Costly observation in a bandit model

with Lucie Ménager
Paris Assas University

May 29, 2015

Motivation

Consider the following social learning situation :

- uncertainty : payoffs are affected by an unknown state of nature θ ;
- costly exploration : agents may learn direct information about θ at some cost ;
- informational externality : agents may learn indirect information about θ from the observation of others.

Bandit problems in game theory

Continuous time and public information Bolton and Harris 1999, Descamps and Mariotti 2004, Keller, Rady, Cripps (KRC) 2005, Keller and Rady 2009 ;

Discrete time and private information Rosenberg, Solan and Vieille 2007, Valimki and Murto 2009, Heidhues, Rady, Strack 2013 ;

Continuous time and private information Rosenberg, Salomon, Vieille 2011, Bonatti and Hrner 2010, Hrner and Samuelson 2010 ;

Applications : Market learning, Research project, finance, matching market...

Bandit problems in game theory

Continuous time and public information Bolton and Harris 1999, Descamps and Mariotti 2004, Keller, Rady, Cripps (KRC) 2005, Keller and Rady 2009 ;

Discrete time and private information Rosenberg, Solan and Vieille 2007, Valimki and Murto 2009, Heidhues, Rady, Strack 2013 ;

Continuous time and private information Rosenberg, Salomon, Vieille 2011, Bonatti and Hrner 2010, Hrner and Samuelson 2010 ;

Applications : Market learning, Research project, finance, matching market...

Exponential bandit machine (Keller Rady Cripps)

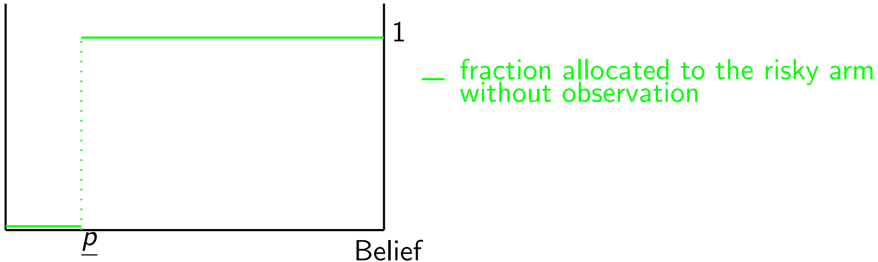
Payoff per unit of ressource :

Two arms
- S : deterministic payoff s
- R
  - Bad : deterministic payoff 0
  - Good : random payoff of mean h with random time of arrival of mean λ

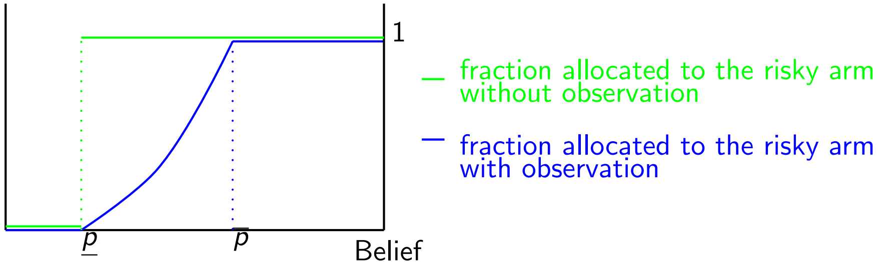
$g := \lambda h$ average payoff of a good risky arm

$0 < s < g$: players strictly prefer the risky arm if it is « good », and the safe arm if it is « bad »

Keller Rady and Cripps, KRC



Inefficiency in KRC.



What we do.

This paper introduces costly observation in KRC. We assume :

- complete observation : actions (experimentation and observation) and outcomes (the arrival of a breakthrough or not)
- irreversibility of stopping observation : once a player stops observing his opponent, he cannot observe him in the future

Remarks

- observation rather than communication

- paying observation rather than buying observation

What we find.

- We find a unique Markovian equilibrium if $c > 0$. This equilibrium is symmetric.

What we find.

- We find a unique Markovian equilibrium if $c > 0$. This equilibrium is symmetric.
- In comparaison with the symmetric equilibrium in KRC :

What we find.

- We find a unique Markovian equilibrium if $c > 0$. This equilibrium is symmetric.
- In comparaison with the symmetric equilibrium in KRC :
  - Players experiment more.

What we find.

- We find a unique Markovian equilibrium if $c > 0$. This equilibrium is symmetric.

- In comparaison with the symmetric equilibrium in KRC :
  - Players experiment more.
  - For some initial beliefs, they play a war of attrition to designate who will free ride and who will work.

What we find.

- We find a unique Markovian equilibrium if $c > 0$. This equilibrium is symmetric.
- In comparaison with the symmetric equilibrium in KRC :
  - Players experiment more.
  - For some initial beliefs, they play a war of attrition to designate who will free ride and who will work.
  - For some beliefs the payoff is unchanged and for other it is strictly smaller.

Model

Players and Actions

- 2 players play with two machines of the same type.
- Time is continuous.
- Discount factor : $r > 0$.
- Observation : $q_t^i \in \{0, 1\}$ with $q_t^i = 1$ iff i observes j in t
- Experimentation : $k_t^i \in [0, 1]$: fraction of the player's resource allocated to R in t

Model

Status

- $\Delta_t^i \in \{A, \bar{A}\}$: status of player i at time t.
- $\Delta_t^i = A$ (active) if player i paid the observation cost in all $t' < t$;
- $\Delta_t^i = \bar{A}$ (inactive) if player i did not pay the observation cost at some $t' < t$.

Because of the irreversibility assumption, $\Delta_t^i = \bar{A} \Rightarrow \Delta_{t'}^i = \bar{A}$ for all $t' \geq t$.

Model

Payoffs

p_t^i : probability assigned by player i at time t to the risky arm being good.

Model

Payoffs

p_t^i : probability assigned by player i at time t to the risky arm being good.

If i observes no breakthrough in $[t, t + dt[$, then

$$dp_t^i = -(k_t^i + q_t^i k_t^j) p_t^i (1 - p_t^i) \lambda dt$$

Model

Payoffs

p_t^i : probability assigned by player i at time t to the risky arm being good.

If i observes no breakthrough in $[t, t + dt[$, then

$$dp_t^i = -(k_t^i + q_t^i k_t^j) p_t^i (1 - p_t^i) \lambda dt$$

Payoffs at date t are, given p_t^i : $(1 - k_t^i)s + k_t^i g p_t^i - q_t^i c$

Model

Payoffs

p_t^i : probability assigned by player i at time t to the risky arm being good.

If i observes no breakthrough in $[t, t+dt[$, then

$$dp_t^i = -(k_t^i + q_t^i k_t^j) p_t^i (1 - p_t^i) \lambda dt$$

Payoffs at date t are, given p_t^i : $(1 - k_t^i)s + k_t^i g p_t^i - q_t^i c$
Intertemporal payoffs :

$$E\left[\int_0^\infty re^{-rt}\left[(1 - k_t^i)s + k_t^i g p_t^i - q_t^i c\right] dt\right]$$

The problem of a player is to determine the sequence of actions $(q_t, k_t)_t$ measurable with respect to the information available at time t, that maximizes her expected discounted payoff in the set of possible actions.

Model

Strategies

We restrict to Markovian strategies whose state variable depends on the subgame played :

Model

Strategies

We restrict to Markovian strategies whose state variable depends on the subgame played :

- when the player is inactive (case $\bar{A}$)

Model

Strategies

We restrict to Markovian strategies whose state variable depends on the subgame played :

- when the player is inactive (case $\bar{A}$)
- when the player is active while his opponent is inactive (case $A\bar{A}$)

Model

Strategies

We restrict to Markovian strategies whose state variable depends on the subgame played :

- when the player is inactive (case $\bar{A}$)
- when the player is active while his opponent is inactive (case $A\bar{A}$)
- when both players are active (case AA)

Subgame : the player is active and his opponent is inactive.

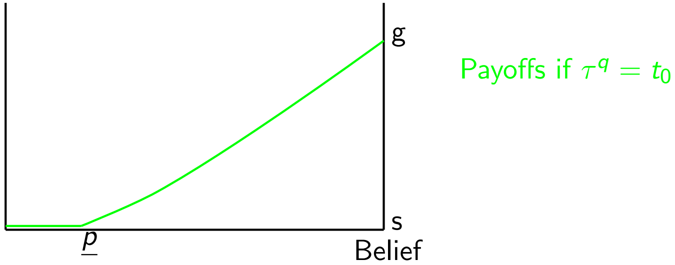
- Let p_0 be the (common) belief at which the subgame starts and the corresponding date t_0.
- Let τ_0 be the date at which player j stops experimenting.
- The i's utility maximization problem is a decision problem with two state variables : the i's current belief and the time it remains before j stops experimenting.
- We look for a random stopping time on stopping observation (τ^q) and an experimentation strategy k_i.

Subgame : the player is active and his opponent is inactive.

- We show that $\tau^q = \{t_0, \tau_0\}$.
- If $\tau^q = t_0$ then i receives the single player payoff.
- If $\tau^q = \tau_0$ then there is a belief $p(\tau_0) > p_0$ such that $k_t = 1$ if $p_t \geq p(\tau_0)$ and $k_t = 0$ otherwise.

Subgame : the player is active and his opponent is inactive.

Payoffs given an initial belief p_0.



Subgame : the player is active and his opponent is inactive.

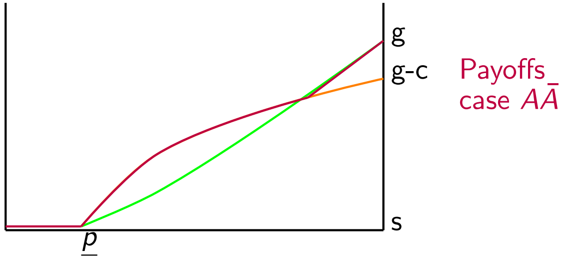
Payoffs given an initial belief p_0.

Subgame : the player is active and his opponent is inactive.

Payoffs given an initial belief p_0.

Subgame : the player is active and his opponent is inactive.

Payoffs given an initial belief p_0.



Subgame : Both players are active.

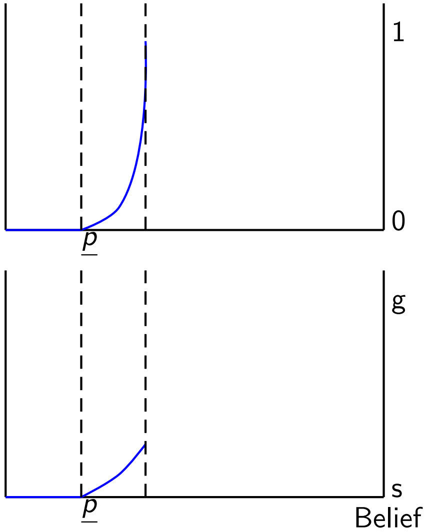
- If at t the subgame is in this category then players have a common belief.
- A natural state variable is then the common belief.
- The Bellman equation is :

$$u(p) = (1-q)u^{\bar{A}}(p)$$

$$+q(s - c + \max_k k(b(u,p) - c(p)) + k^j b(p,u))$$

Subgame : Both players are active.

(constrained)

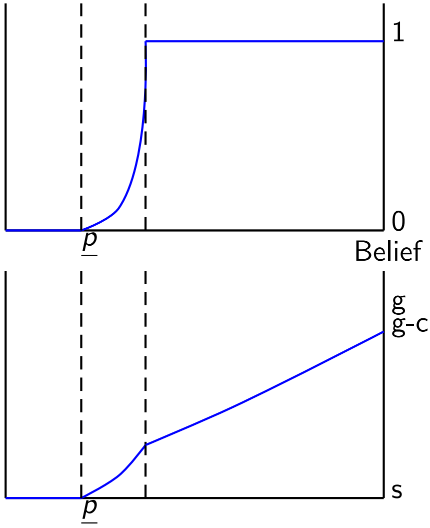


Experimentation strategy $k(p)$
case AA (constrained)

Payoffs
case AA (constrained)

Subgame : Both players are active.

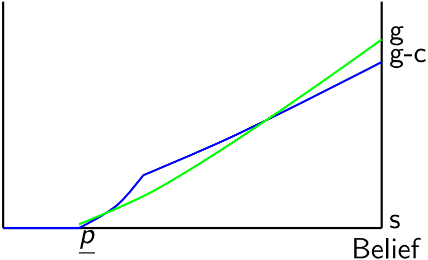
(constrained)



Experimentation strategy $k(p)$ case AA (constrained)

Payoffs case AA (constrained)

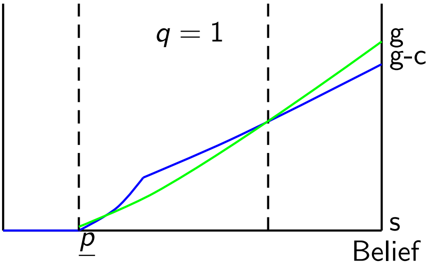
Subgame : Both players are active.



Payoffs case AA (constrained)

Payoffs case $\bar{A}$

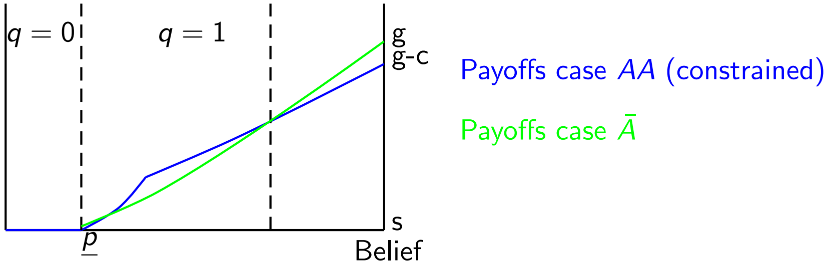
Subgame : Both players are active.



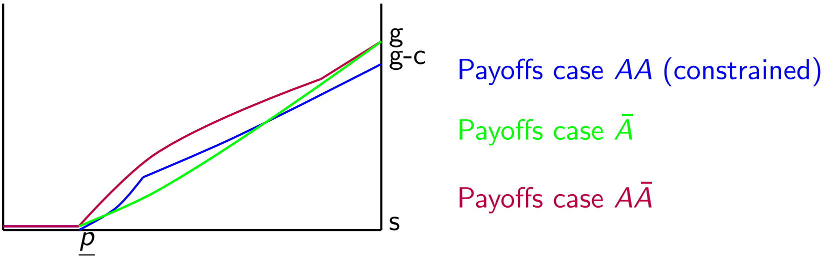
Payoffs case AA (constrained)

Payoffs case $\bar{A}$

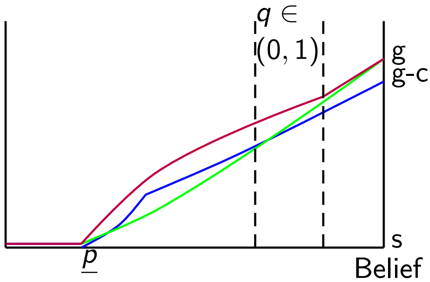
Subgame : Both players are active.



Subgame : Both players are active.

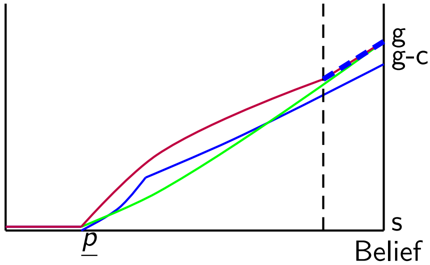


Subgame : Both players are active.



War of attrition

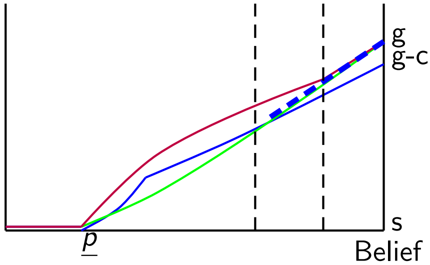
Subgame : Both players are active.



Payoffs symmetric equilibrium

$k = 1, q = 0$

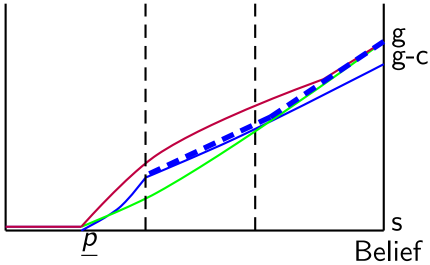
Subgame : Both players are active.



Payoffs symmetric equilibrium

$k = 1, q \in (0, 1)$

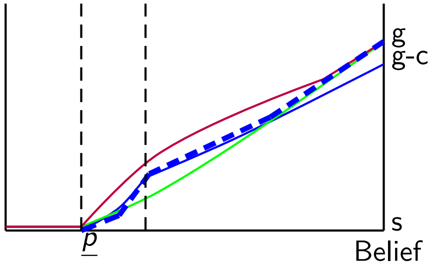
Subgame : Both players are active.



Symmetric equilibrium payoff

$k = 1, q = 1$

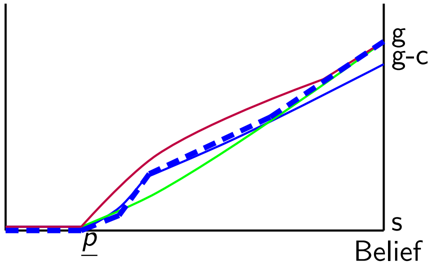
Subgame : Both players are active.



Symmetric equilibrium payoff

$k \in (0, 1), q = 1$

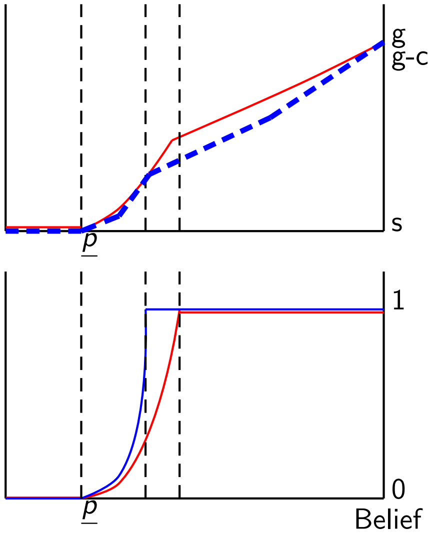
Subgame : Both players are active.



Symmetric equilibrium payoff

$k = 0, q = 0$

Symmetric equilibrium



Payoffs symmetric eq., $c > 0$

Payoffs symmetric eq., $c = 0$.

Experimentation strategy $k(p)$
case $c > 0$

Experimentation strategy, $c = 0$.