratic structure, see [5]. Mean-field games have connections to evolutionary games (see for instance [24]) and large games [2]. Actually, both the *anonymous game* in [24] and the *large game* in [2] build upon the notion of mass interaction and can be seen as a stationary mean field.

This paper is organized as follows. In Section 2, we set up the problem. In Section 3, we establish the main results of the paper. In Section 4, we apply the model to a regret-based setting, and show under incomplete information that nonpositive maximal regrets that are approachable in 1st moment must be one-shot Bayesian equilibria. Finally, in Section 5, we draw concluding remarks.

Notation. We view vectors as columns. For a vector x, we use x_i to denote its ith coordinate component. Occasionally we may write $(x)_{i=1,\dots,m}$ to denote an m-dimensional column vector. For two vectors x and y, we use $x < y$ ($x \le y$) to denote $x_i < y_i$ ($x_i \le y_i$) for all coordinate indices i. We let x^T denote the

[2]Still within the realm of differential games, it is worth noting that the notion of nonanticipative behavior strategies has a long history [4, 14, 38, 35, 39]. Actually, it turns out that classical feedback strategies in differential games are special nonanticipative strategies.

transpose of a vector x, and $\|x\|$ its Euclidean norm. We write $P(x)$ to denote the projection of a vector x on a set X, and $\mathrm{dist}(x, X)$ for the distance from x to X, i.e. $P(x) = \arg\min_{y\in X} \|x - y\|$ and $\mathrm{dist}(x, X) = \|x - P(x)\|$, respectively. We also denote by *conv* the convex hull of a given set of points. ∂_x indicates the first partial derivative with respect to x.

2. **The Model.**

Consider a population game Γ where continuously in time every individual matches with an opponent randomly extracted from a population and receives a vector payoff. Let $A = \{1, 2, \ldots, n\}$ be the finite set of actions of every individual. Payoffs correspond to the elements of the discrete set $M = \{M_{lk}, l, k \in A\}$, where $M_{lk} \in \mathbb{R}^m$ (each entry M_{lk} is an m-dimensional vector). Let $X := conv\{M_{lk} |\, l, k \in A\}$ be the state space, where *conv* denotes the *convex hull*.

This can be modelled as a two-player game where the Row player uses a Markovian pure strategy

$$\sigma : X \times [0, \infty[\to A \text{ such that } a(t) := \sigma(x(t), t),$$

while the Column player plays the mixed strategy $q(t) \in \Delta(A)$ corresponding to the distribution of actions in the population at time t. In the above definition of the Markovian strategy, the state $x(t)$ is the solution of the following differential equation in X:

$$\begin{cases} \dot{x}(t) = u(\sigma(x(t), t), q(t)) - x(t), \\ x(0) = x_0 \in X, \end{cases} \tag{1}$$

where the mapping $(a(t), q(t)) \mapsto u(a(t), q(t)) \in X$ is the instantaneous expected payoff (over the set M) and the initial state x_0 is generated according to a distribution law $\rho_0(x)$. Hence, the state evolves endogenously in response to $\sigma(x(t), t)$'s expected payoff against the mixed strategy $q(t)$.

Equation (1) is in the same spirit as in Hart and Mas-Colell's paper [20] on continuous-time approachability. To see this, consider the time-average expected (over opponent's play) payoff defined as

$$\gamma(s) = \frac{1}{s}\int_0^s u(a(\tau), q(\tau))\mathrm{d}\tau \in \mathbb{R}^m.$$

If we rescale the time window using $s = e^t$, take $x(t) = \gamma(e^t)$ and differentiate with respect to t, we obtain the differential equation (1).[3]

The one-shot vector-payoff game $(A, \Delta(A), u)$ is denoted G, and we will say that the game in continuous time Γ is *based on* G. The game Γ is played over the time interval $[0, T]$. The objective of a player is to approach a given, generally time-varying, target $y : [0, T] \to X$. Then, for each player, consider a running cost $g : X \times X \to [0, +\infty[$, $(x, y) \mapsto g(x, y)$ of the form:

$$g(x, y) \;=\; \frac{1}{2}\left[(y - x)^T\, Q\, (y - x)\right], \tag{2}$$

where $Q > 0$ and symmetric. Also consider a terminal cost $\Psi : X \times X \to [0, +\infty[$, $(x, y) \mapsto \Psi(x, y)$ of the form

$$\Psi(x, y) = \frac{1}{2}(y - x)^T S(y - x), \tag{3}$$

[3] Note that, after rescaling the time window, we have

$$x(0) = \int_0^1 u(\sigma(x(\tau), \tau), q(\tau))\mathrm{d}\tau \in \mathbb{R}^m.$$

where $S > 0$. The payoffs given by (2) and (3) are the current and final distances from the set that each player wishes to approach. As in standard approachability, one wishes to minimize that distance.

The population's approachability problem in its generic form is then the following:

Problem. *Given:*

- *the initial state* x_0*, determined by the density* ρ_0*;*
- *a finite horizon* $T > 0$*;*
- *a suitable running cost* $g : X \times X \to [0, +\infty[$*,* $(x, y) \mapsto g(x, y)$*, as in (2);*
- *a terminal cost* $\Psi : X \times X \to [0, +\infty[$*,* $(y, x) \mapsto \Psi(y, x)$*, as in (3);*
- *and a suitable dynamics for* x*, as in (1),*

solve

$$\inf_{a(\cdot):=\sigma(x(t),t)\in\mathcal{C}} \left\{ J(x_0, a(\cdot), q(\cdot)) = \int_0^T g(x(t), y(t))dt + \Psi(x(T), y) \right\}, \tag{4}$$

where $\mathcal{C}$ *is the set of all measurable functions* $a(\cdot)$ *from* $[0, +\infty[$ *to* A*.*

Thus, Problem 2 analyzes the approachability of a given target in the space of vector payoffs on the part of a population of individuals.

This problem gives rise to two different interpretations, involving an asynchronous or synchronous scenario.

Asynchronous scenario. Following traditional "population-game dynamics" from evolutionary game theory, the state $x(t) \in \mathbb{R}^m$ represents the current average position over the population (though in terms of payoffs, rather than actions). Between time t and $t + dt$, a randomly chosen individual computes $q(t)$ and $x(t)$ and changes his strategy to an action $\sigma(x(t), q(t))$ that draws the state $x(t)$ closer to the target y. Hence, the target in this case is a population one. This would yield the differential equations

$$\left\{ \begin{array}{l} \dot{x} = u(\sigma(x, q), q) - x \\ \dot{q} = \sigma(x, q) - q. \end{array} \right. \tag{5}$$

Synchronous scenario. Following the traditional "mean-field game dynamics", the state $x(t) \in \mathbb{R}^m$ represents, not the average state, but the state of a single player (in analogy to a particle). Between time t and $t+dt$, all individuals simultaneously compute the distribution $q(t) \in \Delta(A)$ of actions in the population and change their strategies to actions $\sigma(x(t), q(t))$ that draw their states closer to the target y. Note that, as the strategies $\sigma(x(t), q(t))$ are Markovian, each player needs to know the distribution ρ at time t to compute $q(t) \in \Delta(A)$. More specifically, we require a probability density function $\rho : X \times [0, +\infty[\to \mathbb{R}$, $(x, t) \mapsto \rho(x, t)$, representing the density of the players whose state is x at time t, and satisfying $\int_{\mathbb{R}} \rho(x, t)dx = 1$ for every t.[4] For the distribution of actions in the population $q(t) \in \Delta(A)$, we have

$$\left\{ \begin{array}{l} q(t) \in \Delta(A)\, s.t.\, q_j(t) = \int_{R_j} \rho(x, t)dx, \\ \qquad R_j := \{x \in \mathbb{R}^m |\, \sigma(x, t) = j\},\, \forall j \in A,\ \forall t \in (0, T]. \end{array} \right. \tag{6}$$

[4] Function ρ is the density distribution of states. We assume that the initial distribution $\rho_0(x)$ is exogenously given with $\rho(x, 0) = \rho_0(x)$. The density $\rho(x, t)$ at any time $t > 0$ is, by contrast, endogenously given based on $\rho_0(x)$ and the best responses of the players used to define their dynamics.

3. **Main Result.** This section outlines the main result of the paper, Theorem 3.1, which establishes conditions for approachability in 1st-moment and introduces self-confirmed equilibrium. Theorems 3.2 and 3.3 elaborate on existence and nonuniqueness respectively. The concept of 1st-moment approachability is a relaxation of standard approachability when Nature (the Column player) is not malevolent. Column's strategy is not exogenous or fixed, but rather it arises from the collective behavior of the system as a whole, i.e. from the average distribution obtained when all players play their best responses. We focus on the synchronous case above, which has a more complex structure.

3.1. **Approachability in 1st-moment.** We will make use of the notion of *projected game*, which we recall next. Let $\lambda \in \mathbb{R}^m$ and denote by $\langle \lambda, G \rangle$ the one-shot zero-sum game whose set of players and their actions are as in game G, with the payoff that the Column player pays to the Row player being $\lambda^T u(a,q)$ for every $(a,q) \in A \times \Delta(A)$. Let $\eta(\lambda)$ be defined as

$$\eta(\lambda) := \min_a \lambda^T u(a,q).$$

In the case of state-dependent payoffs, which occurs when we consider the game whose payoff is

$$f(u(a,q),x) = u(a,q) - x,$$

the above expression for the minimum can be modified to:

$$\begin{array}{rcl} \eta_x(\lambda) & := & \min_a \langle \lambda, f(u(a,q),x) \rangle \\ & = & \min_a \langle \lambda, u(a,q) - x \rangle. \end{array} \tag{7}$$

We recall next the geometric (approachability) principle [9, 38] that lies behind Blackwell's Theorem. To introduce this principle, let Φ be a closed and convex set in $\mathbb{R}^m$ and let $P(x)$ be the projection of any point $x \in \mathbb{R}^m$ (closest point to x in Φ).

Proposition 1. (Blackwell's Approachability Principle) A closed and convex set Φ in $\mathbb{R}^m$ is approachable by the Row player if for every $x \in X$ there exists a strategy $a_1 \in \Delta(A)$ for Row such that (8) holds true for every strategy $a_2 \in \Delta(A)$ of the Column player:

$$\langle x - P(x), u(a_1,a_2) - P(x) \rangle \leq 0. \tag{8}$$

The approachability conditions established in the following theorem are a specialization of the above ones when we take $\Phi = \{y(t)\}$ and under the assumption that Column is committed to playing a mixed strategy $q(t) \in \Delta(A)$, at time $t \in (0,T]$.

Theorem 3.1 (Approachability in 1st-moment, self-confirmed equilibrium)**.** *Let the Row and Column players play $\sigma(x,t)$ and $q(t)$, respectively, where*

$$\left\{ \begin{array}{l} \sigma(x,t) \in \arg\min_{a \in A} \langle \lambda, f(u(a,q),x) \rangle, \quad \lambda = \frac{\partial_x v(x,t)}{\|\partial_x v(x,t)\|} \\ \\ q(t) \in \Delta(A)\, s.t.\, q_j(t) = \int_{R_j} \rho(x,t) dx, \\ R_j := \{x \in X \,|\, \sigma(x,t) = j,\, j = \arg\min_a \langle \lambda, f(u(a,q),x) \rangle\} \\ \qquad \forall j \in A,\ \forall t \in (0,T] \end{array} \right. \tag{9}$$

and where $v(x,t)$ and $\rho(x,t)$ are solutions of the following set of two coupled partial differential equations:

$$\begin{cases} \partial_t v(x,t) + \|\partial_x v\| \eta_x(\lambda) + \frac{1}{2}\langle x-y, Q(x-y)\rangle = 0, \\ \quad \textit{in } \mathbb{R}^m \times [0,T[, \\ \\ v(x,T) = \Psi(y(T),x), \ \textit{in } \mathbb{R}^m, \\ \\ \partial_t \rho(x,t) + div(\rho(x,t)\cdot f(u(\sigma(x,t),q)) = 0, \ \textit{in } \mathbb{R}^m \times [0,T[, \\ \\ \rho(x,0) = \rho_0(x) \ \textit{in } \mathbb{R}^m. \end{cases} \tag{10}$$

The time-varying target $y(t)$ is approachable by the Row player if for every $x \in X$ the Markovian strategy $\sigma(x,t) \in \Delta(A)$ for Row satisfies

$$\langle x-y, u(\sigma(x,t), q(t)) - y\rangle \le 0, \quad \forall\, t \in (0,T]. \tag{11}$$

The proof of this and subsequent results may be found in the appendices. Note that (11) is equivalent to

$$\begin{aligned} \langle x-y, u(\sigma(x,t),q(t)) - y\rangle &= \langle x-y, f(u(\sigma(x,t),q(t)),x) + x - y\rangle \\ &= \langle x-y, f(u(\sigma(x,t),q(t)),x)\rangle + \langle x-y, x-y\rangle \le 0. \end{aligned} \tag{12}$$

The above mirrors Blackwell's Approachability Principle as provided in [8, Corollary 5.1-5.2]. In addition, by taking $\lambda = \frac{x-y}{\|x-y\|} \in \mathbb{R}^m$ and introducing

$$\eta_x(\lambda) := \min_a \langle \frac{x-y}{\|x-y\|}, f(u(a,q(t)),x)\rangle, \forall t \in (0,T], \tag{13}$$

condition (12) can be rewritten as

$$\|x-y\|\eta_x(\lambda) + \langle x-y, x-y\rangle \le 0. \tag{14}$$

Furthermore, note that the value $\lambda = \frac{\partial_x v(x,t)}{\|\partial_x v(x,t)\|}$ used in (9)–(10) is the gradient direction on x. Hence, the term

$$\eta_x(\lambda) := \min_a \langle \frac{\partial_x v(x,t)}{\|\partial_x v(x,t)\|}, f(u(a,q(t)),x)\rangle, \forall t \in (0,T] \tag{15}$$

can be interpreted as the gain floor of the *projected anti-gradient game.*

Remark 1. The value function $v(x,t)$ generalizes the notion of distance from the target. To see this, let us take $v(x,t) = \|x-y\|^2$ and note that the HJB equation in (10) corresponds to the approachability condition (14).

Remarkably, the model (9)-(10) is consistent as $q(t)$ is *self-confirmed.* This means that the mixed strategy $q(t)$ entering the computation of $\eta_x(\lambda)$ reflects the dynamics for $q(t)$ in the expression (6).

Example 1. (Prisoner's Dilemma) Suppose, for instance, that players target the average payoffs across the population. Consider the following game:

	Cooperate	Defect
Cooperate	$(3,3)$	$(0,4)$
Defect	$(4,0)$	$(1,1)$

Figure 1 depicts the payoff space in the continuous-time game based on this Prisoner's Dilemma. Here, the state space is $X = conv\{(3,3),(1,1),(0,4),(4,0)\}$ (the boundary is a solid line), and the target $y = (2,2)$ is the barycenter assuming a uniform distribution. One can visualize the supporting hyperplane H (dot-dashed line) passing through the barycenter, and the vector field $dx(t)$ pointing towards $(\frac{3}{2}, \frac{7}{2})$ for those who cooperate (region below H) and towards $(\frac{5}{2}, \frac{1}{2})$ for those who defect (region above H). The set $conv\{(\frac{3}{2}, \frac{7}{2}), (\frac{5}{2}, \frac{1}{2})\}$ is the set of approachable points with population strategy $q = (\frac{1}{2}, \frac{1}{2})$, and the barycenter is at the equilibrium with uniform distribution over X.

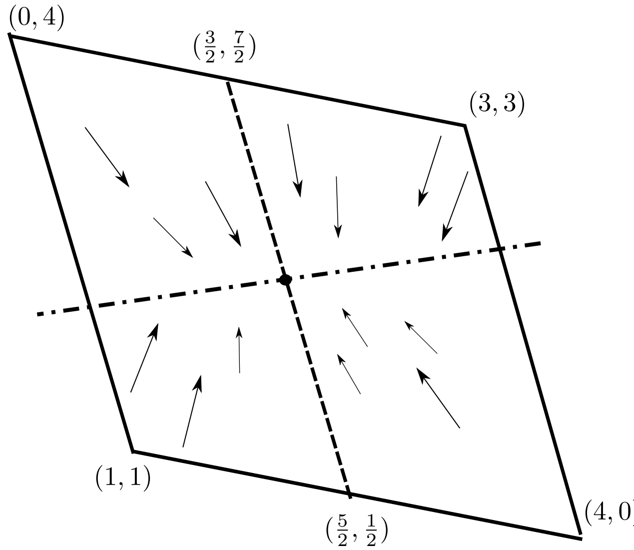

FIGURE 1. Payoff space of Prisoner's dilemma: State space $X = conv\{(3,3),(1,1),(0,4),(4,0)\}$ (boundary a solid line), supporting hyperplane H (dot-dashed line) passing through the barycenter, vector field $dx(t)$ pointing towards $(\frac{3}{2}, \frac{7}{2})$ for those who cooperate (region below H) and towards $(\frac{5}{2}, \frac{1}{2})$ for those who defect (region above H), $conv\{(\frac{3}{2}, \frac{7}{2}), (\frac{5}{2}, \frac{1}{2})\}$ is set of approachable points with population strategy $q = ((\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{2}))$, barycenter is self-confirmed with uniform distribution over X.

3.2. **Existence and nonuniqueness of equilibria.** In this section, we investigate existence and nonuniqueness of equilibria. To do this, we analyze the time-dependence of an estimate error $\nu(t)$, which accounts for the deviation between an estimated density $q(t)$ and the current one $\tilde{q}(t)$ at time t:

$$\nu(t) = q(t) - \tilde{q}(t),$$

where

$$\left\{ \begin{array}{l} \tilde{q}(t) \in \Delta(A)\, s.t.\, \tilde{q}_j(t) = \int_{R_j} \rho(x,t)dx, \\ R_j := \{x \in X \,|\, \sigma(x,t) = j,\, j = \arg\min_a \langle \lambda, f(u(a,q),x)\rangle\} \\ \qquad \forall j \in A,\ \forall t \in (0,T]. \end{array} \right. \tag{16}$$

In other words, R_j is the region where players play the jth action and $\tilde{q}(t)$ is the distribution over actions depending on how many players are in each region R_j for all j. Observe that the time-dependence of $\tilde{q}(t)$ enters in the above through the time-varying nature of the target $y(t)$. Now, according to our procedure, we wish to hypothesize a pair (p,q) that constitutes the input, and obtain a new density $\tilde{q}(p,q)$

as an output. We illustrate the procedure in the special case where $v(x,t) = \|x-y\|^2$, from which we have

$$j = \arg\min_a \langle \lambda, f(u(a,q),x)\rangle = \arg\min_a \langle x-y, u(a,q)-y\rangle.$$

Let us rewrite $y = \sum_{l,k\in A} p_l q_k M_{lk}, \forall p,q \in \Delta(A)$. Then, the expression (16) can be rewritten as

$$\left\{\begin{array}{l} \tilde q(p,q) \in \Delta(A)\, s.t.\, \tilde q_j(p,q) = \int_{R_j} \rho(x,t)dx, \\ R_j := \{x \in X \,|\, \sigma(x,t) = j, \\ j = \arg\min_a \langle x - \sum_{l,k\in A} p_l q_k M_{lk}, u(a,q) - \sum_{l,k\in A} p_l q_k M_{lk}\rangle\} \\ \qquad \forall j \in A,\ \forall t \in (0,T]. \end{array}\right. \tag{17}$$

Eventually, the procedure should return a fixed point. In other words, if we think of an equilibrium as the pair (p^*,q^*) such that $\nu(p^*,q^*) = 0$, existence of an equilibrium is now related to existence of a fixed point for the above procedure, i.e.

$$\tilde q(p_1,q_1) = q.$$

The above means that, given a (p,q) as input to our procedure, the output $\tilde q(p,q)$ coincides with the hypothesized density q. It is natural to represent the above algorithmic procedure as a continuous-time dynamical system and thus to relate convergence to a fixed point to the asymptotic stability of the dynamics. The next assumption introduces conditions for the asymptotic stability to hold.

Assumption 1. There exists a pair $(\dot p, \dot q)$ such that

$$\begin{bmatrix} -\partial_p \tilde q_1 \dot p + \dot q_1 - \partial_q \tilde q_1 \dot q \\ \vdots \\ -\partial_p \tilde q_i \dot p + \dot q_i - \partial_q \tilde q_i \dot q \\ \vdots \\ -\partial_p \tilde q_m \dot p + \dot q_m - \partial_q \tilde q_m \dot q \end{bmatrix} := (-\partial_p \tilde q_i \dot p + \dot q_i - \partial_q \tilde q_i \dot q)_{i=1,\dots,m} \le -\kappa(q - \tilde q). \tag{18}$$

The above describes the possibility of varying (p,q) in order to reduce the estimate error ν, whatever the current error is. The next result establishes the existence of an equilibrium based on the above condition.

Theorem 3.2. (existence) *Let Assumption 1 hold. Then the estimate error decays exponentially fast, i.e.*

$$\nu(t) \le e^{-\kappa t}\nu(0).$$

Essentially, the above theorem shows that if we let the algorithm run for a long time the estimate error asymptotically converges to zero, i.e.

$$\lim_{t\to\infty} \nu = 0,$$

which proves the existence of an equilibrium.

We are now in a position to study nonuniqueness of equilibria. In particular, we provide a variational condition under which the equilibrium is nonunique.

Theorem 3.3. (nonuniqueness) *Starting at an equilibrium where the Lyapunov function* $\mathcal{L} = \frac{1}{2}\nu^T\nu = 0$, *if* $\|\lambda\| = 1$ *for all* $\lambda \in \mathbb{R}^m$, *we have*

$$\begin{array}{c} \min_{\hat p,\hat q} \lambda^T(-\partial_p \tilde q_i \hat p + \hat q_i - \partial_q \tilde q_i \hat q)_{i=1,\dots,m} \\ < 0 < \max_{\hat p,\hat q} \lambda^T(-\partial_p \tilde q_i \hat p + \hat q_i - \partial_q \tilde q_i \hat q)_{i=1,\dots,m}, \end{array} \tag{19}$$

then there exists a $(\dot p, \dot q) = (\hat p, \hat q)$ *such that* $\dot{\mathcal{L}} = 0$ *and thus the current equilibrium is nonunique.*

3.3. **Solution of the mean-field game.** This section investigates the microscopic dynamics of every player given an equilibrium (p, q) and the corresponding target which is the common prior, where the target is denoted by

$$y = \sum_{l,k \in A} p_l q_k M_{lk}.$$

As a result, we obtain that such a dynamics is a "potential" one, in the sense that every player's current average payoff—which we can call the *state* of the player—describes a trajectory along the anti-gradient of a potential function, the latter being the value function of the mean-field game introduced earlier. To this purpose, let us denote by $e(t)$ the deviation between the target y that every player wishes to approach, and the current average payoff $x(t)$, namely

$$e(t) = y - x(t).$$

Given that our running cost is quadratic, from dynamic programming it is natural to assume that the value function also has a quadratic structure. This is a recurrent approach which needs an a posteriori verification of the consistency of the quadratic assumption. In particular, let us assume that the upper bound for the value function takes the form

$$\phi(x,t) = \frac{1}{2} e^T \Phi_t e, \tag{20}$$

where Φ_t is an opportune matrix which is positive definite, i.e. $\Phi_t > 0$. Likewise, consider a quadratic function for the terminal penalty, namely

$$\Psi(x) = \frac{1}{2} e(T)^T \psi e(T).$$

Then, the HJB equation in (22) can be rewritten as

$$\begin{aligned} \partial_t \phi(x,t) + \|\partial_x \phi(x,t)\| l_x[\partial_x \phi(x,t)] + \frac{1}{2} e(t)^T Q e(t) = 0 \text{ in } \mathbb{R}^m \times [0,T[, \\ \phi(x,T) = \Psi(x) \ \forall \ x \in \mathbb{R}^m. \end{aligned} \tag{21}$$

Substituting the expression (20) for the value function in (21) we obtain

$$\begin{aligned} \frac{1}{2} e(t)^T \dot{\Phi}_t e(t) - \frac{1}{2} e(t)^T \Phi_t e(t) + \frac{1}{2} e(t)^T Q e(t) = 0 \text{ in } \mathbb{R}^m \times [0,T[, \\ \frac{1}{2} e(T)^T \Phi_T e(T) = \Psi(x) \ \forall \ x \in \mathbb{R}^m. \end{aligned} \tag{22}$$

The advantage of writing the HJB as above is that all terms are explicitly written as quadratic terms in the error $e(t)$. Considering that the HJB has to hold true for every $e(t)$, we can drop $e(t)$ and thus we have an expression in the only matrix variable Φ_t:

$$\begin{cases} \dot{\Phi}_t - \Phi_t + Q = 0 \text{ in } [0,T[, \\ \Phi_T = \psi \ \forall \ x \in \mathbb{R}^m. \end{cases}$$

The above has the form of a classical differential Riccati equation which can be solved backwardly given the boundary conditions on the matrix in the terminal penalty, $\Phi_T = \psi$.

We can use this result to analyze the microscopic dynamics of each player. Every single player is characterized by the following system of equations involving the evolution of the average payoff (first equation), its best response (second equation), and the expression for the density (third equation):

$$\begin{cases} \dot{x}(t) = \left(\sum_{k\in A} q_k M_{jk} - x(t)\right) dt, \\ \sigma(x,t) = j,\ j = \arg\min_{a\in A} (\Phi_t e(t))^T \left(\sum_{k\in A} q_k M_{ak} - x(t)\right), \\ q \in \Delta(A)\, s.t.\, q_j = \int_{R_j} \rho(x,t)dx, \\ \qquad R_j := \{x \in \mathbb{R}^m \,|\, \sigma(x,t) = j\},\ \forall j \in A. \end{cases} \tag{23}$$

Note that the expression for the best response is obtained from (9), where $\partial_x v$ is now replaced by $\Phi_t e(t)$. This is a straightforward consequence of assuming the value function to be quadratic, as in (20).

4. Application: Regret.

A leading application of games with vector payoffs is in the study of regret-based dynamics, to which we now turn.

4.1. Regret targeting in classical two-player games.

Given a symmetric normal-form game with common action set A and symmetric payoff function $\pi : A \to \mathbb{R}$, let the *regret* of player i from not having played action $k \in A$ under action profile $\alpha \in A^2$ be

$$r(k,\alpha) = \pi(k, \alpha_{-i}) - \pi(\alpha_i, \alpha_{-i}).$$

A straightforward way to justify the vector payoffs introduced earlier is to make them coincide with the regret vector associated to each action profile, i.e.

$$u(\alpha) := \Big(r(k,\alpha)\Big)_{k\in A}.$$

In the following examples, we turn standard games such as the Prisoner's Dilemma, coordination games and Hawk–Dove games into games with regret vectors of type

	Left	Right
Top	$\begin{pmatrix} 0 \\ a \end{pmatrix}$	$\begin{pmatrix} 0 \\ b \end{pmatrix}$
Bottom	$\begin{pmatrix} -a \\ 0 \end{pmatrix}$	$\begin{pmatrix} -b \\ 0 \end{pmatrix}$

and analyse the resulting dynamics of a population targeting expected regret.

Example 2. (Prisoner's Regret) Consider again the Prisoner's Dilemma, and the following bimatrix, which represents the regret vector of the Row player:

	Cooperate	Defect
Cooperate	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$
Defect	$\begin{pmatrix} -1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} -1 \\ 0 \end{pmatrix}$

Putting ourselves in the position of Row, and supposing that the Column player is randomly extracted from the population, we have that if Column is playing D, then if Row switched from D(efect) to C(ooperate), he would lose his payoff of 1, whereas if he stuck to D the regret would be 0. This explains the vector payoff $(-1, 0)$ for the action profile (D, D). Likewise, if Row switched from C to D he would earn a payoff of 1, in comparison with a regret of 0 when sticking to C. This is represented by the regret vector $(0, 1)$ for the action profile (C, D). The reasoning would be analogous if Column were to play C. Note that, at the pure Nash equilibrium (D, D), the regret vector is component-wise nonpositive.

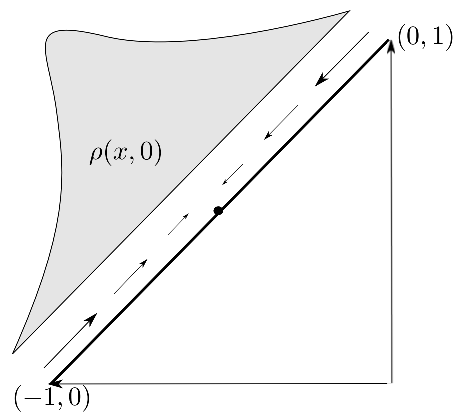

FIGURE 2. Regret space of the Prisoner's dilemma: State space $X = conv\{(-1,0),(0,1)\}$ (solid line), initial distribution $\rho(x,0)$ (grey area), and vector field $dx(t)$ converging to $y = (-0.5, 0.5)$.

Figure 2 depicts the state space $X = conv\{(-1,0),(0,1)\}$ (solid line) in the case with an initial distribution $m(x,0)$ (grey area) of players. The arrows indicate the vector field $dx(t)$ if every player in state $x \in conv\{(-1,0),(-1/2,-1/2)\}$ cooperates, i.e. $a = 1$ and every player in state $x \in conv\{(0,1),(-1/2,-1/2)\}$ defects. The vector field is such that eventually all players converge to the target $y = (-1/2, 1/2)$. Consequently, the distribution converges asymptotically to a Dirac impulse on y.

Example 3. (Coordination game) Consider now the coordination game in the bimatrix on the left, with associated regret-vector game on the right:

	Mozart	Mahler
Mozart	$(2,2)$	$(0,0)$
Mahler	$(0,0)$	$(1,1)$

	Mozart	Mahler
Mozart	$\begin{pmatrix}0\\-2\end{pmatrix}$	$\begin{pmatrix}0\\1\end{pmatrix}$
Mahler	$\begin{pmatrix}2\\0\end{pmatrix}$	$\begin{pmatrix}-1\\0\end{pmatrix}$

In Figure 3, we illustrate the state space $X = conv\{(-1,0),(0,1),(0,-2),(2,0)\}$ (the boundary is a solid line). With a target $y = (0,-1)$, suppose we have a distribution on actions $q = (2/3, 1/3)$, i.e. 2/3 of the population plays Mozart, then $u(Mozart, q) = (0,-1)$ and $u(Mahler, q) = (1, 0)$. The set of approachable points with mixed population strategy $q = (2/3, 1/3)$ is $conv\{(1,0),(0,-1)\}$ (dashed line), i.e. any point in the convex hull of $u(Mozart, q) = (0,-1)$ and $u(Mahler, q) = (1, 0)$. The arrows indicate the vector field $dx(t)$ if every player in state $x \in R_{Mahler} := \{\xi |\, (\xi - y)^T (u(Mahler, q) - y) \leq 0\}$ (grey area) plays Mahler, namely, $a = \sigma(x) = 2$. On the other hand, every player in state $x \in R_{Mozart} := \{\xi |\, (\xi - y)^T (u(Mozart, q) - y) \leq 0\}$ (white area) plays Mozart, i.e. $a = \sigma(x) = 1$. Obviously we need that the integral of the distribution m over R_{Mahler} is consistent with the initial assumption, which means $q_2 = \int_{R_{Mahler}} \rho(x,t)dx = 1/3$. If this occurs, the vector field is such that eventually all players converge to $y = (0,-1)$. Consequently, the distribution converges to a Dirac impulse on y.

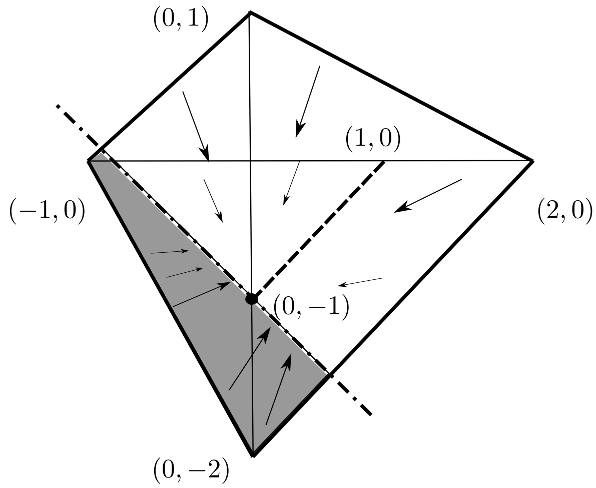

FIGURE 3. Regret space of the coordination game: State space $X = conv\{(-1,0),(0,1),(0,-2),(2,0)\}$ (boundary a solid line), and vector field $dx(t)$ converging to $(1,0)$ (grey area) and $(0,-1)$ (white area), approachable point is $y = (0,-1)$, set of approachable points is $conv\{(1,0),(0,-1)\}$ (dashed line) with mixed population strategy $q = (\frac{2}{3}, \frac{1}{3})$.

Example 4. (Hawk–Dove game) We can likewise transform the Hawk–Dove (or Chicken) game on the left into the corresponding regret-vector game on the right:

	Hawk	Dove
Hawk	$\left(-1,-1\right)$	(4,0)
Dove	(0,4)	$\left(2,2\right)$

	Hawk	Dove
Hawk	$\begin{pmatrix}0\\1\end{pmatrix}$	$\begin{pmatrix}0\\-2\end{pmatrix}$
Dove	$\begin{pmatrix}-1\\0\end{pmatrix}$	$\begin{pmatrix}2\\0\end{pmatrix}$

We have two pure Nash equilibria $(Dove, Hawk)$ and $(Hawk, Dove)$, whose corresponding regret vectors are nonpositive.

More generally, let us now consider the parametric game introduced earlier:

	Left	Right
Top	$\begin{pmatrix}0\\a\end{pmatrix}$	$\begin{pmatrix}0\\b\end{pmatrix}$
Bottom	$\begin{pmatrix}-a\\0\end{pmatrix}$	$\begin{pmatrix}-b\\0\end{pmatrix}$

Figure 4 illustrates the state space $X = conv\{(0,a),(-a,0),(-b,0),(0,b)\}$ (whose boundary is a solid line) where $a < 0 < b$. The target $y = (0,a)$ is in the negative orthant. Here we consider the action distribution $q = (1,0)$, i.e. everybody plays $k = 1$; then $u(Top,q) = (0,a)$ and $u(Bottom,q) = (-a,0)$. The arrows indicate the vector field $dx(t)$, for which eventually all players converge to $y = (0,a)$. Consequently, the distribution converges to a Dirac impulse on y. Note that the supporting hyperplane $H := \{\xi | \, (\xi - y)^T (u(Bottom,q) - y) = 0\}$ (dot-dashed line) intersects X at only one point (the vertex), which is necessary for the vertex to be at the equilibrium by Theorem 3.1.

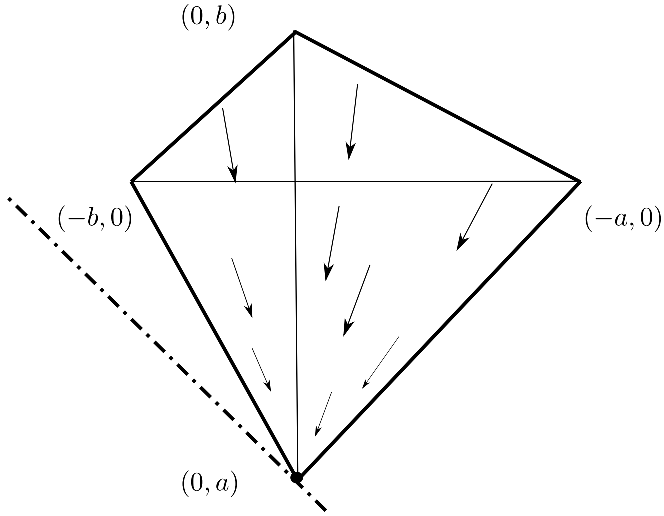

FIGURE 4. Regret space of parametric game with $a < 0 < b$: State space $X = conv\{(0,a),(-a,0),(-b,0),(0,b)\}$ (boundary a solid line), vector field $dx(t)$ converging to $(0,a)$ which is also an approachable vertex with population strategy $q = (1,0)$, supporting hyperplane H (dot-dashed line) intersects X only at one point (the vertex).

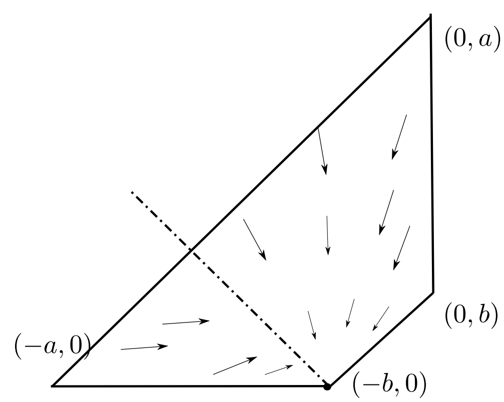

FIGURE 5. Regret space of parametric game with $0 < b < a$: State space $X = conv\{(0,a),(-a,0),(-b,0),(0,b)\}$ (boundary a solid line), supporting hyperplane H (dot-dashed line) passing through the vertex $(-b,0)$, vector field $dx(t)$ converging to $(0,b)$ left of H and to $(-b,0)$ right of H, $conv\{(0,b),(-b,0)\}$ is set of approachable points with population strategy $q = (0,1)$, vertex $(-b,0)$ is not self-confirmed, while vertex $(0,a)$ is self-confirmed with population strategy $q = (1,0)$.

Figure 5 depicts the state space $X = conv\{(0,a),(-a,0),(-b,0),(0,b)\}$ (whose boundary is a solid line) where $0 < b < a$. The target $y = (-b,0)$ is again in the negative orthant. Here we consider the action distribution $q = (0,1)$, i.e. everybody plays $k = 2$; then $u(Top,q) = (0,b)$ and $u(Bottom,q) = (-b,0)$. The arrows indicate the vector field $dx(t)$, for which eventually all players converge to $y = (-b, 0)$. Consequently, the distribution converges to a Dirac impulse on y.[5]

4.2. **Maximum regret and Bayesian equilibrium.** In Hart and Mas-Colell's [20] regret-based dynamics, approachability of the nonpositive orthant implies convergence to Nash equilibrium. This is no longer true for 1st-moment approachability, which drives *expected*—rather than maximum—regret to zero, so that some deviations could still have positive regret. In this section, however, we show how the model can be applied to an incomplete-information setting to yield convergence to Bayesian equilibrium.

Suppose then that the continuous-time population game Γ is based on a game of incomplete information; in particular, we are given a Harsanyi game G (as described in [43]) with state of the world $\omega = (s(\omega); t_1(\omega), t_2(\omega))$ chosen by Nature from a finite set Y using a probability distribution θ. Players then learn their own types $t_i(\omega) \in T_i$, choose actions β_i from a common finite set $B(\omega)$, and receive symmetric payoffs $\varpi_i(\beta; \omega)$, $\beta = (\beta_1, \beta_2)$; the state of nature is $s(\omega) = (B(\omega), \varpi)$, $\varpi = (\varpi_1, \varpi_2)$. Each player i then has a common finite set Σ of (T_i-measurable) pure Bayesian strategies $\sigma_i : Y \to B(\omega)$, which we identify with the action set A in the general framework of Section 2. Given a strategy profile $\sigma \in \Sigma^2$, let the vector payoffs be given by *maximal regrets*,

$$u(\sigma) := \Big(\max_{k \in \Sigma} r(k(\omega), \sigma(\omega)) \Big)_{t_i \in T_i},$$

where the maximum is taken over a player's own strategies, rather than those of his opponent.

Players are continuously rematched against new opponents to play this game G, and a new state of the world is chosen for each such matching. Hence, each play of G is a one-shot game, as distinct from repeated games of incomplete information (see [3] and Ch. 14 of [34]), where the opponents and state remain constant through time. We are thus interested in a population continuously matched to play a one-shot Bayesian game, rather than a dynamic repeated game with learning. Incomplete-information population games have been relatively little studied, with the notable exception of Ely and Sandholm [15, 37], who analyse an evolutionary best-response dynamic with a subpopulation for each possible type. Here, by contrast, we have a single population of rational agents with nonconstant types who adopt one-shot Bayesian strategies through time. Recalling the Google AdWords application from the Introduction, the nonconstant types here correspond to the bidders having an uncertain ranking value in each individual auction, which is quite plausible given their ignorance of the searcher's identity.

1st-moment approachability of the nonpositive orthant in Γ then implies that

$$\mathbb{E}_\theta \max_{k \in \Sigma} \varpi_i(k(\omega), \sigma_{-i}(\omega)) - \varpi_i(\sigma(\omega)) \leq 0.$$

But since the maximum of convex functions is convex, Jensen's inequality implies that the left-hand side is no less than

$$\max_{k \in \Sigma} \mathbb{E}_\theta \varpi_i(k(\omega), \sigma_{-i}(\omega)) - \mathbb{E}_\theta \varpi_i(\sigma(\omega)),$$

[5] However, there is an issue here related to the fact that the vertex y is not at the equilibrium. To see this, note that the supporting hyperplane $H := \{\xi | \, (\xi - y)^T (u(Top, q) - y) = 0\}$ (dot-dashed line) partitions X into two regions, which is necessary for the vertex not to be at the equilibrium by Theorem 3.1.

which is hence also nonpositive. Thus, we have a Nash equilibrium of the Harsanyi game G, which is also a Bayesian equilibrium of the incomplete-information game by Harsanyi's [17] Theorem I.

A simple example is offered by the following game G, where each player's payoffs are randomly determined; with probability 1/2 the Row player $\mathcal{R}$ has the payoffs in the left-hand "l(ow)" matrix, and with probability 1/2 she has the payoffs in the right-hand "h(igh)" matrix:

l	Opera	Football
Opera	3	1
Football	0	2

h	Opera	Football
Opera	1	3
Football	2	0

The Column player $\mathcal{C}$'s payoffs are determined in a symmetric manner. Each player observes her own payoffs, but not those of her opponent. There are thus four possible states of the world $Y = \{\omega_{ll}, \omega_{lh}, \omega_{hl}, \omega_{hh}\}$:

$$\begin{cases} \omega_{ll} = \left(s_{ll}; [\tfrac{1}{2}\omega_{ll}, \tfrac{1}{2}\omega_{lh}], [\tfrac{1}{2}\omega_{ll}, \tfrac{1}{2}\omega_{hl}]\right) \\ \omega_{lh} = \left(s_{lh}; [\tfrac{1}{2}\omega_{ll}, \tfrac{1}{2}\omega_{lh}], [\tfrac{1}{2}\omega_{lh}, \tfrac{1}{2}\omega_{hh}]\right) \\ \omega_{hl} = \left(s_{hl}; [\tfrac{1}{2}\omega_{hl}, \tfrac{1}{2}\omega_{hh}], [\tfrac{1}{2}\omega_{ll}, \tfrac{1}{2}\omega_{hl}]\right) \\ \omega_{hh} = \left(s_{hh}; [\tfrac{1}{2}\omega_{hl}, \tfrac{1}{2}\omega_{hh}], [\tfrac{1}{2}\omega_{lh}, \tfrac{1}{2}\omega_{hh}]\right), \end{cases} \tag{24}$$

each occurring with probability 1/4. Furthermore, there are two possible types of each player,

$$\{R_l, R_h\} = \left\{ \left[\frac{1}{2}\omega_{ll}, \frac{1}{2}\omega_{lh}\right], \left[\frac{1}{2}\omega_{hl}, \frac{1}{2}\omega_{hh}\right] \right\},$$

$$\{C_l, C_h\} = \left\{ \left[\frac{1}{2}\omega_{ll}, \frac{1}{2}\omega_{hl}\right], \left[\frac{1}{2}\omega_{ll}, \frac{1}{2}\omega_{hl}\right] \right\},$$

and each player assigns probability 1/2 to each of her opponent's possible types.

Representing this situation as a Bayesian game, the Row player's vector payoffs are:

	O_l, O_h	O_l, F_h	F_l, O_h	F_l, F_h
O_l, O_h	$\begin{pmatrix}3\\1\end{pmatrix}$	$\begin{pmatrix}2\\2\end{pmatrix}$	$\begin{pmatrix}2\\2\end{pmatrix}$	$\begin{pmatrix}1\\3\end{pmatrix}$
O_l, F_h	$\begin{pmatrix}3\\2\end{pmatrix}$	$\begin{pmatrix}2\\1\end{pmatrix}$	$\begin{pmatrix}2\\1\end{pmatrix}$	$\begin{pmatrix}1\\0\end{pmatrix}$
F_l, O_h	$\begin{pmatrix}0\\1\end{pmatrix}$	$\begin{pmatrix}1\\2\end{pmatrix}$	$\begin{pmatrix}1\\2\end{pmatrix}$	$\begin{pmatrix}2\\3\end{pmatrix}$
F_l, F_h	$\begin{pmatrix}0\\2\end{pmatrix}$	$\begin{pmatrix}1\\1\end{pmatrix}$	$\begin{pmatrix}1\\1\end{pmatrix}$	$\begin{pmatrix}2\\0\end{pmatrix}$

where, for example, O_l, F_h denotes the pure Bayesian strategy $\{\sigma_R(R_l) = \{\text{Opera}\}, \sigma_R(R_h) = \{\text{Football}\}\}$. The Column player's payoffs are symmetric. This game has one pure-strategy equilibrium where Row plays O_l, F_h and Column plays O_l, O_h, and a symmetric one where Row plays O_l, O_h and Column plays O_l, F_h.

Now convert this game into one with maximal-regret payoffs:

	O_l, O_h	O_l, F_h	F_l, O_h	F_l, F_h
O_l, O_h	$\begin{pmatrix}0\\1\end{pmatrix}$	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}1\\0\end{pmatrix}$
O_l, F_h	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}0\\1\end{pmatrix}$	$\begin{pmatrix}0\\1\end{pmatrix}$	$\begin{pmatrix}1\\3\end{pmatrix}$
F_l, O_h	$\begin{pmatrix}3\\1\end{pmatrix}$	$\begin{pmatrix}1\\0\end{pmatrix}$	$\begin{pmatrix}1\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\end{pmatrix}$
F_l, F_h	$\begin{pmatrix}3\\0\end{pmatrix}$	$\begin{pmatrix}1\\1\end{pmatrix}$	$\begin{pmatrix}1\\1\end{pmatrix}$	$\begin{pmatrix}0\\3\end{pmatrix}$

For instance, if Row is playing F_l, O_h and Column is playing O_l, O_h, Row type $\mathcal{R}_l$'s expected payoff is 0, whereas he could have had 3 by playing O_l, O_h, giving a maximal regret of 3; similarly, type $\mathcal{R}_h$'s payoff is 1, whereas he could have had 2 by playing F_l, F_h, giving a maximal regret of 1. 1st-moment approachability of the nonpositive orthant with these maximal-regret payoffs then implies Bayesian equilibrium in G.

From Theorem 3.1 we know that the set of approachable targets is

$$\mathcal{T}(q) = \begin{cases} \{y \mid y \in conv((0,1),(0,0),(3,1),(3,0))\}, & q = (1,0,0,0), \\ \{y \mid y \in conv((0,0),(0,1),(1,0),(1,1))\}, & q = (0,1,0,0), \\ \{y \mid y \in conv((0,0),(0,1),(1,0),(1,1))\}, & q = (0,0,1,0), \\ \{y \mid y \in conv((1,0),(1,3),(0,0),(0,3))\}, & q = (0,0,0,1). \end{cases} \tag{25}$$

This means that, for any pure strategy q, the origin $(0,0)$—and hence Bayesian equilibrium—is reachable and in particular the corresponding strategy is

$$\sigma(x) = \begin{cases} a_i = O_l, F_h & \text{for all } x, \quad q = (1,0,0,0), \\ a_i = O_l, O_h & \text{for all } x, \quad q = (0,1,0,0), \\ a_i = O_l, O_h & \text{for all } x, \quad q = (0,0,1,0), \\ a_i = F_l, O_h & \text{for all } x, \quad q = (0,0,0,1). \end{cases} \tag{26}$$

However, note that none of the above strategies corresponds to a self-confirmed equilibrium according to Theorem 3.1. Indeed, taking the first strategy say, $a_i = O_l, F_h$ for all x when $q = (1,0,0,0)$. But $a_i = O_l, F_h$ for all x implies $R_{O_l,F_h} = X$ and $R_{O_l,O_h} = R_{F_l,O_h} = R_{F_l,F_h} = \emptyset$ which implies in turn $q = (0,1,0,0)$ and this contradicts the assumption $q = (1,0,0,0)$. We can repeat the same reasoning for any other strategy.

5. **Conclusion.** This paper has combined approachability theory, evolutionary games, and mean-field games in a unified framework. The game studied has vector payoffs, a large number of players, and admits a classical mean-field game representation involving two coupled PDEs, the *Hamilton-Jacobi-Bellman equation* and the *advection equation*. We have highlighted multiple contributions. First, we coin the notion of *1st-moment approachability* and analyze the corresponding convergence conditions. Second, we use the mean-field game to introduce *self-confirmed equilibrium*. Third, we discuss existence, nonuniqueness and stability of equilibria as fixed points of the two PDEs.

Future work will involve the stochastic analysis of the same game in the presence of an additional Brownian motion in the dynamics. This would capture uncertainty or model misspecification. In a different direction, we are interested in extending the study to the case where each player can adopt a mixed strategy, which would imply a new definition of density distribution on the space of mixed strategies; so far, the density distribution is defined on the space of pure strategies. A third development will be a further analysis of the connections with the Bayesian approach.

Appendices. Proof of Theorem 3.1.

For the first part, a mean-field game appears in the form of two coupled PDEs intertwined in a forward-backward way as follows:

$$\begin{cases} \partial_t v(x,t) + \inf_a \{\langle \partial_x v(x,t), f(u(a,q),x)\rangle + g(x,y)\} = 0 \text{ in } \mathbb{R}^m \times [0,T[, \\ \\ v(x,T) = \Psi(x,y) \ \forall\, x \in \mathbb{R}^m, \\ \\ \partial_t \rho(x,t) + div(\rho(x,t)\cdot f(u(\sigma(x,t),q),x)) = 0, \\ \\ \rho(0) = \rho_0, \end{cases} \tag{27}$$

where $\sigma(x,t)$ and $q(t)$ are the optimal time-varying state-feedback controls of Row and Column players, respectively, obtained as

$$\begin{cases} \sigma(x,t) \in \arg\min_{a\in A}\{\langle \partial_x v(x,t), f(u(a,q),x)\rangle + g(x,y)\}, \\ \\ q(t) \in \Delta(A)\; s.t.\; q_k(t) = \int_{R_k} \rho(x,t)dx, \\ \qquad R_k := \{x \in \mathbb{R}^m \,|\, \sigma(x) = k\},\ \forall k \in A,\ \forall t \in (0,T]. \end{cases} \tag{28}$$

The first equation in (27) is the *Hamilton-Jacobi-Bellman* (HJB) equation with variable $v(x,t)$ and parametrized in $\rho(\cdot)$. Given the boundary condition on final state (second equation in (27)), and assuming a given population behavior captured by $\rho(\cdot)$, the HJB equation is solved backwards and returns the value function and best-response behavior of the individuals (first equation in (28)) as well as the worst adversarial response (second equation in (28)). The HJB equation is coupled with a second PDE, known as the *advection or transport equation* (third equation in (27)), defined on variable $\rho(\cdot)$ and parametrized in $v(x,t)$. Given the boundary condition on initial distribution $\rho(0) = \rho_0$ (fourth equation in (27)), and assuming a given individual behavior described by u^*, the FPK equation is solved forward and returns the population behavior time evolution $\rho(t)$.

Using (15) we have

$$\inf_a \{\langle \partial_x v(x,t), f(u(a,q),x)\rangle + g(x,y)\} = \|\partial_x v\|\eta_x(\lambda) + \frac{1}{2}\langle x-y, Q(x-y)\rangle$$

We also have

$$\begin{aligned} \sigma(x,t) &\in \arg\min_{a\in A}\{\langle \partial_x v(x,t), f(u(a,q),x)\rangle + g(x,y)\} \\ &= \arg\min_{a\in A}\{\langle \tfrac{\partial_x v(x,t)}{\|\partial_x v(x,t)\|}, f(u(a,q),x)\rangle + g(x,y)\} \\ &= \arg\min_{a\in A}\{\langle \lambda, f(u(a,q),x)\rangle + g(x,y)\}. \end{aligned} \tag{29}$$

It is left to observe that for the vector field we have $f(u(\sigma(x,t),q))$ and this proves the third equation (FPK equation).

For the second part, let us rewrite as $y = \sum_{l,k\in A} p_l q_k M_{lk}$ where $p,q \in \Delta(A)$. Then for every $x \in X$, there exists $\sigma(x) = j$, where

$$j = \arg\min_{a\in A}(x-y)^T(u(a,q)-y)$$

such that (11) holds for all $q \in \Delta(A)$, namely:

$$\langle x-y, u(\sigma(x,t),q)-y\rangle = \langle x-y, \sum_{k\in A} q_k M_{jk} - \sum_{l,k\in A} p_l q_k M_{lk}\rangle \le 0.$$

Now, let $x(t)$, $t \in [0,T]$ be solution of (1). Also, let $\delta(t) = \|x(t)-y\|^2$. We have

$$\dot{\delta}(t) = 2\langle f(p(u,w),x), x(t)-y\rangle = 2\langle u(a,q)-x, x(t)-y\rangle \le -2\delta(t).$$

As a consequence one obtains that

$$\|x(t)-y\| \le \|x(0)-y\|e^{-t}$$

and therefore y is approachable by a population with exponential rate. **Q.E.D.**

Proof of Theorem 3.2. This proof is based on a Lyapunov stability approach. In particular, let us introduce a quadratic (in the error) Lyapunov function

$$\mathcal{L} = \frac{1}{2}\nu^T\nu,$$

and show that its derivative is strictly negative. The time derivative can be decomposed as the sum of two terms involving the gradient of $\mathcal{L}$ with respect to the two variables p and q. More specifically,

$$\begin{aligned}\dot{\mathcal{L}} &= (\partial_p\mathcal{L})^T\dot{p} + (\partial_q\mathcal{L})^T\dot{q}\\ &= \nu^T\dot{\nu} = (q-\tilde{q})^T\Big[\big((\partial_p\nu_i)^T\dot{p}\big)_{i=1,\dots,m} + \big((\partial_q\nu_i)^T\dot{q}\big)_{i=1,\dots,m}\Big]\\ &= (q-\tilde{q})^T(-\partial_p\tilde{q}_i\dot{p} + \dot{q}_i - \partial_q\tilde{q}_i\dot{q})_{i=1,\dots,m}.\end{aligned} \tag{30}$$

From condition (18), we also have

$$\dot{\mathcal{L}} \le -\kappa(q-\tilde{q})^T(q-\tilde{q}) = -\kappa\nu^T\nu,$$

which proves the thesis. **Q.E.D.**

Proof of Theorem 3.3. There exists a $(\dot{p},\dot{q})$ such that

$$\tilde{q}(p+\dot{p}dt, q+\dot{q}dt) = q+\dot{q}dt.$$

The above also means that the error

$$\nu = \tilde{q}(p+\dot{p}dt, q+\dot{q}dt) - (q+\dot{q}dt) = 0.$$

Q.E.D.

REFERENCES

[1] R. J. Aumann. Utility theory without the completeness axiom. *Econometrica*, 30:445–462, 1962.
[2] R. J. Aumann. Markets with a continuum of traders. *Econometrica*, 32(1-2):39–50, 1964.
[3] Robert J. Aumann and Michael B. Maschler. *Repeated Games with Incomplete Information*. MIT Press, 1995.
[4] F. Bagagiolo and D. Bauso. Objective function design for robust optimality of linear control under state-constraints and uncertainty. *ESAIM: Control, Optimisation and Calculus of Variations*, 17:155–177, 2011.
[5] M. Bardi. Explicit solutions of some linear-quadratic mean field games. *Network and Heterogeneous Media*, 7:243–261, 2012.
[6] J. Battelle. *The Search*. Nicholas Brealey Publishing, 2006.
[7] D. Bauso, E. Lehrer, E. Solan, and X. Venel. Attainability in repeated games with vector payoffs. *INFORMS Mathematics of Operations Research*, 40(3):739–755, 2015.
[8] M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. *SIAM Journal on Control and Optimization*, 44(1):338–348, October 2005.

[9] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific J. Math.*, 6(1):1–8, 1956.
[10] F. Blanchini. Set invariance in control a survey. *Automatica*, 35(11):1747–1768, 1999.
[11] L. E. Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica*, 59(1):61–79, 1991.
[12] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning and Games*. Cambridge University Press, 2006.
[13] B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalised second-price auction: selling billions of dollars worth of keywords. *American Economic Review*, 97:242–259, 2007.
[14] N. J. Elliot and N.J. Kalton. The existence of value in differential games of pursuit and evasion. *J. Differential Equations*, 12:504–523, 1972.
[15] Jeffrey C. Ely and William H. Sandholm. Evolution in bayesian games i: Theory. *Games and Economic Behavior*, 53:83–109, 2005.
[16] D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 29:7–35, 1999.
[17] John C. Harsanyi. Games with incomplete information played by 'bayesian' players, i–iii. part ii. bayesian equilibrium points. *Management Science*, 14:320–334, 1968.
[18] S. Hart. Adaptive heuristics. *Econometrica*, 73:1401–1430, 2005.
[19] S. Hart and A. Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98:2654, 2001.
[20] S. Hart and A. Mas-Colell. Regret-based continuous-time dynamics. *Games and Economic Behavior*, 45:375–394, 2003.
[21] M.Y. Huang, P.E. Caines, and R.P. Malhamé. Large population stochastic dynamic games: Closed loop kean-vlasov systems and the nash certainty equivalence principle. *Communications in Information and Systems*, 6:221–252, 2006.
[22] M.Y. Huang, P.E. Caines, and R.P. Malhamé. Large population cost-coupled lqg problems with non-uniform agents: individual-mass behaviour and decentralized ϵ-nash equilibria. *IEEE Trans. on Automatic Control*, 9:1560–1571, 2007.
[23] M.Y. Huang, P.E. Caines, and R.P. Malhamé. Individual and mass behaviour in large population stochastic wireless power control problems: Centralized and nash equilibrium solutions. In *Proc. of the IEEE Conference on Decision and Control*, volume 42, pages 98–103, HI, USA, December 2003.
[24] B. Jovanovic and R. W. Rosenthal. Anonymous sequential games. *Journal of Mathematical Economics*, 17:77–87, 1988.
[25] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. i le cas stationnaire. *Comptes Rendus Mathematique*, 343(9):619–625, 2006.
[26] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. ii horizon fini et controle optimal. *Comptes Rendus Mathematique*, 343(10):679–684, 2006.
[27] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese journal of Mathematics*, 2:229–260, 2007.
[28] E. Lehrer. Allocation processes in cooperative games. *International Journal of game Theory*, 31:341–351, 2002.
[29] E. Lehrer. Approachability in infinite dimensional spaces. *International Journal of game Theory*, 31(2):253–268, 2002.
[30] E. Lehrer. A wide range no-regret theorem. *Games and Economic Behavior*, 42, 2003.
[31] E. Lehrer and E. Solan. Excludability and bounded computational capacity strategies. *Mathematics of Operations Research*, 31(3):637–648, 2006.
[32] E. Lehrer, E. Solan, and D. Bauso. Repeated games over networks with vector payoffs: the notion of attainability. In *Proceedings of the NetGCoop 2011*, Paris, France, October 2011.
[33] E. Lehrer and S. Sorin. Minmax via differential inclusion. *Convex Analysis*, 14(2):271–273, 2007.
[34] Michael Maschler, Eilon Solan, and Shmuel Zamir. *Game Theory*. Cambridge University Press, 2013.
[35] E. Roxin. The axiomatic approach in differential games. *J. Optim. Theory Appl.*, 3:153–163, 1969.
[36] W.H. Sandholm. *Population Games and Evolutionary Dynamics*. MIT Press, 2010.
[37] William H. Sandholm. Evolution in bayesian games ii: Stability of purified equilibrium. *Journal of Economic Theory*, 136:641–667, 2007.

[38] A. S. Soulaimani, M. Quincampoix, and S. Sorin. Approachability theory, discriminating domain and differential games. *SIAM Journal of Control and Optimization*, 48(4):2461–2479, 2009.
[39] P. Varaiya. The existence of solution to a differential game. *SIAM Journal of Control and Optimization*, 5:153–162, 1967.
[40] H.R. Varian. Position auctions. *International Journal of Industrial Organization*, 25:1163–1178, 2007.
[41] N. Vieille. Weak approachability. *Mathematics of Operations Research*, 17:781–791, 1992.
[42] John von Neumann. Zur theorie der gesellschaftsspiele. *Math. Annalen*, 100:295–320, 1928.
[43] S. Zamir. Bayesian games: Games with incomplete information. *Computational Complexity*, 2012.

E-mail address: d.bauso@rug.nl
E-mail address: thomas.norman@magd.ox.ac.uk