1. The online decision problem
2. **Blackwell’s Approachability**
3. The Multiplicative Weight Algorithm
4. Online convex programming
5. Recent topics
The theory of set-approachability in repeated games with vector payoffs was introduced by David Blackwell in 1956. It has since played an important role in the general theory of repeated games, including repeated games with incomplete information, the "folk theorem", Nash dynamics, and online learning.
The theory of set-approachability in repeated games with vector payoffs was introduced by David Blackwell in 1956. It has since played an important role in the general theory of repeated games, including repeated games with incomplete information, the “folk theorem”, Nash dynamics, and online learning.

Following Hannan's paper, Blackwell (1954) observed that no-regret strategies can be elegantly derived using his general approachability framework. We review here these basic results, and briefly mention some more recent developments.
Consider again a two-person repeated-game model between PL1 (the agent) and PL2 (the opponent/environment/Nature), with finite action sets $\mathcal{I}$ and $\mathcal{J}$, actions $i, j$, mixed actions $p, q$, and strategies $\pi^1, \pi^2$ (resp.).

Here, however, the payoffs at each stage are vectors in Euclidean space:

$$v(i, j) \in \mathbb{R}^m$$
The game model

- Consider again a two-person repeated-game model between PL1 (the agent) and PL2 (the opponent/environment/Nature), with finite action sets $\mathcal{I}$ and $\mathcal{J}$, actions $i, j$, mixed actions $p, q$, and strategies $\pi^1, \pi^2$ (resp.).

- Here, however, the payoffs at each stage are vectors in Euclidean space:
  \[ v(i, j) \in \mathbb{R}^m \]

- Recall the notations

  \[ v_t = v(i_t, j_t) \]
  \[ \bar{v}_T = \frac{1}{T} \sum_{t=1}^{T} v(i_t, j_t) \]
  \[ v(p, q) = \sum_{i, j} p_i v(i, j) q_j \]
Approachable Sets

- Let $B$ be a closed set in $\mathbb{R}^m$, which represents a desired region for PL1’s average payoff vector.

**Definition**

A closed set $B$ is *approachable* if there exists a strategy $\pi^1$ for PL1 such that

$$\lim_{T \to \infty} d(\bar{v}_T, B) = 0, \quad \mathbb{P}^{\sigma, \tau}\text{-a.s.}$$

holds for any strategy $\pi^2$ of PL2 (where $d$ is the Euclidean point-to-set distance).

We further require this convergence to be uniform over $\pi^2$, namely

$$\forall \epsilon > 0 \ \exists T \ \text{s.t.} \ \sup_{\sigma} \mathbb{P}\left\{ \sup_{t \geq T} d(\bar{v}_t, B) > \epsilon \right\} \leq \epsilon$$
Sufficient condition

For \( x \notin B \), let \( c_B(x) \) denote a closest point in \( B \) to \( x \), and let \( H(x) \) be the hyperplane through \( c_B(x) \) which is perpendicular to the vector \( x - c_B(x) \).

Theorem (Blackwell)

Suppose that for every \( x \notin B \), there exists a mixed action \( p = p^*(x) \) of PL1 such that the hyperplane \( H(x) \) separates \( x \) from \( \{v(p, q) : q \in \Delta(J)\} \). Then

1. \( B \) is an approachable set.
2. An approaching strategy for PL1 is given by:
   - choose the mixed action \( p_t = p^*(\bar{v}_{t-1}) \) whenever \( \bar{v}_{t-1} \notin B \)
   - play arbitrarily otherwise.

N. Shimkin, Technion
Sufficient condition

For $x \notin B$, let $c_B(x)$ denote a closest point in $B$ to $x$, and let $H(x)$ be the hyperplane through $c_B(x)$ which is perpendicular to the vector $x - c_B(x)$.

**Theorem (Blackwell)**

*Suppose that for every $x \notin B$, there exists a mixed action $p = p^*(x)$ of PL1 such that the hyperplane $H(x)$ separates $x$ from $\{v(p, q) : q \in \Delta(J)\}$. Then*

1. $B$ is an approachable set.

2. An approaching strategy for PL1 is given by:
   - choose the mixed action $p_t = p^*(v_{t-1})$ whenever $v_{t-1} \notin B$
   - play arbitrarily otherwise.

The mixed action $p^*(x)$ can be computed as PL1’s maximin action in the one-shot game with costs projected unto $x - c_B(x)$, namely,

$$p^*(x) = \arg\max_p \min_q \{v(p, q) \cdot (x - c_B(x))\}$$
Proof.

The main step is to recursively bound $\mathbb{E}(d(\bar{v}_t, B)^2)$.

Denote

$$c_t = c_B(\bar{v}_t), \quad d_t = d(\bar{v}_t, B) = ||\bar{v}_t - c_t||$$

Observe that

$$d_{t+1} = ||\bar{v}_{t+1} - c_{t+1}|| \leq ||\bar{v}_{t+1} - c_t||$$

$$(t + 1)d_{t+1} \leq ||t\bar{v}_t + v_{t+1} - (t + 1)c_t||$$

$$= ||t(\bar{v}_t - c_t) + (v_{t+1} - c_t)||$$
Proof.

The main step is to recursively bound $\mathbb{E}(d(\bar{v}_t, B)^2)$. Denote

$$c_t = c_B(\bar{v}_t), \quad d_t = d(\bar{v}_t, B) = ||\bar{v}_t - c_t||$$

Observe that

$$d_{t+1} = ||\bar{v}_{t+1} - c_{t+1}|| \leq ||\bar{v}_{t+1} - c_t||$$

$$(t + 1)d_{t+1} \leq ||t\bar{v}_t + v_{t+1} - (t + 1)c_t||$$

$$\quad = ||t(\bar{v}_t - c_t) + (v_{t+1} - c_t)||$$

But, according to our policy, whenever $d_t > 0$,

$$\mathbb{E} \left( \langle \bar{v}_t - c_t, v_{t+1} - c_t \rangle | \mathcal{F}_t \right) \leq 0$$
Proof.

The main step is to recursively bound $\mathbb{E}(d(\bar{v}_t, B)^2)$.

Denote

$$c_t = c_B(\bar{v}_t), \quad d_t = d(\bar{v}_t, B) = ||\bar{v}_t - c_t||$$

Observe that

$$d_{t+1} = ||\bar{v}_{t+1} - c_{t+1}|| \leq ||\bar{v}_{t+1} - c_t||$$

$$(t + 1)d_{t+1} \leq ||t\bar{v}_t + v_{t+1} - (t + 1)c_t||$$

$$= ||t(\bar{v}_t - c_t) + (v_{t+1} - c_t)||$$

But, according to our policy, whenever $d_t > 0$,

$$\mathbb{E}(\langle \bar{v}_t - c_t, v_{t+1} - c_t \rangle | \mathcal{F}_t) \leq 0$$

Therefore

$$(t + 1)^2 \mathbb{E}(d_{t+1}^2 | \mathcal{F}_t) \leq t^2 d_t^2 + \mathbb{E}(||v_{t+1} - c_t|| | \mathcal{F}_t) \leq t^2 d_t^2 + C_0$$

for some constant $C_0$. The same hold trivially when $d_t = 0$. Iterating, we obtain

$$t^2 \mathbb{E}(d_t^2) \leq C_0 t, \quad \text{or} \quad \mathbb{E}(d_t^2) \leq \frac{C_0}{t}$$

Uniform a.s.-convergence easily follows (as $d_t$ is bounded).
Convex target sets

Theorem

Suppose the set $B$ is closed and convex. Then

1. The condition of the previous theorem is both sufficient and necessary.
2. If $B$ is not approachable, there exists a mixed action $q$ of the opponent so that $v(p, q) \not\in B$ for all $p$.

The last result follows from the minimax theorem, and may be viewed as a generalization of that theorem to vector-valued games. This result also implies the following condition, which is often easy to check.

Corollary (Dual condition) $B$ is approachable if (and only if), for every $q \in \Delta(J)$ there exists $p \in \Delta(I)$ so that $v(p, q) \in B$. 

N. Shimkin, Technion
Convex target sets

**Theorem**

Suppose the set $B$ is closed and **convex**. Then

1. The condition of the previous theorem is both **sufficient and necessary**.
2. If $B$ is not approachable, there exists a mixed action $q$ of the opponent so that $v(p, q) \not\in B$ for all $p$.

The last result follows from the minimax theorem, and may be viewed as a generalization of that theorem to vector-valued games.
Convex target sets

Theorem

Suppose the set $B$ is closed and convex. Then

1. The condition of the previous theorem is both sufficient and necessary.
2. If $B$ is not approachable, there exists a mixed action $q$ of the opponent so that $v(p, q) \not\in B$ for all $p$.

- The last result follows from the minimax theorem, and may be viewed as a generalization of that theorem to vector-valued games.
- This result also implies the following condition, which is often easy to check.

Corollary (Dual condition)

$B$ is approachable if (and only if), for every $q \in \Delta(J)$ there exists $p \in \Delta(I)$ so that $v(p, q) \in B$. 

N. Shimkin, Technion
Recall that a no-regret algorithm must secure
\[ \bar{c}_T \leq c^*(\bar{q}_T) + o(1), \quad \text{where} \quad c^*(q) = \min_p c(p, q) \]

Define the following set:
\[ B_1 = \{(c, q) \in \mathbb{R} \times \Delta(J) : c \leq c^*(q)\} \]

The no-regret requirement is now equivalent to the convergence
\[ d((\bar{c}_T, \bar{q}), B_1) \to 0 \quad (a.s) \]
Recall that a no-regret algorithm must secure
\[ \bar{c}_T \leq c^*(\bar{q}_T) + o(1), \quad \text{where} \quad c^*(q) = \min_p c(p, q) \]

Define the following set:
\[ B_1 = \{ (c, q) \in \mathbb{R} \times \Delta(J) : c \leq c^*(q) \} \]

The no-regret requirement is now equivalent to the convergence
\[ d((\bar{c}_T, \bar{q}), B_1) \to 0 \quad (a.s) \]

We proceed to devise an auxiliary game with vector payoff
\[ v(i, j) = (c(i, j), \delta_j), \]

so that \( \bar{v}_T = (\bar{c}_T, \bar{q}_T) \). It remains only to show that the set \( B_1 \) is approachable in this game.
Observe first that \( B_1 \) is a convex set. Recalling that
\[
B_1 = \{ c, q : c \leq c^*(q) \},
\]
convexity follows immediately from concavity of the Bayes risk
\[
c^*(q) = \min_p c(p, q).
\]

We next verify the dual condition from the last Corollary:
\[
\forall q \exists p \text{ s.t. } v(p, q) \in B_1
\]
Noting that \( v(x, y) = (c(x, y), y) \), this
requires \( c(p, q) \leq c^*(q) \). But this is satisfied (with equality) by choosing
\[
p \in \text{BR}(q) = \arg\min_p c(p, q)
\]
Observe first that $B_1$ is a convex set. Recalling that
$B_1 = \{c, q : c \leq c^*(q)\}$, convexity follows immediately from concavity
of the Bayes risk $c^*(q) = \min_p c(p, q)$.

We next verify the dual condition from the last Corollary:

$$\forall q \exists p \text{ s.t. } v(p, q) \in B_1$$

Noting that $v(x, y) = (c(x, y), y)$, this
requires $c(p, q) \leq c^*(q)$. But this is satisfied (with equality) by choosing
$p \in BR(q) = \arg\min_p c(p, q)$

We note that the actual algorithm is not as simple: At each step, the
agent needs to compute the projection of $\bar{v}_T$ onto $B$, and solve the
zero-sum matrix game with payoff projected onto this direction.
A different formulation that leads to a more explicit algorithm was suggested by Hart and Mas-Collel (2001).

We start again from

$$\bar{c}_T \leq c^*(\bar{q}_T) + o(1), \quad \text{where} \quad c^*(q) = \min_p c(p, q)$$

Clearly this can be written as

$$\bar{c}_T \leq c(i', \bar{q}_T) + o(1), \quad \text{for all} \quad i' \in I$$
A different formulation that leads to a more explicit algorithm was suggested by Hart and Mas-Colell (2001).

We start again from

\[ \bar{c}_T \leq c^*(\bar{q}_T) + o(1), \text{ where } c^*(q) = \min_p c(p, q) \]

Clearly this can be written as

\[ \bar{c}_T \leq c(i', \bar{q}_T) + o(1), \text{ for all } i' \in \mathcal{I} \]

Define the following payoff vector:

\[ v(i, j) = (c(i, j) - c(i', j))_{i' \in \mathcal{I}} \in \mathbb{R}^\mathcal{I} \]

so that

\[ \bar{v}_T = (\bar{c}_T - c(i', \bar{q}_T))_{i' \in \mathcal{I}} \]
A different formulation that leads to a more explicit algorithm was suggested by Hart and Mas-Collel (2001).

We start again from

$$\bar{c}_T \leq c^*(\bar{q}_T) + o(1), \quad \text{where} \quad c^*(q) = \min_p c(p, q)$$

Clearly this can be written as

$$\bar{c}_T \leq c(i', \bar{q}_T) + o(1), \quad \text{for all} \quad i' \in \mathcal{I}$$

Define the following payoff vector:

$$v(i, j) = (c(i, j) - c(i', j))_{i' \in \mathcal{I}} \in \mathbb{R}^\mathcal{I}$$

so that

$$\bar{v}_T = (\bar{c}_T - c(i', \bar{q}_T))_{i' \in \mathcal{I}}$$

No-regret is now equivalent to approachability of the set $B_0 = \mathbb{R}_{-\infty}^\mathcal{I}$ (the negative orthant) in the repeated game with payoff $v$. 
Verifying approachability

- Clearly $B_0$ is convex, and it is again easy to verify the dual condition of the corollary by choosing $p \in BR(q)$.
- Moreover, here we can explicitly find an approaching (hence no-regret) strategy via Blackwell’s primal (separation) condition.
Clearly $B_0$ is convex, and it is again easy to verify the dual condition of the corollary by choosing $p \in BR(q)$.

Moreover, here we can explicitly find an approaching (hence no-regret) strategy via Blackwell’s primal (separation) condition.

Given a point $x \equiv \bar{v}_t \notin B_0$, it is easy to see that

$$x - c_B(x) = x_+ \equiv (x_i)_+ \, i \in I$$

Therefore, the relevant separation condition requires to find $p$ so that, for every $q$,

$$v(p, q) \cdot x_+ \leq 0$$

or

$$c(p, q) - c\left(\frac{x_+}{\|x_+\|_1}, q\right) \leq 0$$
Verifying approachability

- Clearly $B_0$ is convex, and it is again easy to verify the dual condition of the corollary by choosing $p \in BR(q)$.

- Moreover, here we can explicitly find an approaching (hence no-regret) strategy via Blackwell’s primal (separation) condition.

- Given a point $x \equiv \bar{v}_t \notin B_0$, it is easy to see that

$$x - c_B(x) = x_+ \triangleq ((x_i)_+)_{i \in I}$$

Therefore, the relevant separation condition requires to find $p$ so that, for every $q$,

$$v(p, q) \cdot x_+ \leq 0$$

or

$$c(p, q) - c\left(\frac{x_+}{||x_+||_1}, q\right) \leq 0$$

- Clearly, choosing $p = x_+/||x_+||_1$ gives 0 on the LHS.
An explicit no-regret strategy

Substituting $\bar{v}_t = \bar{c}_t - c(\cdot, \bar{q}_t)$ for $x$, we obtain the explicit no-regret strategy:

$$(p_{t+1})_i = \frac{(\bar{c}_t - c(i, \bar{q}_t))_+}{\|(\bar{c}_t - c(\cdot, \bar{q}_t))_+\|_1}$$

That is: the play probability of action $i$ is proportional to the positive regret relative to that action.
Some Generalizations
Hart and Mas-Colell (2001) have generalized Blackwell’s approaching strategy to a whole class of strategies, based on a potential function \( P(x) \) whose gradient provides a projection direction towards the convex set \( B \) from any point \( x \notin B \).

Blackwell’s strategy (based on Euclidean projection) is obtained as a special case for \( P(x) = \min_{y \in B} ||y - x||_2 \), the Euclid distance to \( B \).
Regret-based strategies

- By using the negative-quadrant approachability formulation, a family of no-regret algorithms is derived, which are based on the action-regrets

\[ \bar{R}_t(i) = \bar{c}_t - c(i, \bar{q}_t) \]

For \( p = 2 \) we obtain the previous strategy. However, \( p = \infty \) yields the FTL (or fictitious play) strategy, which is deterministic and does not secure no-regret.
Regret-based strategies

- By using the negative-quadrant approachability formulation, a family of no-regret algorithms is derived, which are based on the action-regrets

\[ \bar{R}_t(i) = c_t - c(i, \bar{q}_t), \quad i \in I \]

- In particular, using the \( \ell_p \) potential (with \( 1 < p < \infty \)) leads to no-regret strategies of the form

\[ (p_{t+1})_i \propto (\bar{R}_t(i)_+)^{p-1} \]
Regret-based strategies

- By using the negative-quadrant approachability formulation, a family of no-regret algorithms is derived, which are based on the action-regrets
  \[ \bar{R}_t(i) = \bar{c}_t - c(i, \bar{q}_t), \quad i \in \mathcal{I} \]

- In particular, using the \( \ell_p \) potential (with \( 1 < p < \infty \)) leads to no-regret strategies of the form
  \[ (p_{t+1})_i \propto (\bar{R}_t(i)_+)^{p-1} \]

- For \( p = 2 \) we obtain the previous strategy. However, \( p = \infty \) yields the FTL (or fictitious play) strategy, which is deterministic and does not secure no-regret.
Internal regret

- Internal regret provides a refined sense of regret, which was found useful in the context of equilibrium dynamics.
- Here one considers possible regret for not having changed action $k$ to another action $\ell$, for all pairs of actions $k, \ell \in \mathcal{I}$.  

Formally, $R_T(k, \ell) = \sum_{t=1}^{T} (c(i_t, j_t) - c(\ell, j_t))^1\{i_t = k\}$

Approachability readily implies existence of strategies that incur no internal regret, in the sense that $R_T(k, \ell) \leq o(T)$ (a.s.) for all action pairs.
Internal regret

- Internal regret provides a refined sense of regret, which was found useful in the context of equilibrium dynamics.
- Here one considers possible regret for not having changed action $k$ to another action $\ell$, for all pairs of actions $k, \ell \in \mathcal{I}$.
- Formally,

$$R_T(k, \ell) = \sum_{t=1}^{T} (c(i_t, j_t) - c(\ell, j_t)) 1\{i_t=k\}$$
Internal regret

- Internal regret provides a refined sense of regret, which was found useful in the context of equilibrium dynamics.
- Here one considers possible regret for not having changed action $k$ to another action $\ell$, for all pairs of actions $k, \ell \in \mathcal{I}$.
- Formally,

$$ R_T(k, \ell) = \sum_{t=1}^{T} (c(i_t, j_t) - c(\ell, j_t))1\{i_t=k\} $$

- Approachability readily implies existence of strategies that incur no internal regret, in the sense that $R_T(k, \ell) \leq o(T)$ (a.s.) for all action pairs.
Internal regret

We can verify that using Blackwell’s dual condition. Consider the payoff vector \( (v_{k\ell}) \) with components:

\[
v_{k\ell}(i, j) = (c(k, j) - c(\ell, j))1_{\{i=k\}}
\]

so that

\[
v_{k\ell}(i, q) = (c(k, q) - c(\ell, q))1_{\{i=k\}}
\]
We can verify that using Blackwell’s dual condition. Consider the payoff vector \((v_{k\ell})\) with components:

\[
v_{k\ell}(i, j) = (c(k, j) - c(\ell, j))1_{\{i=k\}}
\]

so that

\[
v_{k\ell}(i, q) = (c(k, q) - c(\ell, q))1_{\{i=k\}}
\]

The corresponding target set \(B_0\) is the negative quadrant. Evidently, given any \(q\), choosing a best-response \(i^* \in \text{argmin}_i c(i, q)\) ensures \(v(i^*, q) \leq 0\), which establishes approachability of \(B_0\).
Calibration

- Calibration, or calibrated forecasts (Foster and Vohra, ’98) provides a stronger sense of on-line prediction performance, which is intended specifically for probabilistic forecasts. Here the probability vector $p_t$ itself is issued as the forecast for the opponent’s (discrete) action $j_t$. 

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{\{p_t \in Q\}} (\delta z_t - p_t) = 0 \quad (a.s.)
\]

Approachability has been a fundamental tool in establishing calibration results as well. However, computationally efficient algorithms are just starting to emerge.
Calibration

Calibration, or calibrated forecasts (Foster and Vohra, '98) provides a stronger sense of on-line prediction performance, which is intended specifically for probabilistic forecasts. Here the probability vector $p_t$ itself is issued as the forecast for the opponent’s (discrete) action $j_t$.

An on-line forecaster is calibrated if for every Borel measurable set $Q \subset \Delta(J)$ and any strategy of the opponent, it holds that

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} 1\{p_t \in Q\}(\delta_{z_t} - p_t) = 0 \ (a.s.)$$
Calibration

- Calibration, or calibrated forecasts (Foster and Vohra, ’98) provides a stronger sense of on-line prediction performance, which is intended specifically for probabilistic forecasts. Here the probability vector $p_t$ itself is issued as the forecast for the opponent’s (discrete) action $j_t$.

- An on-line forecaster is *calibrated* if for every Borel measurable set $Q \subset \Delta(J)$ and any strategy of the opponent, it holds that

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} 1\{p_t \in Q\}(\delta_{z_t} - p_t) = 0 \ (a.s.)$$

- Approachability has been a fundamental tool in establishing calibration results as well. However, computationally efficient algorithms are just starting to emerge.
In some applications, the (convex) set $B$ to be approached is complex, and the projection step of Blackwell’s strategy becomes computationally hard. On the other hand, the response map $p^*(q)$ which satisfies $v(p^*(q), q)$ may still be simple (recall that existence of such a map is necessary for a convex set to be approachable).

In other case the set $B$ may be specified indirectly through the response function.

We next outline an approaching strategy that avoids the need for projections, but rather relies on a given response function $p^*$. 
Recall the original game with actions $i \in \mathcal{I}$, $j \in \mathcal{J}$, and payoffs $v(i, j) \in \mathbb{R}^m$. An approaching strategy must ensure
$\lim_{T \to \infty} d(\bar{v}_T, B) = 0$ for the given target set $B$.

Define the following auxiliary game:
- PL1’s actions: $(i, j^*) \in \mathcal{I} \times \mathcal{J}$
- PL2’s actions: $(i^*, j) \in \mathcal{I} \times \mathcal{J}$
- Payoffs: $w((i^*, ji^*)) = v(i^*, j^*) - v(i, j)$
- Average payoffs: $w_t = v(i_t, j_t) - v(i^*_t, j^*_t) \equiv v_t - v(i^*_t, j^*_t)$. 
**Lemma.** There exists a strategy for PL1 in the auxiliary game that ensures (a.s.),

$$\lim_{T \to \infty} \left( \bar{v}_T - \frac{1}{T} \sum_{t=1}^{T} v(p^*_t, q^*_t) \right) = 0$$

**Proof:** It is sufficient to show that $\bar{w}_t \to 0$, or equivalently that the set $B = \{0\}$ is approachable in the auxiliary game. We employ Blackwell’s dual condition. Given any mixed action $\eta$ of PL2 with marginals $p^*, q$, PL1 can simply choose the mixed action $\mu$ with marginals $p = p^*$, $q^* = q$ to obtain $w(\mu, \eta) = 0$.

Since this holds for arbitrary $p^*_t$, we can return to the original game, and choose (virtually) $p^*_t$ so that $v(p^*_t, q^*_t) \in B$, namely $p^*_t = p^*(q^*_t)$. This implies that the second sum in the Lemma is in $B$, and we are (almost) done.
Approachability w/o projection

- It still remains to specify PL1’s strategy that satisfies the lemma. We use Blackwell’s strategy for this problem.

- Recall that the payoff is $w$, and $B_0 = \{0\}$ (note that the original set $B$ is not involved here). Given a point $\bar{w}_{t-1}$, the vector from this point to $B_0$ is $w_{t-1}$ itself. Hence we require a mixed action $\mu = (p, q^*)$ so that, for all $\eta = (p^*, q)$,
  \[
  \bar{w}_{t-1} \cdot w(\mu, \eta) \leq 0
  \]
  or
  \[
  \bar{w}_{t-1} \cdot (v(p, q) - v(p^*, q^*)) \leq 0
  \]
  This holds if we choose $p_t$ (resp. $q_t^*$) as the minimax (resp. minimax solution of the matrix game with projected payoffs $a(i, j) = \langle \bar{w}_{t-1}, v \rangle(i, j)$, namely
  \[
  p_t = \arg\min_p \max_q \langle \bar{w}_{t-1}, v \rangle(p, q)
  \]
  \[
  q_t^* = \arg\max_p \min_q \langle \bar{w}_{t-1}, v \rangle(p, q)
  \]
Stochastic (or Markov) games, introduced by Shapley, are the multi-player extension of Markov decision processes.

Extensions of Blackwell’s approachability to stochastic games are described in S. & Shwartz (1993), Milman (2000), Kamal (2010).

No-regret algorithms for stochastic games have been considered in Mannor & S. (2003), Yu, Mannor & S. (2009), Even-Dar, Kakade & Mansour (2009).