Games and Economic Behavior 82 (2013) 636–657

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

# Strategic learning in teams

Nicolas Klein *

*Université de Montréal and CIREQ*

A R T I C L E I N F O

*Article history:*
Received 3 June 2011
Available online 17 October 2013

*JEL classification:*
C73
D83
O32

*Keywords:*
Strategic experimentation
Three-armed bandit
Exponential distribution
Poisson process
Bayesian learning
Markov perfect equilibrium
R&D teams

A B S T R A C T

This paper analyzes a two-player game of strategic experimentation with three-armed exponential bandits in continuous time. Players play bandits of identical types, with one arm that is safe in that it generates a known payoff, whereas the likelihood of the risky arms' yielding a positive payoff is initially unknown. When the types of the two risky arms are perfectly negatively correlated, the efficient policy is an equilibrium if and only if the stakes are high enough. If the negative correlation is imperfect and stakes are high, there exists an equilibrium that leads to efficiency for optimistic enough *prior* beliefs.



## 1. Introduction

When faced with two competing hypotheses, economic agents often have to strike a balance between optimally using their current information on the one hand, and investing in the production of new information on the other hand. When doing so, they have to take into account the impact of their decisions not just on themselves, but on their partners and competitors also; indeed, the latter may benefit from the information a given agent produces. Think e.g. of two jurisdictions, or two hospitals, which are faced with a certain disease that could either be caused by a virus or by a bacterium.[1] Now, either hospital has to decide which of the two competing hypotheses to investigate, e.g. by administering either an antibiotic or an anti-viral drug. Once one of them has found out which hypothesis is true, both benefit from the discovery. Thus, while the costs of experimentation have to be borne privately, any information an agent produces is a public good. This makes for a situation in which a player's experimentation decisions are strategic, in that they affect the other player's payoffs.

I model this trade-off as a three-armed strategic bandit problem.[2] Specifically, I consider two players operating three-armed exponential bandits in continuous time. One arm is safe in that it yields a known flow payoff, whereas the other two arms are risky, i.e. they can be either good or bad. The risky arms are meant to symbolize two mutually incompatible hypotheses. In the baseline case, I assume that it is common knowledge that exactly one of them is good. In Section 6, I extend the analysis to the case in which there is some chance that both hypotheses may be bad, while maintaining the

* Correspondence to: Université de Montréal, Département de Sciences Économiques, C.P. 6128 succursale Centre-ville; Montréal, H3C 3J7, Canada.
*E-mail address:* kleinnic@yahoo.com.

[1] See Klein and Rady (2011).

[2] For an overview of the bandit literature, see Bergemann and Välimäki (2008).

0899-8256/$ – see front matter 
http://dx.doi.org/10.1016/j.geb.2013.09.008

assumption that the two hypotheses cannot be true at the same time; i.e. it could be that the disease is neither caused by a virus nor by a bacterium (it could e.g. be genetic), but it is certainly not caused by both. Players are playing exact carbon copies of the same bandit machine; conditional on the state of the world, draws are iid between the players (i.e. players are playing so called *replica bandits*). The bad risky arm never yields a positive payoff, whereas a good risky arm yields positive payoffs after exponentially distributed times. As the expected payoff of a good risky arm exceeds that of the safe arm, players will want to know which risky arm is good. As either player's actions, as well as the outcomes of his experimentation, are perfectly publicly observable, there is an incentive for players to free-ride on the information the other player is providing; information is a public good.

Observability, together with a common prior, implies that the players' beliefs agree at all times. As only a good risky arm can ever yield a positive payoff, all the uncertainty is resolved as soon as either player has a breakthrough on a risky arm of his and beliefs become degenerate at the true state of the world. As all the payoff-relevant strategic interaction is captured by the players' common belief process, I restrict players to use stationary Markov strategies with their common posterior belief as the state variable, thus making my results directly comparable to those in the previous strategic experimentation literature.

Players have to bear experimentation costs privately; the benefit, by contrast, is public. There are thus obvious incentives for players to free-ride on their partner's experimentation. Indeed, inefficiency because of free-riding has been a staple result of the literature on strategic experimentation with bandits. Using distributional assumptions similar to ours, Keller et al. (2005) e.g. have shown that in the game with positively correlated bandits, the efficient benchmark is never sustainable in equilibrium.[3] In Klein and Rady (2011), however, full efficiency, regarding both the amount and the speed of experimentation, is an equilibrium if, and only if, stakes are *below* a certain threshold. In the present setting, by contrast, I show that, if players know that exactly one risky arm is good, free-riding incentives can be completely overcome if and only if the stakes *exceed* a certain threshold; in this case, there exists an efficient equilibrium. The result extends to the imperfect negative correlation setting of Section 6 in the sense that, for high enough stakes, there exists an equilibrium that leads to players' behaving efficiently provided their *prior* beliefs are optimistic enough.

The rest of the paper is structured as follows: Section 2 reviews some related literature; Section 3 introduces the model; Section 4 analyzes the utilitarian planner's problem; Section 5 analyzes the non-cooperative game, exhibiting a necessary and sufficient condition for the existence of an efficient equilibrium; Section 6 investigates the robustness of our main result to a more general correlation structure; and Section 7 concludes. Some auxiliary results and proofs are provided in Appendices A and B.

## 2. Related literature

Chatterjee and Evans (2004) analyze a treasure-hunting game in discrete time, where it is common knowledge that exactly one of several projects is good. As in my model, they allow players to switch projects at any point in time. The game ends as soon as one of the players finds the treasure. As in my model, they find that, for high enough stakes, there exists an efficient equilibrium. The forces leading to this result are very different from those at work in my model, though. The nub is that the Chatterjee and Evans (2004) game also involves payoff externalities, in the sense that if player 1 finds the treasure it is lost to player 2. In actual fact, their winner takes all, and hence internalizes all the social gains of his discovery. In my model, by contrast, externalities are purely informational in nature; when one agent makes a discovery, the other agent fully profits from the information generated by this breakthrough, as he will imitate his partner's successful approach in the future. Thus, Chatterjee and Evans' (2004) model may be better-suited e.g. to the analysis of experimentation by rival firms competing for market share; mine may be more appropriate to the case of different jurisdictions investigating the impact of various treatment options for a particular disease, or if e.g. one wants to analyze free-riding incentives by scientists working for the same firm or in the same lab, and the like.

My model lets agents choose themselves which hypothesis to explore; Klein and Rady (2011) by contrast assign one hypothesis each to either player. The comparison between my model and Klein and Rady (2011) would suggest that delegation was a good idea if the stakes at play were high; if the stakes are low, however, it can be better to assign one hypothesis each to either player, the comparison suggests. Indeed, it is notable that, depending on the circumstances, firms or institutions seem to pursue quite different approaches in this respect. For instance, subsequently to marked growth in the number of its research laboratories and facing increasing competitive pressures, 3M, which arguably makes more low-stakes products such as adhesives and abrasives, moved to restrict scientists' discretion over their work, which had traditionally been very vast (see Bartlett and Mohammed, 1995). By contrast, firms in the arguably more high-stakes pharmaceutical sector moved in the opposite direction. Swiss pharmaceutical giant Novartis, for instance, entered into a multi-million five-year agreement with the Department of Microbial and Plant Biology at Berkeley, CA, delegating project decisions to a committee being comprised of five experts, only two of whom were Novartis employees (see Lacetera, 2009)—a scheme that can reasonably be interpreted as a commitment device on the part of Novartis to delegate project choice to their scientific partners in academia. A somewhat similar deal had earlier been signed by Thousand Oaks, CA, based pharmaceutical company Amgen

[3] See also Bolton and Harris (1999, 2000), and Keller and Rady (2010).