Bandit Models and Blotto Games

August 10, 2011

Caroline Désirée Thomas

Thesis submitted for a Ph.D. in Economics.

Department of Economics

University College London

To Alexander.

I thank Martin Cripps and Guy Laroque for their supervision.
I am grateful to the members of my thesis committee, Godfrey Keller and Lars Nesheim,
for helpful and constructive comments.

I, Caroline Désirée Thomas, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

In this thesis we present a new take on two classic problems of game theory: the "multiarmed bandit" problem of dynamic learning, and the "Colonel Blotto" game, a multidimensional contest.

In Chapters 2-4 we treat the questions of experimentation with congestion: how do players search and learn about options when they are competing for access with other players? We consider a bandit model in which two players choose between learning about the quality of a risky option (modelled as a Poisson process with unknown arrival rate), and competing for the use of a single shared safe option that can only be used by one agent at the time.

We present the equilibria of the game when switching to the safe option is irrevocable, and when it is not. We show that the equilibrium is always inefficient: it involves too little experimentation when compared to the planner solution. The striking equilibrium dynamics of the game with revocable exit are driven by a strategic option-value arising purely from competition between the players. This constitutes a new result in the bandit literature. Finally we present extensions to the model. In particular we assume that players do not observe the result of their opponent's experimentation.

In Chapter 5 we turn to the n-dimensional Blotto game and allow battlefields to have different values. We describe a geometrical method for constructing equilibrium distribution in the Colonel Blotto game with asymmetric battlefield values. It generalises the 3-dimensional construction method first described by Gross and Wagner (1950). The proposed method does particularly well in instances of the Colonel Blotto game in which the battlefield weights satisfy some clearly defined regularity conditions. The chapter also explores the parallel between these conditions and the integer partitioning problem in combinatorial optimisation.

Contents

Chapter 1

Introduction

In this thesis we present a new take on two classic problems of game theory: the "multiarmed bandit" problem of dynamic learning, and the "Colonel Blotto" game, a multidimensional contest. In Chapters 2-4 we treat the questions of experimentation with congestion: how do players search and learn about options when they are competing for access with other players? In Chapter 5 we explore the n-dimensional Blotto game when battlefield values can vary.

In the standard multiarmed bandit problem, a player faces several objects delivering stochastic payoffs. Each object is associated with a distribution that is unknown to the player, but about which he can learn by sampling. This class of problems is commonly used to illustrate the situation faced by a player whose information about, say, his valuation for an object is slow to arrive. At each trial the player faces the trade-off between "exploitation", i.e. maximising his expected payoff based on information he already has, and "exploration", i.e. looking for information about other sources of payoff.

In this thesis we consider an extension of the multiarmed bandit problem in which two players are simultaneously learning about their (independent) valuation of options, and can get in each other's way: while one player is experimenting with one option, the other player can use any other option but this one.

This model addresses the question of how players search and learn about options when they are in competition with other players. Consider an agent who searches for an option with which to be matched: a job, a spouse, a second-hand car, a flat-share. Information about the quality of a match is slow to arrive. In this context it is natural to think about the option as a one-armed bandit. If there are other agents in the market engaging in similar search and only one agent at a time can access an option, we refer to this phenomenon as congestion. For instance to learn about the quality of a second-hand car, you need to take it for a test-drive. While you are doing this no other agent can.

Spending time learning about the quality of an option is costly in that it involves the risk of losing access to other options. While you are test-driving one car, other agents may be buying other cars without you having had the opportunity to test these. This can be thought of as the opportunity-cost of learning. At the same time, you may now be willing to spend more time learning about that one car if you knew that another potential buyer is interested in it. If you leave it, he is likely to buy it, making it henceforth unavailable to you. There is a pressure exerted by the "second in line".

These sorts of considerations are common in all kinds of strategic situations. In the race towards developing a new technology a firm's incentive to invest in a new research project rather than relying on some averred method depends on the strategy of the competitor if the market can only support one producer. The love interest of a pretender may be enhanced by the presence of a rival. When looking for a parking-space, do we take the one we just spotted or continue driving in the hope of finding a space closer to the cinema, but at the risk of losing the first one?

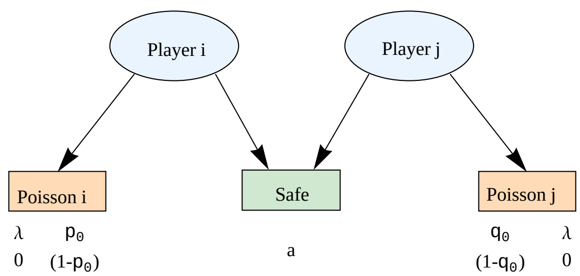
Furthermore, the fact that buyers may come to face these strategic situations has been internalised by some markets. For instance, web-sites like Amazon or Opodo will tell you when there is only one copy of this book left in stock, or only one seat left on that airplane, thus bringing to your attention that browsing one more book or flight may come at the cost of losing the one you just considered. We could also think of tips for wedding-gown sales provided in the feminine press for so-called 'bride-runs': during wedding-gown sales, brides-to-be are advised to ask friends along who can then hold on to gowns they don't want to return to the floor, where other potential buyers may take them.

The notion of congestion we consider bears some resemblance to the idea of *exploding offer*. A buyer either has to settle for an option now, or risk losing it to someone else, and then wait until they maybe leave the option before he can access it. Rather than the offer expiring exogenously, it expires because someone else has taken it.

The aim of the first chapters is to propose a simple model of experimentation with congestion, in which to analyse the trade-offs from strategic interaction. We are not attempting to capture some particular market as best we can, but rather, to offer a framework in which we can isolate the new *strategic considerations* that emerge as a result of congestion. To this end, we consider a model in which two players choose between learning about the quality of their risky option and switching to a common option. In this market, congestion arises if both players want to be matched with the common option. To underline the strategic incentives around the common option, we assume that nothing can be learned from it: it is "safe" in that it delivers a known constant flow-payoff.

Risky options are modelled as Poisson processes whose arrival rates are unknown to the players. We also assume that the arrival rates of the risky options are independent so that each player can learn nothing about the quality of his risky option from the actions or payoffs of the other player. Each risky option is either "good" and yields a lump-sum payoff of 1 at rate λ to the player activating it, or it is "bad" and always yields zero. This assumption makes the motion of beliefs monotonic: as long as a risky option is activated and does not produce a success, the belief about the quality of that option decreases. Once an option has produced a success, it is known to be good. So in our model, the strategic interaction takes place as players wait for the first Poisson event. During the game, players observe each other's actions and payoffs and so share a common belief about the qualities of the risky options.

The safe option yields a flow payoff of $0 < a < \lambda$ with certainty to the player occupying it and this is common knowledge. A player who occupies the safe option gains absolute priority over its use; his opponent can then only use the safe option if the first player leaves it and returns to his risky option. If both simultaneously decide to move from their risky option to the safe option, then a tie-break rule specifies the probability with which either player gains access to the safe option.



Each player has access to his risky option and to a shared safe option that can only be occupied by one player at a time. Player i's risky option is good with prior probability p_0, player j's with prior probability q_0.

Our main result is two-fold: First, we show that strong preemption motives arising as part of the strategic interaction mean that the equilibrium always involves inefficiently low levels of experimentation and unraveling of the exit decision. This was to be expected in light of the literature on preemption games (Fudenberg and Tirole (1985)), and our model affords us a clear illustration of the mechanism leading to the inefficiency. The second, more striking result is that when the exit decision is revocable, in equilibrium a player may strategically block the safe option temporarily in order to *force the other player to experiment.* This is possible because the first player can *commit* to leaving the safe option eventually, even as his opponent's demand for the safe option *intensifies.*

The main finding of this part of the thesis lies in recognising that there exists a *strategic option-value* associated with occupying an option. This has a number of interesting implications. First, it gives rise to behaviour excluded in the standard bandit model, in particular the temporary interruption of experimentation. This is because congestion may make options more attractive than they would be without. Second, it implies that preemption need not be irreversible.

To our knowledge, this model is the first to present the disappearance of an option from a multi-armed bandit as the result of strategic interaction. Dayanik et al. (2008) examine the performance of a generalised Gittins Index for the case in which a player must decide at each point in time which of N arms to activate, knowing that arms may *exogenously* break down, and thereby disappear from the choice set, temporarily or permanently. In particular they observe that the potential disappearance of arms may disrupt learning as the optimal policy is increasingly biased towards maximising one's payoff based on current information ("exploitation") and away from acquiring new information ("exploration") as the probabilities of breakdown of arms increase.

In the economic literature, multi-armed bandit models have been augmented with various other strategic complications. For instance, Keller, Rady, and Cripps (2005) assume that all players want to learn about the same risky arm whose payoff realisations are publicly observable so that players have an incentive to free-ride on the experimentation of others. Murto and Välimäki (2011) assume that the qualities of different arms are correlated but their payoff realisations are private information to the players who only observe one another's decision to continue experimenting or stop. They find that information aggregates in distinct burst of activity, or "exit waves". In particular, multi-armed bandit models have been widely used to model job search (Jovanovic (1979)) and more recently to describe within-firm job allocations and trial periods for new recruits once their wage contract has been set. The explicit modelling of prices can then be dispensed with. For instance, Camargo and Pastorino (2010) point out that incentive pay is not widespread when employment happens at a probationary stage.

The bandit problem translates into the job assignment example as follows: Assuming that the productive characteristics of a new recruit are not perfectly observable, but that information about a worker's ability can be acquired by observing the worker's performance on a given task, the employer trades off the profit loss he may incur if the new recruit is ill-suited to the task with the benefit of acquiring new information about that worker's skill. If the worker does not know his own skill, he faces a similar problem.

In the context of our model, consider a firm in which two workers have been recruited to perform identical jobs. Each worker does not yet know his level of skill at that particular task, and both workers' skills are independent. If he discovers that he is skilled, a worker's expected payoff is positive, if he is unskilled, his payoff is zero. At any time, a worker can ask for the support of a scarce management resource. In that case credit is irrevocably shared and the worker earns less than if he were skilled and succeeded by himself, but more than if he were unskilled and trying to work by himself. Crucially, the manager can assist only one worker at a time.

If the initiative to assist workers lies with the manager, he behaves like a social planner. We find that he would optimally let workers try to solve the task by themselves for longer than he would if there were only one worker. If on the other hand, it is the worker's decision to solicit the manager's assistance, workers face a strategic situation in which the trade-off between learning and collecting a payoff is supplemented by a race to the safe option. We find that in equilibrium, the threat of congestion makes workers act increasingly myopically, leading to extreme inefficiencies.

More generally in the context of two-sided matching markets in which information about the quality of a match arrives slowly, the inefficient unravelling caused by the incentive to anticipate the decision of opponents is well documented, for instance in markets for lawyers (Posner et al. (2001)) or gastroenterologists (Niederle and Roth (2009)). A popular example in the economic literature is the US market for new doctors (Roth (1984)). In the early 1940's hospitals would hire medical students as future interns or residents two years in advance of their graduation, so that the matching was done before crucial information about students (such as skills or preferences for a particular medical specialisation) became available. The results in this thesis may contribute to better understanding pathologies of decentralised matching markets, in which agents only gradually learn about the quality of their match.

To illustrate the equilibrium when the decision to switch to the common option is assumed to be revocable, we can think of the village sweetheart who has two suitors. Only one suitor at a time may date the sweetheart, or they may pursue their search for a partner in the city, where there is no congestion. In equilibrium we find that the suitor who is most likely to be successfully paired in the city will date the village sweetheart first with the sole aim of deterring the rival, who is then forced to search in the city. If the rival were successfully paired there, the first suitor would be able to also search in the city or return to the village sweetheart without fear of rivals. But in equilibrium we find that the first suitor will eventually leave the sweetheart and search in the city *even if* the rival's claim to the village sweetheart is not dropped, but intensified.

The thesis is organised as follows: In Chapter 2 we formally model the risky and the safe options, the evolution of beliefs about the quality of the risky options as well as the rules of precedence for access to the congested safe option. These will constitute the building blocks for subsequent Chapters. We then present a set of efficient benchmarks. When there is no congestion, the planner problem reduces to a single-player two-armed bandit problem. We define the myopic and the optimal threshold beliefs, which will be recurring concepts throughout the thesis. When there is congestion we describe the planner solution for the case where the decision to allocate a player to the safe option is irrevocable and then when that decision is revocable.

In Chapter 3 we consider the two-player games in which we present the trade-offs from strategic interaction and derive the Markov Perfect Equilibria of the games, again distinguishing between the cases of irrevocable and revocable exit, for which we have provided efficient benchmarks.

In Chapter 4 we consider extensions to the two-player games, and look in particular at the game with irrevocable exit in which, this time, payoffs are *private*: a player can observe his opponent's behaviour, but does not observe whether he has already had a success or not. This section combines results and conjectures, as this extension constitutes work in progress, and sets the course for future research.

In Chapter 5 we consider the Colonel Blotto Model. Budget-constrained multidimensional allocation problems were amongst the very first ones considered in game theory. The first version can be found in Borel and Ville **?**. This problem and similar ones later came to be known as "Colonel Blotto" games, after Gross and Wagner's approach to the allocation problem (Gross and Wagner (1950)).

In the simplest version of the Colonel Blotto game, two generals want to capture three equally valued battlefields. Each general disposes of one divisible unit of military resources. The generals have to simultaneously allocate these resources among the three battlefields. A battlefield is captured by a general if he allocates more resources there than his opponent. The goal of each general is to maximise the number of captured battlefields.

In Chapter 5 we describe a geometrical method for constructing equilibrium distribution in the Colonel Blotto game with asymmetric battlefield values. The appeal of geometrical methods for constructing n-dimensional distributions subject to restrictions on their support and their margins lies in the relative simplicity with which they describe complicated multi-dimensional objects. The drawback is that they may fail to generate the full set of distributions satisfying given restrictions on support and margins. This downside is limited when that set is well defined, as it is here, so that the exercise becomes to generate instances of these well-defined objects.

The method presented in this chapter generalises to the n-dimensional case a construction method first proposed by Gross and Wagner. It does particularly well in instances of the Colonel Blotto game in which the battlefield weights satisfy some clearly defined regularity conditions. Though these conditions constrain the set of games in which this method reliably generates equilibrium strategies, they are less restrictive than the condition of symmetry across all battlefields (Laslier and Picard (2002)). Moreover, their implications suggest directions for further research.

Noticing that the conditions obtained can be interpreted as the requirement that there exists a coalition such that every battlefield is pivotal suggests a parallel between behaviour of candidates seeking to maximise plurality and candidates seeking to maximise probability of victory, though this chapter leaves the exact relationship between these games an open question. We consider the parallel with the constrained integer partitioning problem, or "bin-packing" problem particularly exciting.

Chapter 2

Experimentation with Congestion - Model and Benchmarks

2.1 Introduction

In this part we define the main component of the congestion game. In Section 2.2 we present the model on which Chapters 2-4 of this thesis are based. It is intended as a general and simple model of the phenomenon of congestion described in Chapter 1, and designed to outline the new strategic considerations that emerge as a result of congestion, in particular when compared with the standard multi-armed bandit model.

In Sections 2.3 we present a set of benchmarks, starting with the single-player problem. This is akin to the standard multi-armed bandit model in our model, and resumes the intuitions and result of that model. In particular, given the setup, there is no option-value associated with being able to return to his Poisson process for a player occupying the safe option. Two further benchmarks summarizing the planner solutions to the two-player games analysed in Chapter 3 are presented: the planner solution when switching to the safe option is irrevocable, and when it is not.

2.2 Model

In this section, we define the basic elements of the model on which further parts of this thesis will build: the risky option and the motion of beliefs about the quality of a risky option, the safe (potentially congested) option and the precedence rules determining access to the safe option. In all sections, time is continuous, ρ denotes the common discount rate, and each player maximises his expected discounted payoff over an infinite time horizon.

Risky option: Each risky option is either "good" and yields a lump-sum payoff of 1 at Poisson rate λ to the player activating it, or it is "bad" and always yields zero. The quality of each option is independently drawn at the beginning of the game: player i's risky option is good with probability p_0 and player j's risky option is good with probability q_0. This is common knowledge. Once a risky option has produced a success, it is known to be good. As long as a risky option produces only unsuccessful trials, the belief about that option being good decreases.

Beliefs: Payoffs are publicly observed, so given the players' common prior (p_0, q_0) about the qualities of player i and player j's risky options respectively, players share a common posterior at each date $t \geq 0$ denoted (p_t, q_t). If over the time interval $[t, t+dt)$, $dt > 0$, a player, say i, activates his risky option without it producing a success, the belief about player i's option at $t + dt$ is, by Bayes' rule,

$$p_{t+dt} = \frac{p_t\ e^{-\lambda dt}}{p_t\ e^{-\lambda dt} + 1 - p_t}.$$

This is decreasing in dt: the longer the player experiments without a success, the less optimistic he becomes about his risky option being good. When dt is small we obtain that $p + dp = \frac{p(1-\lambda dt)}{1-p\lambda dt}$. The law of motion followed by the belief when the risky option is activated over the time interval $dt \to 0$ and produces only unsuccessful trials is then

$$dp = -p(1-p)\lambda dt. \tag{2.1}$$

Notice that this expression is maximised when $p = 1/2$ and that when priors are different beliefs don't move at the same rate. Once a risky option has produced a success, the common belief about that option is equal to 1 and remains there forever. At any date $t \geq 0$ the expected arrival rate on player i's (j's) risky option is $p_t\lambda$ ($q_t\lambda$). Whenever $p_t \neq q_t$ we refer to the player with the highest expected arrival rate as the more *optimistic* player and to his opponent as the more *pessimistic* player.

Safe option: The safe option yields a flow payoff of a with certainty to the player occupying it and this is common knowledge. We choose $a \in (0, \lambda)$ with the implication that when the risky option is known to be good, it is strictly preferred to the safe option, and vice-versa when a risky option is known to be bad.

Precedence rule: While each player has exclusive and unconstrained access to his risky option, both players have access to the safe option, but it can only be activated by one player at a time. A player who occupies the safe option gains absolute priority over its use; his opponent can then only use the safe option if the incumbent player leaves it and returns to his risky option. If both players simultaneously switch from their risky option to the safe option, then a tie-break rule allocates the safe option to player i with probability $\iota \in (0, 1)$.

2.3 Benchmarks

In this section we present a series of planner problems intended as efficient benchmarks for the models of strategic interaction in Chapter 3. First we consider the situation in which there is no congestion on the safe option (Section 2.3.1). Each player then faces an identical two-armed bandit problem with one risky and one safe arm. This problem is standard and has often been analysed in the previous literature. We use it as a framework to introduce concepts and methods that are recurrent throughout this thesis. The socially optimal policy is to let each player experiment with his risky option for high enough beliefs. If the risky option produces a success, the player should never switch to the safe option. If it does not and the player becomes sufficiently pessimistic about the quality of his risky option, he should permanently switch to his safe option when his belief hits the threshold value $p_V > 0$, which we refer to as the *single-player optimal threshold belief*.

We also define the *single-player myopic threshold belief*, $p_M > p_V$, below which the immediate payoff from the safe option exceeds the immediate payoff from the risky option. The belief p_M is the optimal threshold of an infinitely impatient or "myopic" player. In contrast a non-myopic player finds it optimal to continue playing the risky option on the interval (p_V, p_M) in the hope of it producing a success as long as he is able to return to the safe option at a later date: for the patient player the available safe option generates a positive option-value, making experimentation beyond the myopic threshold worthwhile.

The two remaining planner problems present new results, and set the efficient benchmark for the game analysed in Chapter 3. We then assume that the safe option can be played by at most one player at a time. If a risky option is known to be good, it is optimal never to let the player who is activating it switch to the safe option. If neither option produces a success, the planner will eventually allocate one player to the safe option. In Section 2.3.2 we assume that the decision to let one player choose the safe option cannot be revoked, even if the other risky option should produce a success. In Section 2.3.3, we assume that the planner can do this without restrictions.

When this is the case, the planner allocates the player with the lowest belief, say player j, to the safe option once the belief about his risky option being good hits a threshold. This threshold is always *below* the single-player threshold, p_V. This is because the safe option provides an option-value for both players: allocating player j to the safe option costs player i the option-value. This is internalised by the planner who therefore delays the exit of player j. The higher the belief of the optimistic player, the lower the option-value of the safe option for him and the closer the socially optimal exit belief of the pessimistic player to the single-player optimum.

When exit is revocable, the planner problem is akin to a standard multi-armed bandit problem. At each date the planner may activate two out of three arms (two risky, one safe) over a time interval $\Delta > 0$ so as to maximise his expected discounted payoff. The planner solution is analog to the Gittins Index policy: he either allocates both players to their risky options or allocates the player with the lowest expected Poisson arrival rate to the safe option. We present the solution to the planner's problem as $\Delta \to 0$.

2.3.1 No congestion - Single-player model

First assume that there are two safe options. The planner maximises the joint payoff of both players. Since the qualities of the risky options are uncorrelated and players cannot hinder one another's access to the safe option, the planner problem is equivalent to solving two single-player problems: a player, say player i, has access to his risky option and to the safe option as described in Section 2.2.

This single-player problem is standard, and has been analysed in a setup very similar to ours by Keller et al. (2005). We do repeat this analysis so as to introduce notation as well as concepts that will be relevant in further sections, in which we will derive novel results. In particular, we use the framework of the relatively straightforward single-player problem to carefully spell out (in Appendix 2.4.1) the method for solving ordinary differential equations that will be used throughout this thesis.

Formally, the single agent solves the following dynamic problem: at each date t, he chooses which option to activate from the set $\{S, R\}$, where S and R denote the safe and risky options respectively. The state is summarised by the belief p_t.

For $p_t < 1$, i.e. for histories in which the risky option has not yet produced a success let k_t denote the probability with which the agent activates the risky option during the time interval $[t, t + dt)$. The player chooses a path $\{k_t\}_{t \geq 0}$ that maximises his expected payoff:

$$\mathbb{E}\left[\int_0^\infty e^{-\rho t}\ [k_t\ p_t\lambda + (1 - k_t)\ a]\ dt \mid p_0\right].$$

Notice that if the player were to play myopically ($\rho \to \infty$), he would only compare the immediate payoff from playing R with the immediate payoff from playing S. We call the "myopic stopping belief", p_M, the belief at which the myopic player finds it optimal to irreversibly switch to the safe option:

$$p_M = \frac{a}{\lambda}.$$

In contrast, a more patient player ($\rho < \infty$) will experiment with the risky option in the hope of discovering that it is good. Let $V(p)$ denote the value function associated with this problem. By Bellman's Principle of Optimality the value function $V(p)$ solves the following dynamic programme: for all $p \in [0, 1]$

$$V(p) = \max\{L^R V(p), L^S V(p)\}$$

with

$$\begin{aligned} L^S V(p) &:= \tfrac{a}{\rho} \\ L^R V(p) &:= p\lambda dt(1+(1-\rho dt)\tfrac{\lambda}{\rho}) + (1-p\lambda dt)(1-\rho dt)V(p+dp) \end{aligned}$$

where we have used the approximation $e^{-\rho dt} \simeq (1-\rho dt)$. When a trial on the risky option does not produce a success dp is defined in Equation 2.1.

We solve the agent's problem in Appendix 2.4.1 and obtain the threshold belief at which the agent optimally switches to the safe option:

$$p_V = \frac{a\rho}{\lambda(\rho+\lambda-a)}.$$

Throughout this thesis, we will refer to p_V as the *single-player optimal threshold*, and to p_M as the *single-player myopic threshold*. Notice that $p_V < p_M$, that both are increasing as the value of the safe option, a, increases and that p_V tends to p_M as $\rho \to \infty$. Lemma 1 describes the optimal behaviour in the single-player game and presents the value function. The detail of the proof can be found in Appendix 2.4.1.

Lemma 1. *For $p > p_V$, playing the risky option is optimal and*

$$V(p) = p\ \frac{\lambda}{\rho} + (1-p)\ \frac{a-\lambda p_V}{\rho\ (1-p_V)}\ \left(\frac{1-p}{p}\ \frac{p_V}{1-p_V}\right)^{\frac{\rho}{\lambda}},$$

while for $p \leq p_V$, playing the safe option is optimal and $V(p) = \frac{a}{\rho}$.

The first term, $p\frac{\lambda}{\rho}$, is the payoff from activating the risky option forever. The second term reflects the option-value of being able to switch to the safe option. It increases as p decreases, i.e. as the player becomes more pessimistic about the quality of the risky option. It is equal to zero when $p = 1$ and strictly positive for all $p \in [0, 1)$ which is why, for beliefs $p \in (p_V, p_M)$ the patient player continues to experiment with the risky option even though he would be maximising his immediate payoff by switching to the safe option. When $p = p_V$, the expected payoff from the risky option is so low, that the player prefers switching to the safe option.

2.3.2 Planner solution - Irrevocable exit

We now consider the planner problem in a model where two players each have access to a risky option as described in the previous section, but there is only one safe option that can be occupied by at most one player at a time. The social planner maximises the sum of both players' payoffs. At each date, he has the choice between letting both players experiment (RR) or retiring one player to the safe option irrevocably so that the other player must continue to experiment on his risky option forever (RS).

If there were two safe options, the planner solution would be to let each player follow the single-player optimal policy derived in Section 2.3.1. Here, however, the planner may only retire one player to the safe option. There is now an additional option-value compared with the single-player game: suppose a player's option is good but has not yet produced a success. If the player switches to the safe option, not only does he forego his own profit from the good option, there is now the additional loss of his opponent's option-value from being able to switch to the safe option. Because such mistakes are more costly here, there will be more experimentation than in the single-player game. We show that it is optimal for the planner to eventually retire the most pessimistic player (Lemma 2) and to make both player experiment beyond their single-player threshold (Lemma 3).

Let p_t and q_t respectively denote the belief at t that player i's and player j's risky options are good. Each belief follows the laws of motion described in Section 2.3.1. The state at t is summarised by the vector of beliefs $(p_t, q_t) \in [0,1]^2$.

Lemma 2. *If in state (p,q) the policy RS is optimal, the planner necessarily allocates the pessimistic player to the safe option.*

Proof: Assume by way of contradiction that the policy which allocates the player with belief $\max(p,q)$ to the safe option in state (p,q) is optimal when $p \neq q$. The joint continuation utility in state (p,q) is then $\min(p,q)\frac{\lambda}{\rho} + \frac{a}{\rho} < \max(p,q)\frac{\lambda}{\rho} + \frac{a}{\rho}$. So the policy which allocates the player with belief $\max(p,q)$ to the safe option in state (p,q) is dominated by the policy which retires the more pessimistic player in that state. □

We now formally describe the planner's problem. Because we have assumed that $0 < a < \lambda$, it is by design optimal never to retire a player whose risky option has produced a success. If only one risky option has produced a success, the joint payoff is then maximised by letting the other player follow the optimal single-player policy. As long as neither risky option has produced a success, i.e. for states such that $(p,q) \in [0,1)^2$, let $\kappa_t \in [0,1]$ denote the probability with which the planner makes both players activate their risky options during the time interval $[t, t+dt)$. Then with probability $(1-\kappa_t)$ the planner irrevocably retires the pessimistic player to the safe option. The planner chooses a path $\{\kappa_t\}_{t\geq 0}$ subject to the constraint that exit is irrevocable, so as to maximise the expected joint payoff:

$$\mathbb{E}\left[\int_0^\infty e^{-\rho t}\ \ (\kappa_t(p_t+q_t)\lambda + (1-\kappa_t)[\max(p_t,q_t)\ \lambda +\ a])\ \ dt \mid (p_0,q_0)\right],$$

where $(p_0,q_0) \in [0,1)^2$ is the vector of prior beliefs. Let $\mathcal{W}(p,q)$ denote the value function associated with this problem. It solves the following dynamic program: for all $(p,q) \in [0,1)^2$,

$$\mathcal{W}(p,q) = \max_{\kappa}\left\{\kappa\ L^{RR}\mathcal{W}(p,q) + (1-\kappa)\ L^{RS}\mathcal{W}(p,q)\right\} \tag{2.2}$$

where, by Lemma 2, we have

$$L^{RS}\mathcal{W}(p,q) := \max(p,q)\frac{\lambda}{\rho} + \frac{a}{\rho}.$$

The payoff to the policy RR satisfies:

$$\begin{aligned} L^{RR}\mathcal{W}(p,q,\kappa) := \quad & p\lambda dt\ q\lambda dt\ 2\tfrac{\lambda+\rho}{\rho} + (1-p\lambda dt)(1-q\lambda dt)(1-\rho dt)\ \mathcal{W}(p',q') \\ & +p\lambda dt\ (1-q\lambda dt)\big[\tfrac{\lambda+\rho}{\rho} + (1-\rho dt)V(q')\big] \\ & +q\lambda dt\ (1-p\lambda dt)\big[\tfrac{\lambda+\rho}{\rho} + (1-\rho dt)V(p')\big] \end{aligned}$$

where $V(p)$ denotes the value function of the single-player game (Lemma 1) and $p' = p + dp$ where dp is defined in Equation 2.1.

Solving the planner's problem (Appendix 2.4.2), we find the set of threshold beliefs at which the planner irrevocably allocates the player with the lowest belief to the safe option, forcing the other player to experiment on his risky option forever. That set of threshold beliefs is depicted in Figure 1 below.

Lemma 3. *In states (p,q) such that $p \geq q$, it is optimal for the planner to irrevocably retire the player with the lowest belief (player j) to the safe option if and only if*

$$q \leq \frac{a\ \rho}{\lambda\ (\lambda + \rho - a + \rho\ V(p) - p\ \lambda)} \leq q_V,$$

where the threshold is equal to q_V when $p = 1$. Otherwise, he optimally lets both players activate their risky options. Conversely for states such that $p \leq q$.

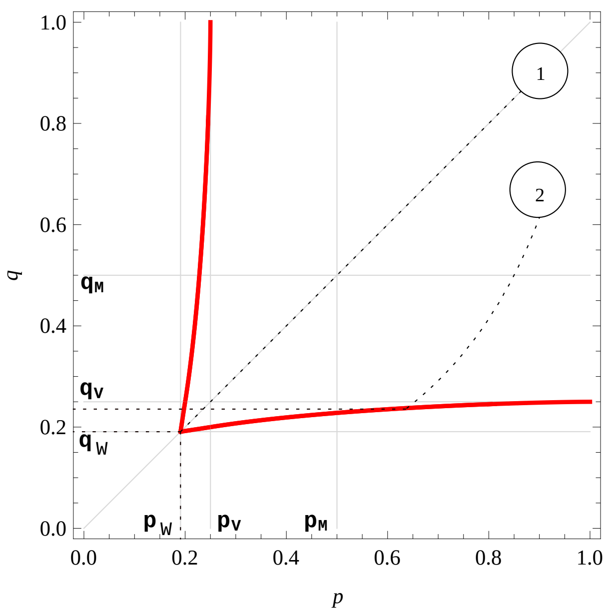


Figure 1: *Threshold beliefs in the planner problem with irrevocable exit.*

This planner solution offers a good insight about what is at stake in a model of experimentation with congestion. Regardless of the prior (p_0, q_0), the pessimistic player always experiments for longer than in the single-player case. This is because his switching to the safe option would cancel the option-value it affords to the optimistic player. That option-value increases when the optimistic player's belief falls, increasing the discrepancy between the pessimistic player's exit belief and the optimal single-player threshold. That discrepancy is maximised when $p_0 = q_0$. In contrast, when the belief of the optimistic player tends to one, the option-value for him of being able to switch to the safe option tends to zero and the pessimistic player's exit belief tends to his single-player optimal threshold belief.

We can already perceive that the externality imposed on his opponent by the player who takes the safe option is larger when the opponent is more pessimistic about his Poisson process. This happens when priors are closer. From this observation we can already conjecture that the competition for the safe option will be more intense the closer the priors.

Let us look at some typical trajectories of the state in this planner solution. If $p_0 = q_0$ (trajectory 1 in Figure 1) then as long as both players play their risky option without success we have $p_t = q_t$. The planner then optimally allocates either player to the risky option when the beliefs hit the threshold value

$$p_{\mathcal{W}} := \frac{1}{2\lambda}\left[\lambda + \rho - \sqrt{(\lambda+\rho)^2 - 4a\rho}\right] < p_V. \tag{2.3}$$

If he allocates, say, player i to the safe option, p remains forever equal to $p_{\mathcal{W}}$. Player j meanwhile is forced to experiment forever. If his risky option is bad q, will gradually decrease towards zero following the law of motion $dq = -q(1-q)\lambda dt$. If his risky option is good it will eventually produce a success.

If $p_0 > q_0$ (trajectory 2 in Figure 1) then as long as both players play their risky option without success the state moves following the trajectory depicted above. Once the state reaches the threshold described in Lemma 3 (thick red boundary in Figure 1) the planner allocates player j to the safe option. Then q remains constant while p gradually decreases to zero if player i's risky option is bad, or jumps to one with positive probability if the option is good.

2.3.3 Planner solution - Revocable exit

We now consider the planner's problem when the decision to retire one player to the safe option is revocable. The planner is de facto playing a multi-armed bandit problem: at each date the planner chooses to activate two out of three arms (two risky, one safe) over a time interval $\Delta > 0$ so as to maximise his expected discounted payoff. The optimal policy, following which the planner either allocates both players to their risky options or allocates the player with the lowest expected Poisson arrival rate to the safe option, will therefore be the equivalent of the Gittins Index[1] policy for our setting. In light of this, Lemma 4, the analogue to Lemma 2 in the previous section, seems trivial: an arm with a higher expected arrival rate produces a higher Gittins index. We present the solution[2] to the planner's multi-armed bandit problem as

[1] For a good summary of Gittins' pairwise interchange argument, see Frostig and Weiss (1999).

[2] Here we concentrate on the discrete-time approximation as the time interval tends to zero and will use the intuition from a discrete-time problem. Notice however that because it involves the planner alternating between two options, the existence of that solution is problematic in continuous time. Indeed the optimum in continuous time is achieved by mixing at each point in time. This is a well-known issue with continuous-time dynamic problems. For a good exposition of the issue and the way to address it, see Bellman (1957) Chapter 8.

$\Delta \to 0$. As in the previous section, the state at t is summarised by the vector of beliefs $(p_t, q_t) \in [0,1]^2$.

Lemma 4. *If the policy RS is optimal in state (p,q), then the planner allocates the pessimistic player to the safe option.*

Proof: Trivial in view of the optimality of the Gittins Index Policy: A risky option's Gittins index is increasing in its expected arrival rate. □

We now formally describe the planner's problem. For states such that $(p,q) \in [0,1)^2$, let $\bar{\kappa}_t \in [0,1]$ denote the probability with which the planner makes both players activate their risky options during the time interval $[t, t+dt)$. With probability $(1-\bar{\kappa}_t)$ the planner lets the player with the highest posterior belief at t activate their risky option during the time interval $[t, t+dt)$, while the player with the lowest posterior belief activates the safe option. The planner chooses a path $\{\bar{\kappa}_t\}_{t\geq 0}$ that maximises the expected joint payoff:

$$\mathbb{E}\left[\int_0^\infty e^{-\rho t} \ \left(\bar{\kappa}_t(p_t+q_t)\lambda + (1-\bar{\kappa}_t)[\max(p_t,q_t)\ \lambda + \ a]\right) \ dt \mid (p_0,q_0)\right],$$

where $(p_0, q_0) \in [0,1)^2$ is the vector or prior beliefs. Let $\mathcal{U}(p,q)$ denote the value function associated with this problem. It solves the following dynamic program: for all $(p,q) \in [0,1)^2$,

$$\mathcal{U}(p,q) = \max\{L^{RR}\mathcal{U}(p,q), L^{RS}\mathcal{U}(p,q)\}. \tag{2.4}$$

Here RR denotes the policy whereby both players play their risky option and RS the policy where the planner allocates the player with the lowest belief to the safe option, while the player with the highest belief experiments on his risky option.

We first derive the payoff from playing the policy RS forever. Consider states $p \geq q$ such that player i's risky option has a higher probability of generating a success than player j's option. As long as neither risky option produces a success, the policy RS involves first making the player with the high belief (player i) activate his risky option while player j occupies the safe option. Then q does not evolve while p decreases towards q following the law of motion for active options: $dp = -p\lambda(1-p)dt$.

Once $p = q$, the planner alternates the players on the safe option, generating the payoff $\mathcal{A}(p)$ as described in Appendix 2.4.3. Then, for $p \geq q$, the payoff to the policy RS is $(1 - e^{-\rho s})\left(\frac{a}{\rho} + p\,\frac{\lambda}{\rho}\right) + e^{-\rho s}\ \mathcal{A}(q)$, where $s = \frac{1}{\lambda}\ln\left[\frac{1-q}{q}\frac{p}{1-p}\right]$, which equals zero for $p = q$. Simplifying, we have that for all (p, q) such that $p \geq q$,

$$L^{RS}\mathcal{U}(p,q) := \left(1 - \left(\frac{1-q}{q}\frac{p}{1-p}\right)^{\frac{\rho}{\lambda}}\right)\left(\frac{a}{\rho} + p\,\frac{\lambda}{\rho}\right) + \left(\frac{1-q}{q}\frac{p}{1-p}\right)^{\frac{\rho}{\lambda}}\mathcal{A}(q),$$

with the corresponding expression holding for $p \leq q$.

To get an intuition about $\mathcal{A}(p)$, consider discrete-time planner problem in the state $p = q$. As seen in the previous section, if exit is irrevocable then once the planner follows policy RS he always allocates the same player, say j, to the safe option. The belief p about the quality of player i's risky option then decreases at rate $dp = -p(1-p)\lambda\Delta$, for a positive but small time interval Δ. In contrast, when exit is revocable, the planner can alternate players on the safe option. He can therefore let the players successively play their risky option in state p, thus getting twice as many trials at each belief p as when exit is irrevocable. The law of motion of beliefs is then $dp = -\frac{1}{2}p(1-p)\lambda\Delta$. Notice that, as with irrevocable exit, when $p \to 0$, the value of the RS policy, $\mathcal{A}(p)$, tends to $\frac{a}{\rho}$, which is the value of the multi-armed bandit problem when both risky options are known to be bad and there is only one safe option.

The payoff to the policy RR satisfies:

$$\begin{aligned} L^{RR}\mathcal{U}(p,q) := \quad & p\lambda dt\ q\lambda dt\ 2\tfrac{\lambda+\rho}{\rho} + (1-p\lambda dt)(1-q\lambda dt)(1-\rho dt)\ \mathcal{U}(p',q') \\ & + p\lambda dt\ (1-q\lambda dt)\big[\tfrac{\lambda+\rho}{\rho} + (1-\rho dt)V(q')\big] \\ & + q\lambda dt\ (1-p\lambda dt)\big[\tfrac{\lambda+\rho}{\rho} + (1-\rho dt)V(p')\big] \end{aligned}$$

where $V(p)$ denotes the value function of the single-player game (Lemma 1) and $p' = p + dp$ where dp is defined in Equation 2.1. The set of threshold beliefs at which the planner allocates the most pessimistic player to the safe option is depicted below.

Lemma 5. *The solution to the planner problem with revocable exit is depicted below: For states (p,q) in the shaded area, including the boundary, the planner optimally allocates the player with the lowest belief to the safe option over a period $\Delta > 0$. For states in the white area, the planner optimally lets both players activate their risky option over a period $\Delta > 0$. On the boundary, the planner is indifferent between the two policies. For $\Delta \to 0$, the planner solution is depicted in Figure 2.a. below.*

Proof: See Appendix 2.4.4.

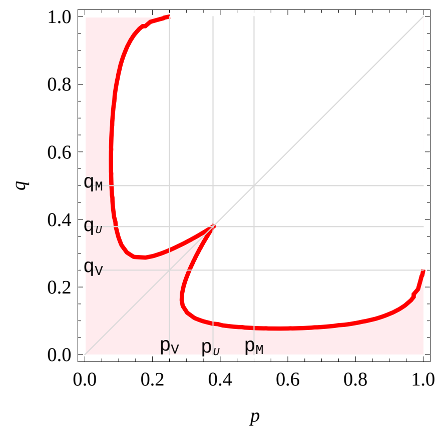
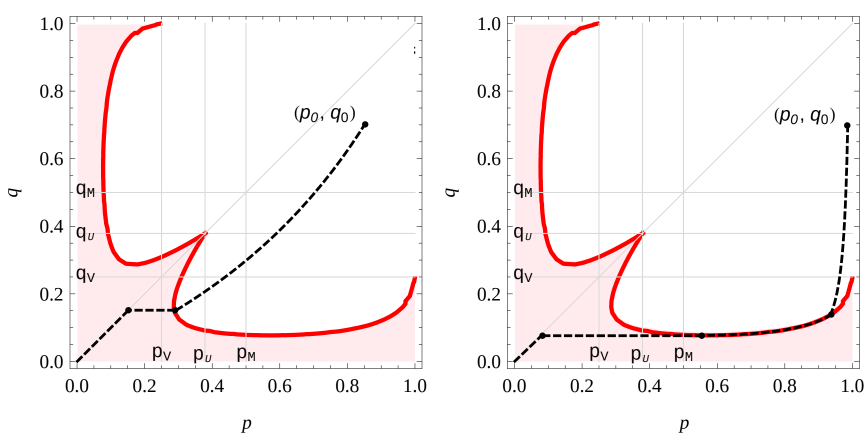


Figure 2.a: *Set of states in which policy RS is optimal (shaded area) and threshold beliefs in the planner problem with revocable exit.*

In Figure 2.a, consider the portion of the graph such that $p \geq q$. When $p = 1$, the socially optimal threshold belief is q_V, the threshold belief in the single-player game. This is because when $p = 1$, player i knows with certainty that his risky options is good, and so he will never threaten the safe option, so that player j effectively plays as in the single-player game. When $p = q$, the threshold belief $p_{\mathcal{U}} = q_{\mathcal{U}}$ is derived in Appendix 2.4.4.

We now illustrate typical trajectories of the state in the planner solution for the case where both Poisson processes are indeed bad. When the priors are relatively close, as in Figure 2.b, the planner lets both players experiment while the state is in the white region. Once the state hits the boundary, the planner allocates the player with the lowest expected arrival rate (in this case player j) to the safe option and lets the other player experiment until the state hits the 45 degree line. At that point, both players' expected arrival rates are equalised, and the planner alternates the players on the safe option.

If however the priors are such that one player is ex-ante much more likely than his opponent to have a good Poisson process, as in Figure 2.c, the planner first lets both players experiment as long as the state is in the white region. Again, once the state hits the boundary, the optimal regime-change requires that the planner allocate the player with the lowest expected arrival rate (player j) to the safe option. But then the state moves back into the white region, and the planner lets both players experiment again. As the time-interval δ becomes very small, this policy approximates the optimal policy in continuous time, which for such states requires moving *along* the boundary, as depicted in Figure 2.c. Eventually, the state stops moving back into the white region and the planner allocates player j to the safe option until the state hits the 45 degree line. The planner then alternates the players on the safe option.



Figures 2.b and 2.c: *Typical trajectories of the state in the planner solution. Black dots indicate the prior,* (p_0, q_0) *and subsequent regime-changes.*

Finally, notice the interesting discontinuity implied by the planner solution: fix player i's belief at $p_{\mathcal{U}}$ and consider the various optimal regimes depending on q. When $q = q_{\mathcal{U}}$, the planner finds it optimal to alternate the players on the safe option, thereby slowing down the decay in expected joint payoff. For $q > q_{\mathcal{U}}$ however, player j is still too optimistic relative to player i for the planner to alternate the players on the safe option, and player i is still too optimistic to be assigned the safe option without alternating with player j. Invert the reasoning for $q < q_{\mathcal{U}}$ but such that the sate falls in the white region. When q is such that $(p_{\mathcal{U}}, q)$ belongs to the boundary - which happens for q much below the single-player threshold - the planner retires player j to the safe option until the posteriors are equalised and the alternates the players on the safe option.

2.4 Appendix Chapter 2

2.4.1 Single-player - Value function

The derivation of the value-function for the single-player problem mirrors Keller et al. (2005). Consider states in which playing the risky option is optimal, so that $V(p) = L^R V(p) \geq L^S V(p)$. Using $V(p') = V(p+dp) = V(p) + V'(p)dp = V(p) - p(1-p)\lambda V'(p)dt$, we obtain the following ordinary differential equation for the value function:

$$p\lambda(1-p)\ V'(p) + (p\lambda + \rho)\ V(p) = p\lambda\ \frac{\lambda+\rho}{\rho}.$$

Solving, we obtain the solutions:

$$V_C(p) = p\ \frac{\lambda}{\rho} + C_V(1-p)\left(\frac{1-p}{p}\right)^{\frac{\rho}{\lambda}}$$

where C_V is the constant of integration. For all C_V, $V_C(p)$ is continuous and differentiable at $p \in [0,1]$.

At $p = 0$, the risky option is known to be bad, so the expected payoff from activating it is 0. Playing the safe option is therefore optimal at $p = 0$. At $p = 1$, the risky option is known to be good, and playing the risky option is optimal. So $V(0) = L^S V(0)$ and $V(1) = L^R V(1)$.

Assume there exists some belief $p_V \in (0,1)$ at which the player switches from the risky to the safe option. By continuity of the value function we then have $V_C(p_V) = L^S V(p_V) = \frac{a}{\rho}$ (value-matching). This regime change is optimal if and only if $V_C'(p_V) = L^S V'(p_V) = 0$ (smooth-pasting). At p_V, $L^S V(p_V)$ then constitutes a particular solution to the differential equation above. We obtain:

$$p_V = \frac{\rho a}{\lambda(\rho + \lambda - a)},$$

which is indeed below p_M, the myopic stopping belief. Finally, the constant of integration is then $C_V = \frac{a - p_V \lambda}{\rho p_V} \left(\frac{p_V}{1-p_V}\right)^{\frac{\rho}{\lambda}}$, for which $V_C(p)$ is increasing and convex on $[p_V, 1]$.

We conclude that for $p > p_V$, playing the risky option is optimal and

$$V(p) = p\,\frac{\lambda}{\rho} + (1-p)\,\frac{a - \lambda p_V}{\rho\,(1-p_V)}\,\left(\frac{1-p}{p}\,\frac{p_V}{1-p_V}\right)^{\frac{\rho}{\lambda}},$$

while for $p \le p_V$, playing the safe option is optimal and $V(p) = \frac{a}{\rho}$.

Notice finally that the regime switch is indeed optimal, as for $p \in [0,1)$, $V(p) > p\,\frac{\lambda}{\rho}$, the payoff from never switching to the safe option.

2.4.2 Irrevocable Exit - Planner Solution

Let $\mathcal{W}_p(p,q)$ denote the partial derivative of $\mathcal{W}(p,q)$ with respect to p. Similarly for q. Consider the states $(p,q) \in [0,1]^2$ in which having both players activate their risky option is optimal, so that $\mathcal{W}(p,q) = L^{RR}\mathcal{W}(p,q) \ge L^{RS}\mathcal{W}(p,q)$. We obtain the following partial differential equation for the value function:

$$\begin{aligned}(p\lambda + q\lambda + \rho)\,\mathcal{W}(p,q) + p\lambda(1-p)\,\mathcal{W}_p(p,q) + q\lambda(1-q)\,\mathcal{W}_q(p,q) \\ = p\lambda\left[\tfrac{\lambda+\rho}{\rho} + V(q)\right] + q\lambda\left[\tfrac{\lambda+\rho}{\rho} + V(p)\right].\end{aligned} \tag{2.5}$$

Letting $\tilde{\mathcal{W}}(s) \equiv \mathcal{W}(p(s), q(s))$, where $p(s) = \frac{p_0 e^{-\lambda s}}{1-p_0+p_0 e^{-\lambda s}}$, $q(s) = \frac{q_0 e^{-\lambda s}}{1-q_0+q_0 e^{-\lambda s}}$ and noticing that $\frac{d\tilde{\mathcal{W}}}{ds} = \frac{dp}{ds}\,\mathcal{W}_p + \frac{dq}{ds}\,\mathcal{W}_q$, we obtain the following ordinary differential equations in for $\tilde{\mathcal{W}}(s)$:

$$\begin{aligned}\tilde{\mathcal{W}}'(s) - (p(s)\lambda + q(s)\lambda + \rho)\,\tilde{\mathcal{W}}(s) = -p(s)\lambda\left[\tfrac{\lambda+\rho}{\rho} + V(q(s))\right] \\ -q(s)\lambda\left[\tfrac{\lambda+\rho}{\rho} + V(p(s))\right]\end{aligned} \tag{2.6}$$

Notice that when integrating terms including $V(.)$ on the right-hand-side, the single-player game threshold values $p_V = q_V = \frac{a\rho}{\lambda(\lambda+\rho-a)}$ will come to matter. Solving, we obtain the family of solutions for the value function:

$$\tilde{\mathcal{W}}_C(s) = H(p(s), q(s)) + H(q(s), p(s)) + \frac{p(s)\ q(s)}{p_0\ q_0\ e^{(2\lambda+\rho)s}}\ C_{\tilde{\mathcal{W}}}$$

where $C_{\tilde{\mathcal{W}}}$ is a constant of integration, and

$$H(x,y) = \begin{cases} x\ \frac{\lambda}{\rho} + x\ y \left(\frac{1-y}{y}\right)^{\frac{\lambda+\rho}{\lambda}} \frac{a-\lambda p_V}{\rho p_V} \left(\frac{p_V}{1-p_V}\right)^{\frac{\lambda+\rho}{\lambda}}, & p_V \leq y \\ \\ x\ \frac{\lambda}{\rho} \left(y\ \frac{\lambda+\rho+a}{2\lambda+\rho} + (1-y)\ \frac{\lambda+\rho+a}{\lambda+\rho}\right) & p_V \geq y. \\ +x\ (1-y) \left(\frac{1-y}{y}\right)^{\frac{\lambda+\rho}{\lambda}} \frac{a\lambda}{(\lambda+\rho)(2\lambda+\rho)} \left(\frac{p_V}{1-p_V}\right)^{\frac{\lambda+\rho}{\lambda}}, & \end{cases}$$

For all $C_{\tilde{\mathcal{W}}}$, $\tilde{\mathcal{W}}_C(s)$ is continuous and differentiable in s.

Let $L^{RS}\tilde{\mathcal{W}}(s) \equiv L^{RS}\mathcal{W}(p(s), q(s)) = \max(p(s), q(s))\frac{\lambda}{\rho} + \frac{a}{\rho}$. Consider the priors $p_0 \geq q_0$ both tending to 1. Then the payoff from letting both players activate their risky option tends to $\frac{2\lambda}{\rho} > \frac{a+\lambda}{\rho}$, and allocating both players to their risky option is optimal. Consider the case in which both risky options are bad so that as long as both players experiment, $\forall s \geq 0,\ 1 > p(s) \geq q(s)$ and as $s \to \infty$, both $p(s) \geq q(s)$ tend to zero. At that point, the expected payoff from letting both players activate their risky option tends to $0 < \frac{a}{\rho}$ and allocating one player to the safe option is optimal.

Assume that there exists some date $s_{\tilde{\mathcal{W}}} \geq 0$ at which the planner finds it optimal to irrevocably allocate the player with the lowest belief to the safe option. By the continuity of $\tilde{\mathcal{W}}$ we then have $\tilde{\mathcal{W}}_C(s_{\tilde{\mathcal{W}}}) = L^{RS}\tilde{\mathcal{W}}(s_{\tilde{\mathcal{W}}})$ (value-matching), and the regime change is optimal if and only if $\tilde{\mathcal{W}}'_C(s_{\tilde{\mathcal{W}}}) =$

$L^{RS}\tilde{\mathcal{W}}'(s_{\tilde{\mathcal{W}}})$ (smooth-pasting). Then, at $s_{\tilde{\mathcal{W}}}$, $L^{RS}\tilde{\mathcal{W}}(s_{\tilde{\mathcal{W}}})$ constitutes a particular solution to the differential equation above. The optimal switching date $s_{\tilde{\mathcal{W}}}$ solves, for $p(s_{\tilde{\mathcal{W}}}) \geq q(s_{\tilde{\mathcal{W}}})$:

$$q(s_{\tilde{\mathcal{W}}}) = \frac{a\ \rho - p(s_{\tilde{\mathcal{W}}})\ \lambda\ (\rho\ V(q(s_{\tilde{\mathcal{W}}})) - a)}{\lambda\ (\lambda + \rho + \rho\ V(p(s_{\tilde{\mathcal{W}}})) - p(s_{\tilde{\mathcal{W}}})\ \lambda - a)}.$$

For all $p(s_{\tilde{\mathcal{W}}}) \in [0,1]$ this equation admits solutions $q(s_{\tilde{\mathcal{W}}}) \in [q_{\mathcal{W}}, q_V]$ (where $q_{\mathcal{W}}$ is defined in Equation 2.3) so that $V(q(s_{\tilde{\mathcal{W}}})) = \frac{a}{\rho}$ and we obtain the expression in Lemma 3. The set of solutions is depicted in Section 2.3.2 in the belief space (p,q) . This defines a particular solution to the differential equation, which allows us to compute a closed-form expression for $\mathcal{W}(p,q)$.

2.4.3 Social Planner, Revocable Exit: Payoff from implementing policy RS forever when $p = q$.

RS denotes the policy whereby the planner always allocates the player with the lowest belief to the safe option, while the player with the highest belief experiments on the risky option. In the states (p,q) where the beliefs of the two players are equal $(p = q)$, the payoff to the policy RS satisfies:

$$\begin{aligned}\mathcal{A}(p) = \quad & adt + p\lambda dt \left[\tfrac{\lambda+\rho}{\rho} + (1-\rho dt)V(q)\right] \\ & +(1-p\lambda dt)(1-\rho dt)\Big[adt + q\lambda dt\left[\tfrac{\lambda+\rho}{\rho} + (1-\rho dt)V(p')\right] \\ & \qquad +(1-q\lambda dt)(1-\rho dt)\mathcal{A}(p')\Big]\end{aligned}$$

where $V(p)$ is the value function in the single-player game. Using $p = q$, $p' = p - p\lambda(1-p)dt$, $\mathcal{A}(p') = \mathcal{A}(p) - p\lambda(1-p)\mathcal{A}'(p)dt$ and eliminating terms $\in \mathcal{O}(dt^2)$, we obtain the following ordinary differential equation for $\mathcal{A}(p)$:

$$\mathcal{A}'(p) + \frac{2(p\lambda+\rho)}{p\lambda(1-p)}\,\mathcal{A}(p) = \frac{2a}{p\lambda(1-p)} + \frac{2}{(1-p)}\left[\frac{\lambda+\rho}{\rho} + V(p)\right].$$

Notice that when integrating the right-hand side, because it includes the function $V(p)$, the single-player threshold p_V will come to matter. Assuming that, if neither risky option ever produces a success, the policy RS is played forever, i.e. until $p \to 0$, at which point $\mathcal{A}(0) = \frac{a}{\rho}$, we obtain the solution:

$$\mathcal{A}(p) = e^{-\int f(p)\,dp}\int_0^p e^{\int f(x)\,dx}\; g(x) \;\; dx$$

where $f(p){:=}\frac{2(p\lambda+\rho)}{p\lambda(1-p)}$, $g(p) := \frac{2a}{p\lambda(1-p)}+\frac{2}{(1-p)}\left[\frac{\lambda+\rho}{\rho}+V(p)\right]$. Notice that $e^{\int f[x]\,dx}\big|_{x=0} = 0$. Solving, we obtain the following expression for $\mathcal{A}$:

$$\mathcal{A}(p) = \begin{cases} \frac{a}{\rho} + \frac{p\lambda}{\rho}\;\frac{(2\lambda+2\rho-\lambda p)}{(\lambda+2\rho)} & p_V \geq p \\ \\ \frac{a}{\rho}\left(1 - \frac{p\lambda(2\lambda+2\rho-\lambda p)}{(\lambda+\rho)(\lambda+2\rho)}\right) & p_V \leq p \\ +\frac{p\lambda}{\rho}\left(\frac{(2\lambda+2\rho-\lambda p)}{(\lambda+2\rho)} + \frac{p\lambda}{(\lambda+\rho)} + \frac{2(a-\lambda p_V)}{(\lambda+\rho)}\;\Omega(p,p_V,\frac{\lambda+\rho}{\lambda})\right) & \\ +\left(\frac{a}{\rho}\frac{p_V\lambda(2\lambda-p_V\lambda+2\rho)}{(\lambda+\rho)(\lambda+2\rho)} + \frac{p_V\lambda}{\rho}\frac{p_V\lambda-2a}{(\lambda+\rho)}\right)\left(\Omega(p,p_V,\frac{\lambda+\rho}{\lambda})\right)^2 & \end{cases}$$

where $\Omega(p,q,\alpha) = \frac{p}{q}\left(\frac{1-p}{p}\frac{q}{1-q}\right)^{\alpha}$ and p_V is the optimal stopping belief in the single-player game. □

2.4.4 Social Planner, Revocable Exit

In this section we describe the steps to derive the set of threshold beliefs in the planner problem with revocable exit, as illustrated in Figure 2. The method resembles the one used in Appendix 2.4.2 to derive the set of threshold beliefs in the planner problem with irrevocable exit. For all (p, q), the Bellman equation (2.4) for the planner's problem becomes

$$\mathcal{U}(p,p) = \max\{L^{RR}\mathcal{U}(p,p), L^{RS}\mathcal{U}(p,p)\}$$

where $L^{RS}\mathcal{U}(p,p) = \left(1 - \left(\frac{1-q}{q}\frac{p}{1-p}\right)^{\frac{\rho}{\lambda}}\right)\left(\frac{a}{\rho} + p\ \frac{\lambda}{\rho}\right) + \left(\frac{1-q}{q}\frac{p}{1-p}\right)^{\frac{\rho}{\lambda}}\mathcal{A}(q)$, and $L^{RR}\mathcal{U}(p,p)$ solves the ordinary differential equation (2.6), which is the ODE for $\mathcal{W}$ in the social planner problem with irrevocable exit. In Appendix 2.4.2 we have derived the family of solutions $\mathcal{W}_C$ to that ODE. We obtain the boundary in Figure 2.a by assuming that there exists some date $s_{\mathcal{U}}$ after which the planner finds it optimal to follow the policy RS forever, so that we can consider $L^{RS}\mathcal{U}(p(s_{\mathcal{U}}), q(s_{\mathcal{U}}))$ as a particular solution to ODE (2.6). Solving for $s_{\mathcal{U}}$ we then obtain the boundary in (p, q) space depicted in Figure 2.a.

As illustrated in Figure 2.c, for priors $1 > p_0 \gg q_0 > 0$, in our discrete-time approximation to the continuous-time optimum the planner will alternate between policies RS and RR, moving in broken horizontal and vertical lines about the boundary. As $\Delta \to 0$, this approximation gets aver closer to the continuous-time optimal policy, which involves the planner mixing so as to make the state move "along" the boundary. Our value-matching and smooth-pasting conditions are then in fact identifying an interval of dates over which the planner mixes so as to be indifferent between allocating the pessimistic player to the safe option or to his risky option over any time interval dt.

Finally, when $p = q$, the threshold belief $p_{\mathcal{U}} = q_{\mathcal{U}}$ satisfies

$$\begin{aligned} a(\lambda+\rho)(\lambda+2\rho) - (a-\lambda)\lambda^2 p_{\mathcal{U}}^2 \left(\tfrac{1-p_{\mathcal{U}}}{p_{\mathcal{U}}}\tfrac{p_V}{1-p_V}\right)^{\frac{2(\lambda+\rho)}{\lambda}} &= \\ p_{\mathcal{U}}\lambda\left(2(\lambda+\rho-a)(\lambda+\rho) + p_{\mathcal{U}}\lambda(a+\rho)\right) & \\ + p_{\mathcal{U}}\lambda(\lambda+2\rho)2(a-\lambda p_V)\tfrac{p_{\mathcal{U}}}{p_V}\left(\tfrac{1-p_{\mathcal{U}}}{p_{\mathcal{U}}}\tfrac{p_V}{1-p_V}\right)^{\frac{(\lambda+\rho)}{\lambda}} &. \end{aligned} \tag{2.7}$$

Chapter 3

Experimentation with congestion - Two-player Games

3.1 Introduction

We now consider the game in which there is congestion: as long as one player plays the safe option, it is unavailable to the other player. The players now interact strategically. They not only face the trade-off between exploration and exploitation, as in the single-player case, they must now consider the possibility of their opponent blocking their access to the safe option, temporarily or permanently. As a consequence, players will now have preemption motives. In section 3.2 we assume that once a player chooses to play the safe option, he may not return to his risky option. In this way, the decision to retire to the safe option is irrevocable. In Section 3.3, we will relax this assumption. Then a player can decide to temporarily occupy the safe option, before returning to his risky option. In each case we illustrate the equilibrium dynamics by considering some typical equilibrium state-trajectories.

In Section 3.2 we consider the strategic situation for which the planner problem analysed in Chapter 2 sets the efficient benchmark. We saw that because a player irrevocably switching to the safe option cancels the option-value it affords the other player, it is socially optimal for the pessimistic player to experiment for longer than in the single-player game. When players act strategically and compete for access to the safe option, to remedy the threat of being deprived of this option-value, they both have incentives to preempt the other player's switch. In equilibrium, the pessimistic player switches to the safe option in a state such that the optimistic player has no preemption motives. When there is sufficient competition between the players this will involve the pessimistic player switching to the safe option when the optimistic player's belief equals the myopic threshold.

The equilibrium will therefore be inefficient in the sense that the player capturing the safe option does so too early compared with the efficient threshold. When we intensify the degree of competition (by setting the priors closer to one another) this inefficiency increases until, for $p_0 = q_0$, the players behave myopically and completely disregard the option-value associated with experimenting on the risky option.

When exit is revocable (section 3.3), the player occupying the safe option is able to return to his risky option if his opponent's experimenting results in a success. In that case, relieved from the opponent's pressure on the safe option, the first player can achieve the utility of the single-player game. A player now has incentives to postpone his own experimenting and occupy the safe option so as to force his opponent to experiment in the hope of his producing a success and dropping his claim to the safe option.

In equilibrium, when there is sufficient competition for the safe option, the player with the highest expected arrival rate (the "optimist") temporarily occupies the safe option and forces the pessimist to experiment for a given duration of time. That duration increases with the competition for the safe option. Moreover it is such that the pessimist is always forced to experiment for longer than he would have in the single-player game. That duration is, however, finite and if the pessimist's experimenting is unsuccessful, the optimist eventually resumes his own experimenting, freeing up the safe option for the pessimist. This result may be surprising in light of intuitions from the standard multi-armed bandit problem in which, in the context of our model, a player would never return to a risky option he has rejected in the past.

3.2 Irrevocable Exit

We now consider the game in which two players each have access to a risky option and there is only one safe option that can be occupied by at most one player at a time. The risky and the safe options, as well as the rules of precedence are as described in Chapter 2. We assume that exit is once-and-for-all: once a player occupies the safe option, he may not switch back to his risky option. Under this condition, the assumption that the congested option is safe is without loss of generality: it could also be a risky bandit with expected arrival rate a.

Each player faces the trade-off between exploration and exploitation as described in the single-player game. Additionally, a player takes into account the fact that he loses the option-value from being able to switch to the safe option at a later date if his opponent occupies the safe option. As a result, in this game, there will be preemption motives leading to the unraveling of the exit decision.

We derive the Markov Perfect Equilibrium of this game and compare it with the planner solution derived in Chapter 2. We find that in equilibrium, the pessimistic player captures the safe option, and does so when the optimistic player's beliefs are greater than or equal to his myopic threshold belief. Though the pessimistic player would like to experiment until his belief reaches p_V, he is better-off exiting in a state in which his opponent has no preemption motives. The allocation of the safe option is efficient in that it goes to the same player as in the planner solution. However, the amount of experimentation by the pessimistic player is always inefficiently low. The closer the priors of the players, the greater the competition for access to the safe option, and the more inefficient the equilibrium.

Let us formally describe each player's problem. At each date, a player either chooses to activate his risky option over the time interval $[t + dt)$ (R) or to irrevocably switch to the safe option (S)[1] so as to maximise his expected discounted payoff. As in previous sections, the state is summarised by the vector of posterior beliefs $(p_t, q_t) \in [0, 1]^2$.

Because we have assumed that $0 < a < \lambda$, retiring to the safe option is

[1]Notice that the set of possible actions is not history dependent: we assumed that if a player switches to the safe option when that is already occupied by the opponent, the player "bounces" back to his risky option.

strictly dominated for a player whose risky option has produced a success. If only one risky option has produced a success, the other player maximised his expected discounted payoff by following the optimal single-player policy. We define a (Markovian) strategy $k^i(.)$ for player i to be the mapping $k^i : [0,1)^2 \to [0,1]$ from states (p_t, q_t) to k^i_t, the probability that player i plays his risky option at t. A (Markov-Perfect) equilibrium is a pair of strategies $(k^i(.), k^j(.))$ such that the strategy of player i maximises his expected discounted payoff conditional on the strategy of player j (subject to the constraint that exit is irrevocable), and vice-versa.

As in the previous sections, $V(.)$ denotes the value function in the single-player game. Let $\mathsf{W}(.)$ denote the value function in the two-player game with irrevocable exit. Given that, as long as neither player is occupying the safe option, player j uses the Markovian strategy $k^j(.)$ and plays his risky option in state (p,q) with probability $k^j(p,q)$, player i's value function solves the dynamic problem:

$$\mathsf{W}(p,q;k^j) = \max_{k^i(p,q)\in[0,1]} \{k^i(p,q)\ L^R\mathsf{W}(p,q;k^j) + (1-k^i(p,q))\ L^S\mathsf{W}(p,q;k^j)\}$$

where

(3.1)

$$\begin{aligned} L^S\mathsf{W}(p,q;k^j) :=\ & k^j(p,q)\ \tfrac{a}{\rho} + (1-k^j(p,q))\ T^i(p,q), \\ L^R\mathsf{W}(p,q;k^j) :=\ & p\lambda dt \left(1 + e^{-\rho dt}\tfrac{\lambda}{\rho}\right) \\ & +(1-p\lambda dt)\Big((1-k^j(p,q))\ e^{-\rho dt}p'\tfrac{\lambda}{\rho} \\ & \qquad +k^j(p,q)\ e^{-\rho dt}\left[q\lambda dt V(p') + (1-q\lambda dt)\mathsf{W}(p',q';k^j)\right]\Big), \end{aligned}$$

and with p', q' as defined in Chapter 2. The corresponding expression holds for player j. Because ties are broken in favour of player i with probability ι, player i and j's payoffs from a tie are respectively:

$$T^i(p) = \iota\ \frac{a}{\rho} + (1-\iota)\ p\ \frac{\lambda}{\rho}, \quad T^j(p) = (1-\iota)\frac{a}{\rho} + \iota\ p\ \frac{\lambda}{\rho}.$$

We now derive the unique[2] equilibrium of this game. We first show that there can be no equilibrium in mixed strategies (Lemma 6). Disregarding equilibria in weakly dominated strategies, we then present the Markov Perfect Equilibrium in the two-player game with irrevocable exit (Theorem 1). This equilibrium is inefficient, and we describe how it falls short of the planner solution derived in Chapter 2.

Lemma 6. *There exists no positive time interval $[t, t+dt)$, $dt > 0$ on which both players best-respond to one another by playing strictly mixed strategies.*

Proof: Suppose player j plays a strategy that lets him exit with positive probability at two distinct dates. If in state (p, q) player j switches to the safe option with strictly positive probability, player i can only be indifferent between his two pure strategies, when his belief is $p = p_M$. So there is no strictly positive time-interval over which player i is indifferent between switching to the safe option and continue activating his risky option. The detail of the proof can be found in Appendix 3.4.1. □

As long as player j switches to the safe option with strictly positive probability, player i is essentially trading off the payoff from winning a tie-break and irrevocably switching to the safe option, a/ρ, with the payoff from being stuck forever on this risky option, $p\lambda/\rho$. In states (p, q) such that $p = p_M$,

[2]Up to variations in weakly dominated strategies, which do not affect the equilibrium allocation or exit date.

the myopic exit belief in the single-player game, these payoffs are equalised and player i is indifferent between the outcomes, while in states (p, q) such that $p \neq p_M$, player i has strict preferences for either option. The remainder of this section hinges on this observation.

From Lemma 6, we conclude that the equilibrium strategies, given an initial state (p_0, q_0), involve either player switching to the safe option with certainty at some date $t \geq 0$ when the state is $(p(t), q(t))$. As argued in detail in Appendix 3.4.2 such an instantaneous switch cannot be an equilibrium in states (p, q) such that $p < p_M$, $q < q_M$, as a player's opponent then has strict incentives to preempt the player's switch. Over that support, there would be unraveling of the exit decisions as players try to preempt one another's switch. The preemption motives only disappear once at least one player is indifferent between switching to the safe option and staying with his risky option. Conversely, for beliefs above the myopic threshold, irrevocably switching to the safe option is strictly dominated by the strategy whereby the player commits to his risky option forever.

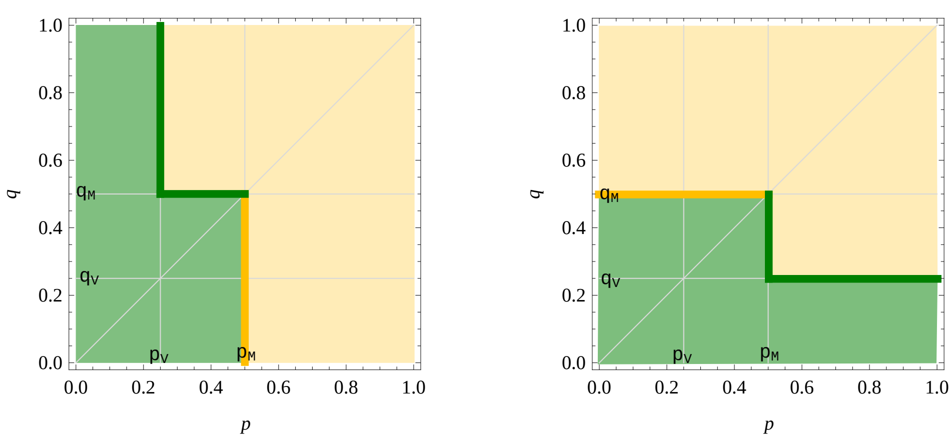


Figure 3: *Equilibrium strategies of player i and player j when exit is irrevocable. If a state (p, q) is in the green (dark) area, the player plays the safe option, if it is in the orange (light) area, the player plays his risky option.*

Theorem 1. *Consider the strategy profile illustrated in Figure 3:*

$$k^i(p,q) = \begin{cases} 0 \quad \textit{if} \begin{cases} q < q_M, \ p < p_M, \\ q = q_M, \ p \leq p_M, \\ q > q_M, \ p \leq p_V, \end{cases} \\ 1 \quad \textit{else.} \end{cases}$$

$$k^j(p,q) = \begin{cases} 0 \quad \textit{if} \begin{cases} p < p_M, \ q < q_M, \\ p = p_M, \ q \leq q_M, \\ p > p_M, \ q \leq q_V, \end{cases} \\ 1 \quad \textit{else.} \end{cases}$$

This constitutes the unique MPE of the game (up to variations in weakly dominated strategies for histories in which the safe option has already been allocated, so that they do not affect the allocation of the objects, given an initial state).

Proof: See Appendix 3.4.2. An intuition of the proof is given in the following illustrations. □

We now illustrate the resulting allocation and compare it with the planner solution for the case in which both risky options are in fact bad, so that beliefs never jump to one. Notice that in moving from case 1 to case 3, i.e. as the discrepancy in priors increases, the equilibrium exit belief of the pessimistic player gets closer to his single-player threshold - and also to the socially optimal exit belief.

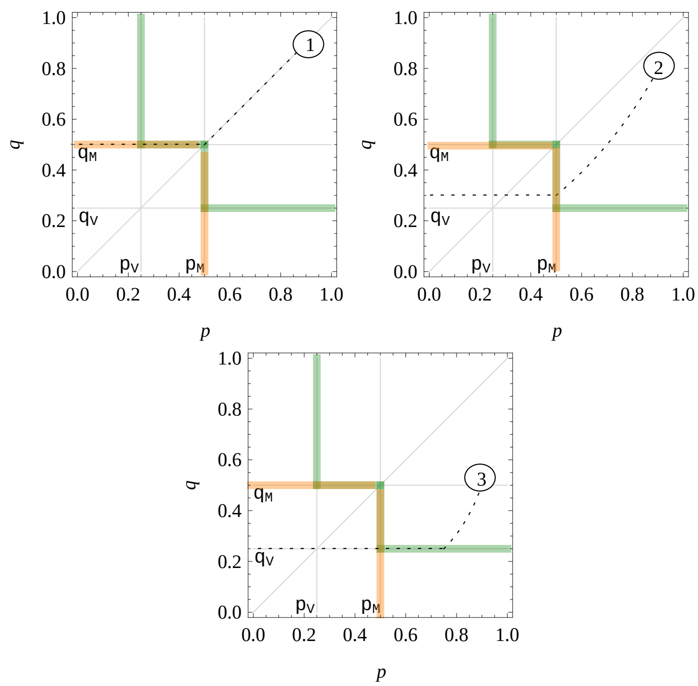


Figure 4: *In Case 1, the prior is* $p_0 = q_0 > p_M$. In Case 2, the priors $p_0 > q_0$ are such that at date $t > 0$ satisfying $p_t = p_M$ we have that $q_t > q_V$, while in Case 3 we have that $q_t \leq q_V$.

Case 1: If the prior is $p_0 = q_0 > p_M$, then in equilibrium both players switch to the safe option when beliefs reach the single-player myopic threshold belief, $p_M = \frac{a}{\lambda}$ and the safe option is allocated in a tie break (illustrated for player i winning the tie-break). At that point, both players are indifferent between activating their risky option and switching to the safe option, as long as their opponent switches with strictly positive probability. Because players cannot be indifferent between switching when beliefs are p_M and switching at a later date, both players switch at p_M with probability 1. Switching at an earlier date is strictly dominated.

In the planner solution, both players would only have switched to the safe option at $p_{\mathcal{W}} < p_V$. The extreme inefficiency here comes from the fact that competition from the other player is most intense when $p_0 = q_0$. As we will see in the next two cases, when one player is more pessimistic than the other, the inefficiency is mitigated.

Case 2: Here in equilibrium, player i uses the strategy whereby he continues activating his risky option for all $p \geq p_M$ and player j switches to the safe option with certainty when $p = p_M$. As long as player j switches with positive probability when $p = p_M$, player i is indifferent between playing R and S in that state, and has no incentive to preempt player j's exit.

Notice that player j, who is more pessimistic than player i, is allocated the safe option with certainty, and the belief about his risky option remains constant forever, while the belief about player i's risky option gradually decreases according to the law of motion for active options: $dp = -p\lambda(1-p)dt$.

The more pessimistic player j is relative to player i , the closer the exit belief of player j becomes to q_V, and the less inefficient the equilibrium. This is intuitive: if a player is more optimistic that another, he poses less of a threat to his opponent, who is then under less pressure to secure the safe option, and can experiment for longer.

Case 3: Here player i is so optimistic relative to player j that even when the belief about player j's risky option reaches the single-player threshold

q_V, the belief about player i's risky option is still above the myopic threshold belief p_M, and player i strictly prefers activating his risky option to switching to the safe option regardless of player j's action.

Player j then effectively plays a single-player game and switches to the safe option when $q = q_V$. The inefficiency is even lower than in the previous two cases, and as $p_0 \to 1$, the equilibrium tends to the planner solution.

Even though all equilibria described above are inefficient in the sense that there is less experimenting than in the planner solution, they are efficient in the sense that the safe option is always allocated to the most pessimistic player. The inefficiency of the level of experimentation is maximised when $p_0 = q_0$. In that case, the lost option-value to the optimist is the highest conditional on the exit date of the pessimist. In the two-player game, this intensifies competition and makes the pessimist exit earlier, while in the planner solution, the loss of the option-value is internalised by the planner who then postpones the exit of the pessimistic player.

3.3 Revocable Exit

In this section we assume that a player who is occupying the safe option may later return to the risky option. That is, the decision to switch to the safe option is revocable. We have seen in the previous section that when exit is irrevocable, the more pessimistic player, say player j, is the first to switch to the safe option in equilibrium and player i is forced to experiment with his risky option forever. If player i's experimenting results in a success, then for him switching to the safe option is dominated. Player j is then relieved of the threat of congestion and is de facto facing the single-player problem. If the state is such that $q > q_V$, player j would then like to return to his risky option and resume his experimenting.

While this is not possible with irrevocable exit, when exit is revocable this is the most desirable outcome for a player. So much so that in equilibrium players have incentives to temporarily interrupt their own experimentation and force their opponent to experiment with the sole aim of eliminating the threat of congestion. Let it be noted that there are no informational externalities to an opponent's success as the qualities of the players' risky options are independent.

Let us formally describe each player's problem. At each date, a player either chooses to activate his risky option (R) or the safe option (S) over the time interval $[t + dt)$. We assume that if a player switches to the safe option when it is already occupied by the opponent, the player "bounces" back to his risky option. Each player tries to maximise his expected discounted payoff. As in previous sections, the state at date t is summarised by the vector of posterior beliefs $(p_t, q_t) \in [0, 1]^2$.

As before, once a risky option has produced a success, the player occupying it never finds it optimal to switch to the safe option and the other player optimally plays as in the single-player game. We define a (Markovian) strategy $\bar{k}^i(.)$ for player i to be the mapping $\bar{k}^i : [0, 1)^2 \to [0, 1]$ from states (p_t, q_t) to $\bar{k}^i_t$, the probability that player i plays his risky option at t over the time interval $[t + dt)$. A (Markov-Perfect) equilibrium is a pair of strategies $(\bar{k}^i(.), \bar{k}^j(.))$ such that the strategy of player i maximises his expected discounted payoff conditional on the strategy of player j, and vice-versa.

Let $\mathsf{U}(.)$ denote the value function in the two-player game with revocable exit. Conditional on player j using the Markovian strategy $\bar{k}^j(.)$ player i's value function solves the dynamic problem:

$$\mathsf{U}(p,q;\bar{k}^j) = \max_{\bar{k}^i(p,q)\in[0,1]} \{\bar{k}^i(p,q)\ L^R\mathsf{U}(p,q;\bar{k}^j) + (1-\bar{k}^i(p,q))\ L^S\mathsf{U}(p,q;\bar{k}^j)\}$$

where

(3.2)

$$\begin{aligned} L^S\mathsf{U}(p,q;\bar{k}^j) :=\ & \left[1-(1-\bar{k}^j(p,q))(1-\iota)\right] \\ & \left(adt + e^{-\rho dt}[q\lambda dt\ V(p) + (1-q\lambda dt)\ \mathsf{U}(p,q';\bar{k}^j)]\right) \\ & +(1-\bar{k}^j(p,q))(1-\iota) \\ & \left(p\lambda dt\left(1+e^{-\rho dt}\tfrac{\lambda}{\rho}\right) + (1-p\lambda dt)e^{-\rho dt}\ \mathsf{U}(p',q;\bar{k}^j)\right) \end{aligned}$$

$$\begin{aligned} L^R\mathsf{U}(p,q;\bar{k}^j) :=\ & p\lambda dt\left(1+e^{-\rho dt}\tfrac{\lambda}{\rho}\right) \\ & +(1-p\lambda dt)\Big((1-\bar{k}^j(p,q))\ e^{-\rho dt}\ \mathsf{U}(p',q;\bar{k}^j) \\ & \quad +\bar{k}^j(p,q)\ e^{-\rho dt}\left[q\lambda dt\ V(p') + (1-q\lambda dt)\ \mathsf{U}(p',q';\bar{k}^j)\right]\Big), \end{aligned}$$

and with p', q' as defined in Chapter 2. The corresponding expressions hold for player j. Notice that for $\bar{k}^j = 1$, $L^R\mathsf{U}(p,q;1)$ solves the same differential equation as $L^R\mathsf{W}(p,q;1)$ in the previous section.

We derive the Markov Perfect Equilibrium of this game (Theorem 2). Disregarding equilibria in weakly dominated strategies, this equilibrium is unique in the two-player game with revocable exit. All proofs are relegated to the appendix and we concentrate on describing the equilibrium dynamics, drawing parallels with the equilibrium in Section 3.2 when pertinent. To this end, we first define some notation (Section 3.3.1) that will then serve to define the equilibrium strategies and illustrate the equilibrium dynamics (Section 3.3.2).

3.3.1 Notation

Let us first define the functions $S(.,.)$, $R_0(.,.)$ and $R_1(.,.,.)$. For each function, the first argument is the current belief about the quality of the risky option of the player to whom the payoff accrues. The second argument is the current belief about the quality of his opponent's risky option. For $(p,q) \in [0,1]^2$,

$$S(x,y) := \frac{a}{\rho} + y\ \frac{\lambda}{\lambda+\rho}\left[V(x) - \frac{a}{\rho}\right],$$

$$R_0(x,y) := x\ \frac{\lambda}{\rho},$$

$$R_1(x,y,\sigma) := x\ \frac{\lambda}{\rho}\ (1 - e^{-(\lambda+\rho)\sigma}) + \frac{a}{\rho}\ e^{-\rho\sigma}(1 - x + xe^{-\lambda\sigma}).$$

The function $S(.,.)$ denotes the utility of occupying the safe option until the opponent's experimenting produces a success and then to play as in the single-player game, collecting payoff $V(x)$. The first term, $\frac{a}{\rho}$, is the utility of having to play the safe option forever. The second term reflects the option-value of being able to adopt the optimal single-player behaviour should the opponent's experimenting prove successful. This is decreasing in y, the probability of the opponent's risky option being good.

There are two channels through which that option-value can be nullified. The first obtains if $y \to 0$ so that the opponent's experimenting never produces a success and the player occupying the safe option never gets the opportunity to behave as a single-player. The second obtains if x is below the single-player threshold belief, so that $V(x) = \frac{a}{\rho}$. In this case, even if the opponent's experimenting produces a success the player occupying the safe option does not return to his risky option.

The function $R_0(.,.)$ denotes the utility of being forced to experiment forever. This is the payoff accruing to a player if his opponent occupies the safe option forever or until the first player's experimenting is successful. The function $R_1(.,.,\sigma)$ denotes the utility of being forced to experiment for a duration of time σ before regaining access to the safe option and occupying it forever. This is the payoff accruing to a player if his opponent occupies the safe option and leaves it after a duration of time σ. In all cases, the belief about the quality of the risky option which is not being activated remains constant over time. It will become clear as we construct the equilibrium why those are the only payoffs we need to consider.

We now define boundaries in $[0,1]^2$ that will be relevant in describing the equilibrium strategies and illustrating the equilibrium dynamics. Let $B_0(.)$ denote the function that satisfies, for all $(x,y) \in [0,1]^2$,

$$x \leq B_0(y) \quad \Leftrightarrow \quad R_0(x,y) \leq S(x,y). \tag{3.3}$$

The set of states $\{(p,q) : p = B_0(q)\}$ is illustrated for $q \leq q_M$ in Figure 5. Notice that $B_0(q)$ is increasing in q and that $B_0(0) = p_M$.

In states (p,q) such that $q = B_0(p)$, player i is indifferent between being forced to experiment until successful, achieving the payoff $R_0(p,q)$, and forcing his opponent to experiment until successful, and achieving the payoff $S(p,q)$. We show in Appendix 3.4.3 that when $q \leq q_M$, if player j occupies the safe option, he never leaves it unless player i's experimenting produces a success. Player i therefore indeed faces the choice above when $q \leq q_M$ and trades off the payoffs $R_0(p,q)$ and $S(p,q)$.

At this point, let us highlight one striking feature of the equilibrium dynamics by making the naive assumption that a player never returns to his risky option unless his opponent's experimentation produces a success. Such a strategy seems in line with intuitions from the standard bandit model: once a player leaves his risky option, he never returns to it. Moreover, why would a player have stronger incentives to leave the safe option if his opponent's experimenting is unsuccessful? As the opponent becomes more pessimistic about the quality of his risky option, his demand for the safe option intensifies. The first player would then be more likely to permanently lose access to the safe option if he were to leave it than when he occupied it in the first place.

We find however, that in equilibrium, there are states in which the player occupying the safe option will eventually leave it even if he is certain that his opponent will then occupy it permanently. For $q \leq q_M$ consider any state $(\hat{p}, \hat{q})$ such that $S(\hat{p}, \hat{q}) > R_0(\hat{p}, \hat{q})$, and assume that player i is occupying the safe option - his preferred choice. As player j experiments unsuccessfully the common belief about the quality of his risky option decreases, while the belief about player i's risky option remains constant. In Figure 5, the state evolves towards the p-axis along a vertical trajectory. If the initial state $(\hat{p}, \hat{q})$ is such that $\hat{p} \leq p_M$, then the subsequent states never leave the set $\{(p,q) : S(p,q) \geq R_0(p,q)\}$ and player i indeed never leaves the safe otpion. If instead the initial state is such that $\hat{p} > p_M$, the subsequent states eventually fall into the set $\{(p,q) : S(p,q) < R_0(p,q)\}$, and player i prefers returning to his risky option, even if that means losing access to the safe option forever.

The crucial point is that $p > p_M$. Recall that for these beliefs we have that $p\frac{\lambda}{\rho} > \frac{a}{\rho}$, and player i prefers occupying his risky option forever to occupying the safe option forever. The ability to proceed as in the single-player game if the opponent is successful augments the payoff to choosing the safe option by $q\ \frac{\lambda}{\lambda+\rho}\left[V(p) - \frac{a}{\rho}\right] \geq 0$, making the safe option more attractive relative to the risky option than with irrevocable exit. For q sufficiently high, player i may therefore be willing to occupy the safe option in states in which he would prefer the risky option once and for all when exit is irrevocable. The additional term however decreases with q, the likelihood of the opponent's risky option being good, and eventually player i switches back to his risky option, and our naive assumption proves incorrect.

In states such that $q \in [0, q_M]$ and $p \in [B_0(0), B_0(q_M)]$, therefore, player j knows that player i will only *temporarily* force him to experiment. More precisely, if player i occupies the safe option in state (p,q) such that $q \in [0, q_M]$ and $p \in [B_0(0), B_0(q)]$, he will leave it after a time span σ_p^q of unsuccessful experimenting by player j, where σ_p^q satisfies

$$p = B_0\left(\frac{q\ e^{-\lambda\sigma_p^q}}{q\ e^{-\lambda\sigma_p^q} + 1 - q}\right).$$

Therefore in state (p,q) player j's expected payoff from being forced to experiment for the duration σ_p^q before being able to switch to the safe option is $R_1(q, p, \sigma_p^q)$.

Notice that the payoff to player i from forcing player j to experiment for the duration σ_p^q is :

$$\begin{aligned}(1-q+qe^{-\sigma_p^q\lambda})\left((1-e^{-\sigma_p^q\rho})\tfrac{a}{\rho}+e^{-\sigma_p^q\rho}\ p\tfrac{\lambda}{\rho}\right)\\ +q(1-e^{-\sigma_p^q\lambda})\tfrac{a}{\rho}\ +q\ \tfrac{\lambda}{\lambda+\rho}(1-e^{-\sigma_p^q(\rho+\lambda)})\left(V(p)-\tfrac{a}{\rho}\right),\end{aligned}$$

which simplifies to

$$\frac{a}{\rho}+q\ \frac{\lambda}{\lambda+\rho}\left[V(p)-\frac{a}{\rho}\right]=S(p,q),$$

the utility to player i of forcing j to experiment until he produces a success and then playing as in the single-player game. The intuition is simple: when in state $(p,B_0^{-1}(p))$ player i switches back to his risky option, he is indifferent between doing so and keeping the safe option, so his continuation utility at that date is equal to $S(p,B_0^{-1}(p))$.

As before, the subscripts M and V respectively denote the single-player myopic threshold and optimal threshold. Let $B_1(.)$ denote the function that satisfies, for all $(x,y)\in[0,1]\times[y_M,B_0(x_M)]$,

$$x\leq B_1(y)\quad\Leftrightarrow\quad R_1(x,y,\sigma_y^x)\leq S(x,y). \tag{3.4}$$

The set of states $\{(p,q): p=B_1(q)\}$ is illustrated for $q_M\leq q\leq B_0(p_V)$ in Figure 5.

Notice that for $x\searrow x_M$, where $x_M=\frac{a}{\rho}$ denotes the myopic threshold of the player being forced to experiment, the duration for which he is forced to experiment $\sigma_{x_M}^y\to\infty$ and $R_1(x_M,y,\sigma_{x_M}^y)\to R_0(x_M,y)$ so that the function

$$\begin{cases}R_0(x,y), & \text{if } x\leq x_M,\\ R_1(x,y,\sigma_x^y) & \text{if } x\geq x_M,\end{cases}$$

is continuous in x.

We are now ready to define, for $y \leq B_0(x_V)$,

$$B(y) = \begin{cases} B_0(y) & \text{if } 0 \leq x \leq x_M, \\ B_1(y) & \text{if } x_M \leq x \leq B_0(y_V). \end{cases} \tag{3.5}$$

The role $B(.)$ plays in the proof of theorem 2 is analogous to the one the myopic threshold belief plays when exit is revocable: in equilibrium, one player switches to the safe option in a state where his opponent is indifferent between also switching and pursuing his experimentation.

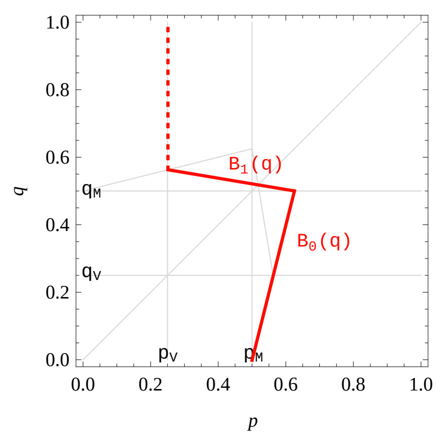


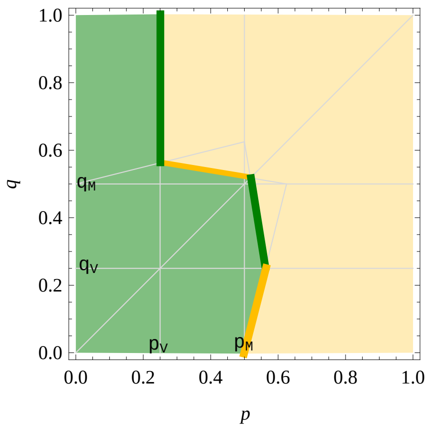
Figure 5: *Illustration of the boundary $B(q)$:*
For $q < q_M$, player i is better-off on the safe option than being forced to experiment until he produces a success whenever: $S(p,q) \geq R_0(p,q) \;\Leftrightarrow p \leq B_0(q)$.
For $q_M \leq q \leq B_0(p_V)$, player i is better-off on the safe option than being forced to experiment temporarily whenever: $S(p,q) \geq R_1(p,q) \;\Leftrightarrow p \leq B_1(q)$.

3.3.2 Equilibrium and Dynamics

We now derive the Markov Perfect Equilibrium of the two-player game with revocable exit (Theorem 2). Disregarding equilibria in weakly dominated strategies, this equilibrium is unique. All proofs are relegated to the appendix and we concentrate on describing the mechanics of the equilibrium, drawing parallels with the equilibrium in Section 3.2 when pertinent.

As a first step, (Appendix 3.4.3) we show that for the set of states (p, q) such that $p \leq B(q)$ and $q \leq B(p)$ both players have incentives to preempt one another's exit ($B(.)$ is defined in equation 3.5). This is relatively straightforward: for states in the above set, each player prefers occupying the safe option to being forced by his opponent to experiment, temporarily or permanently. Moreover, when his belief reaches the single-player threshold a player attempts to capture the safe option, at which point his opponent is better-off preempting his switch and the process unravels.

Theorem 2 describes the equilibrium strategies, illustrated below. The proof is relegated to the appendix, though we will illustrate it in this section by presenting typical equilibrium trajectories of the beliefs.



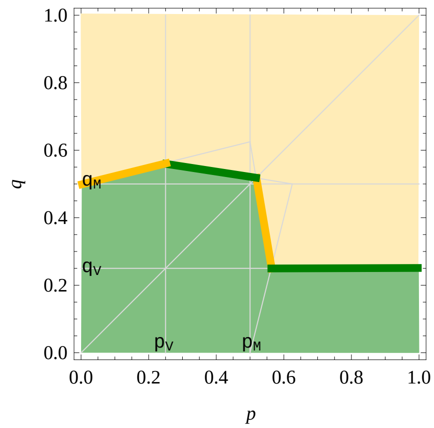


Figure 6: *Equilibrium strategies of player i and player j when exit is revocable. If a state (p, q) is in the green (dark) area, the player plays the safe option, if it is in the orange (light) area, the player plays his risky option.*

Theorem 2. *Consider the strategy profile illustrated above:*

$$\bar{k}^i(p,q) = \begin{cases} 0 & \text{if} \begin{cases} p \leq p_V, \\ p > p_V, \ p < B(q), \ q \leq B(p), \\ p = p_{\mathsf{U}}, q = q_{\mathsf{U}} \end{cases} \\ 1 & \text{else.} \end{cases}$$

$$\bar{k}^j(p,q) = \begin{cases} 0 & \text{if} \begin{cases} q \leq q_V, \\ q > q_V, \ q < B(p), \ p \leq B(q), \\ q = q_{\mathsf{U}}, p = p_{\mathsf{U}} \end{cases} \\ 1 & \text{else.} \end{cases}$$

where $q_{\mathsf{U}} = p_{\mathsf{U}}$ satisfy $B_1(p_{\mathsf{U}}) = B_1(q_{\mathsf{U}})$ This constitutes the unique MPE of the game (up to variations in weakly dominated strategies for histories in which the safe option has already been allocated, so that they do not affect the allocation of the objects, given an initial state).

Proof: See Appendix 3.4.3.□

We now illustrate the resulting equilibrium dynamics. The duration for which one player can force the other to experiment in equilibrium increases as competition intensifies (as priors get closer). In cases where priors are very different so that one player's risky option is much more likely to be of good quality, competition for the safe option is so low that the pessimistic player can play as in the single-player game. In all equilibria, if the player whose risky option is initially (at $t = 0$) least likely to be of good quality does not experiment successfully, he eventually gains access to the safe option.

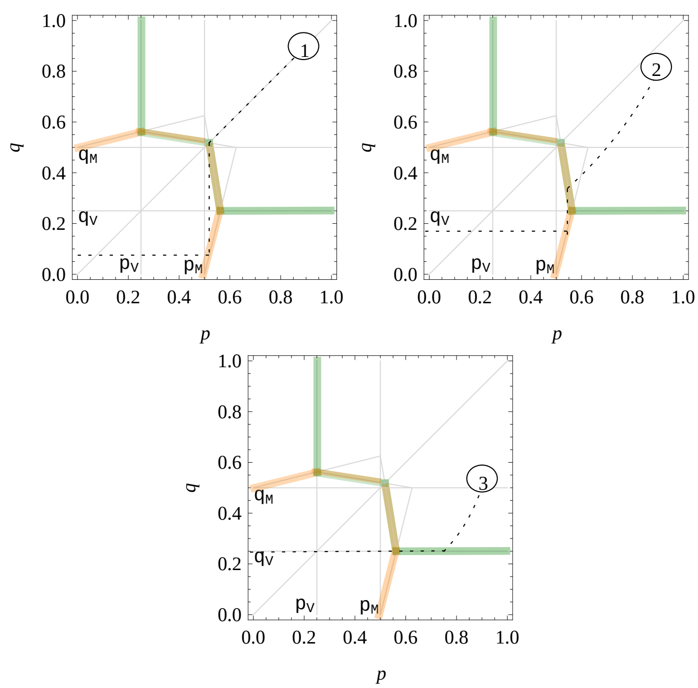


Figure 7: *In Case 1, the prior is* $p_0 = q_0 > p_M$. In Case 2, the priors $p_0 > q_0$ are such that at date $t > 0$ satisfying $q_t = B_1(p_t)$ we have $q_t > q_V$. In Case 3, the priors $p_0 > q_0$ are such that at date $t > 0$ satisfying $q_t = q_V$ we have $q_t \geq B_1(p_t)$.

Case 1: $p_0 = q_0 > \frac{a}{\rho}$**.** In equilibrium both players switch to the safe option when beliefs reach the state $p_{\mathsf{U}} = q_{\mathsf{U}}$ satisfying $B_1(p_{\mathsf{U}}) = B_1(q_{\mathsf{U}})$. At that point, both players are indifferent between being forced temporarily to activate their risky option and switching to the safe option, as long as their opponent switches to the safe option. Notice that $p_{\mathsf{U}} > p_M$, and that the player who is not allocated the safe option in the tie-break (here illustrated to be player j) will be forced to experiment unsuccessfully for longer than in any equilibrium with asymmetric priors.

Case 2: $p_0 > q_0$ **are such that at** $t > 0$ **satisfying** $q_t = B_1(p_t)$ **we have** $q_t > q_V$**.** Here in equilibrium, player j uses the right-continuous (in p) strategy whereby he continues activating his risky option for all $q \geq B_1(p)$ and player i switches to the safe option with certainty when $q = B_1(p)$. As long as player i switches with positive probability when $q = B_1(p)$, player j is indifferent between playing R and S in that state.

Notice that it is player i, the player most likely to experiment successfully, who is the first to capture the safe option, thus forcing player j to experiment. This is feasible because player i finds it optimal to eventually let player j occupy the safe option if his experimentation does not result in a success: If q_t falls too low, the prospect for player i of being able to achieve the single-player value vanishes, and he prefers resuming his own experimenting as his belief p is above the myopic threshold p_M. Notice also that in this equilibrium the player with the lowest prior is forced to experiment until his belief falls below the single-player optimal threshold.

Case 3: $p_0 > q_0$ are such that at $t > 0$ satisfying $q_t = q_V$ we have $q_t \geq B_1(p_t)$. Here player i is so optimistic relative to player j that even when the belief about player j's risky option reaches the single-player threshold q_V, the belief about player i's risky option is still above the boundary $B(q_V)$ and player i strictly prefers activating his risky option to switching to the safe option. Player j then effectively plays a single-player game and switches to the safe option when $q = q_V$. In this equilibrium, because of insufficient competition for the safe option, there is no alternating and it is the initially pessimistic player who captures the safe option once and for all.

3.4 Appendix Chapter 3

3.4.1 Lemma 6

Assume by way of contradiction that there exists and interval of time $[t, t + dt)$, $dt > 0$, on which player j plays S and R both with positive probability, and player i is indifferent between S and R. Then

$$\begin{aligned}
&L^R\mathsf{W}(p,q,k^j(p,q)) = \\
&\qquad p\lambda dt \left(1 + e^{-\rho dt}\tfrac{\lambda}{\rho}\right) \\
&\qquad +(1-p\lambda dt)\left((1-k^j(p,q))\; p' e^{-\rho dt}\tfrac{\lambda}{\rho}\right. \\
&\qquad\qquad \left. +k^j(p,q)\; e^{-\rho dt}\left[q\lambda dt V(p') + (1-q\lambda dt)L^R\mathsf{W}(p',q'^j(p',q'))\right]\right).
\end{aligned}$$

For $dt \to 0$ this condition becomes

$$L^R\mathsf{W}(p,q,k^j(p,q)) = (1-k^j(p,q))\; p\;\frac{\lambda}{\rho} + k^j(p,q)\; L^R\mathsf{W}(p,q,k^j(p,q))$$

For $k^j(p,q) \neq 1$ this holds if and only if $L^R\mathsf{W}(p,q,k^j(p,q)) = p\;\frac{\lambda}{\rho}$. Then

$$L^R\mathsf{W}(p,q,k^j(p,q)) = L^S\mathsf{W}(p,q,k^j(p,q)) \Leftrightarrow p = \frac{a}{\lambda}$$

The player is then only indifferent between his two actions when his belief is equal to he myopic belief, i.e. at one particular date, but not over an interval of time $dt > 0$. $\square$

3.4.2 Proof of Theorem 1

In what follows, fix an arbitrary initial state (p_0, q_0) such that $p_0 \geq q_0$, $p_M < p_0 < 1$, $q_M < q_0 < 1$. We will now derive an expression for the expected discounted utility of player i when both player i and j play their risky options from date $t = 0$ to date $t = \tau$ and player j exits at τ.

When both players play their risky options ($k_t^i = k_t^j = 1$), let $w(p, q)$ denote $L^R \mathsf{W}(p, q, 1)$, player i's utility from playing R, and $w_p(p, q)$, $w_q(p, q)$ its partial derivatives with respect to p and q respectively. $V(.)$ denotes the single-player value function. Simplifying Equation 3.1, $w(p, q)$ satisfies:

$$\begin{aligned}(p\lambda + q\lambda + \rho)\ w(p,q) + p\lambda(1-p)\ w_p(p,q) + q\lambda(1-q)\ w_q(p,q) \\ = \ p\lambda\ \tfrac{\lambda+\rho}{\lambda} + q\lambda\ V(p).\end{aligned}$$

Letting $\tilde{w}(s) := w(p(s), q(s))$, and noticing that $\frac{d\tilde{w}}{ds} = \frac{dp}{ds} w_p + \frac{dq}{ds} w_q$, we obtain the following ODE for $\tilde{w}(s)$:

$$\tilde{w}'(s) + f(s)\tilde{w}(s) = g(s),$$

with

$$\begin{aligned} f(s) &:= -(p\lambda + q\lambda + \rho), \\ g(s) &:= \ -p\lambda\ \tfrac{\lambda+\rho}{\lambda} - q\lambda\ V(p). \end{aligned}$$

Solving this ODE using definite integration, for $\tau \geq 0$, we obtain the solutions:

$$\tilde{w}(0;\tau) = \tilde{w}(\tau)\ e^{\int f(\tau) d\tau} - \int_0^\tau e^{\int f(s) ds}\ g(s)\ ds.$$

Solving explicitly, we obtain:

$$\begin{aligned} \tilde{w}(0;\tau) = \ & e^{-\rho\tau}(p_0 e^{-\lambda\tau} + 1 - p_0)(q_0 e^{-\lambda\tau} + 1 - q_0) \\ & \qquad \left[\tilde{w}(\tau) - p_\tau \tfrac{\lambda}{\rho} - K\ p_\tau q_\tau \left(\tfrac{1-p_\tau}{p_\tau}\right)^{\frac{\lambda+\rho}{\lambda}}\right] \\ & + p_0 \tfrac{\lambda}{\rho} + K\ p_0 q_0 \left(\tfrac{1-p_0}{p_0}\right)^{\frac{\lambda+\rho}{\lambda}}, \end{aligned}$$

with $K = \frac{a - \lambda p_V}{p_V \rho} \left(\frac{p_V}{1-p_V}\right)^{\frac{\lambda+\rho}{\lambda}}$ and p_V denoting the single-player optimal exit belief.

Because we assumed that $p_M < p_0 < 1$, $q_M < q_0 < 1$, $\tilde{w}(0)$ is a strictly increasing function of $\tilde{w}(\tau)$. If player j exits at date τ, then for some arbitrary $\Delta > 0$,

- if player i exits at $\tau + \Delta$, then $\tilde{w}(\tau) = p_\tau \frac{\lambda}{\rho}$,
- if player i exits at τ, then $\tilde{w}(\tau) = \iota \, \frac{a}{\rho} + (1 - \iota) \; p_\tau \frac{\lambda}{\rho}$,
- if player i exits at $\tau - \Delta$, then $\tilde{w}(\tau - \Delta) = \frac{a}{\rho}$ and in the limit, as $\Delta \to 0$, $\tilde{w}(\tau) \to \frac{a}{\rho}$

For $\iota \in (0, 1)$, the order of magnitude of these terms depends solely on the position of p_τ relative to the myopic exit belief, p_M. Player i is only indifferent between these three options when $p_\tau = p_M$.

When $p_\tau > p_M$, player i strictly prefers letting player j occupy the safe option and being stuck on his risky option forever, to occupying the safe option himself. So there can be no equilibrium in which player i switches to the safe option with certainty at τ' such that $p_{\tau'} > p_M$.

When $p_\tau < p_M$, player i is strictly better-off anticipating player j's move to the safe option, and letting the other player switch to the safe option is never a best response for player i on that support. There can therefore be no equilibrium[3] in which a player switches to the safe option with certainty in state (p, q) such that $p < p_M$, $q < q_M$, since the other player would respond by "undercutting" him.

[3] Unless the other player exits at a date such that player i is indifferent, or strictly prefers staying on his risky option. In that case, regardless of the exit date prescribed by his strategy, he never gains access to the safe option, so the allocation is not sensitive to his exit date. Having noticed this kind of multiplicity of equilibria, we henceforth only consider equilibria in strategies that are not weakly dominated.

Notice furthermore that the term

$$e^{-\rho\tau}(p_0 e^{-\lambda\tau}+1-p_0)(q_0 e^{-\lambda\tau}+1-q_0)\left[\frac{a}{\rho}-p_\tau\frac{\lambda}{\rho}-K\ p_\tau q_\tau\left(\frac{1-p_\tau}{p_\tau}\right)^{\frac{\lambda+\rho}{\lambda}}\right],$$

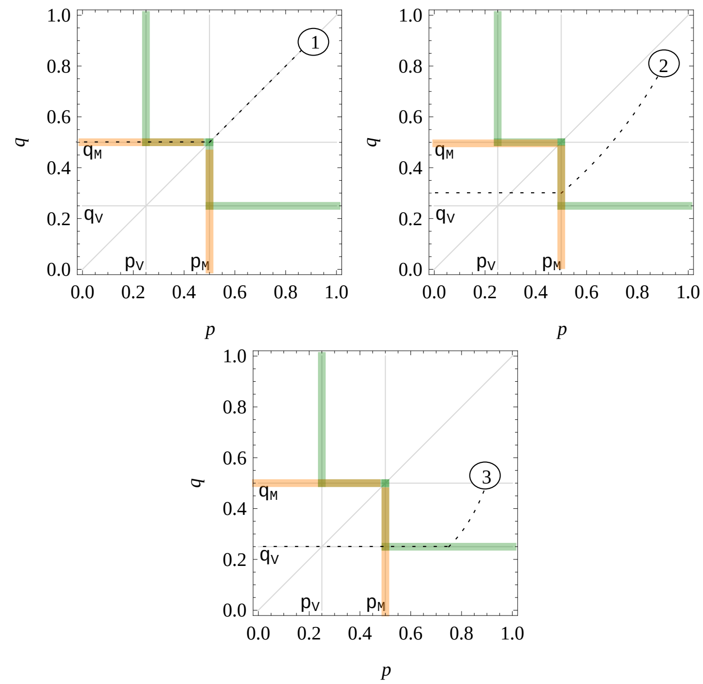
and therefore $\tilde{w}(0)$, are strictly increasing in τ for $p_\tau > p_V$ (they are maximised when $p_\tau = p_V$), i.e. for $p_\tau > p_V$, players gain from experimenting for longer.

One implication is that, conditional on exiting before player j, player i then maximises his utility with respect to his exit date. If $p_\tau < p_V$, player i optimally switches to the safe option at date $\tau' < \tau$ such that $p_{\tau'} = p_V$. If on the other hand $p_M > p_\tau \geq p_V$, then player i would like to exit at the latest possible date preceding player j's exit. In discrete time, this strategy would be unambiguous: player i would exit at date $\tau - 1$. In continuous time however, it only exists if player j's strategy is right-continuous[4] in p (left-continuous in time) so that an optimal exit date for player i does exist: $\max\{t \in \mathbb{R} : 0 \leq t \leq \tau\} = \tau$. If player j's strategy is left-continuous[5] in p (right-continuous in time), then player i always benefits from postponing his exit by some infinitesimal dt, and his optimal exit date, $\max\{t \in \mathbb{R} : 0 \leq t < \tau\}$, does not exist.

For the remainder of the argument we consider the three generic cases illustrated in the figures in section 3.2, and reproduced here.

[4] $k^i(p,q) = \begin{cases} 1 & \text{if } p \geq p_\tau \\ 0 & \text{if } p < p_\tau \end{cases}.$

[5] $k^i(p,q) = \begin{cases} 1 & \text{if } p > p_\tau \\ 0 & \text{if } p \leq p_\tau \end{cases}$



Case 1: $p_0 = q_0$**.** Following the arguments above, the only equilibrium is for both players to play their risky option when $p > p_M$ and to switch to the safe option in state $p = p_M$.

$$k^i(p,q)_{\text{Case 1}} = \begin{cases} 1 & \text{if } p > p_M \\ 0 & \text{if } p \leq p_M \end{cases}, \quad k^j(p,q)_{\text{Case 1}} = \begin{cases} 1 & \text{if } q > q_M \\ 0 & \text{if } q \leq q_M \end{cases}$$

They then face a tie-break in which either player is allocated the safe option with positive probability. If player i gains access to the safe option, the belief about his risky option remains p_M forever, while the belief about player j's risky option gradually decreases (all the way to zero, if the option is bad.)

Case 2: $p_0 \geq q_0$ **and such that when** $p_t = p_M$**,** $q_t > q_V$**.** Following the arguments above, player i will only optimally move to the safe option if his belief is p_M, and he is indifferent between being allocated either option forever. As noted above, for player j to have a best response, player i's strategy must be right-continuous in p.

Assume this were not the case and player i played S, then anticipating player i's switch by some positive time-interval $\Delta > 0$ would be a profitable deviation for player j. There would, however, be no best response (in continuous time), as player j would prefer anticipating player i's exit by $\frac{\Delta}{2}$ rather than Δ. In equilibrium, player i plays R when $p = p_M$ and player j is best-responding by switching to the safe option at $p = p_M$.

We therefore have that

$$k^i(p,q)_{\text{Case 2}} = \begin{cases} 1 & \text{if } p \geq p_M \\ 0 & \text{if } p < p_M \end{cases}, \quad k^j(p,q)_{\text{Case 2}} = \begin{cases} 1 & \text{if } p > p_M \\ 0 & \text{if } p \leq p_M \end{cases}$$

In that case, player j, who is more pessimistic than player i, is allocated the safe option with certainty, and the belief about his risky option remains constant forever, while the belief about player i's risky option gradually decreases (all the way to zero, if the option is bad.)

Case 3: $p_0 \geq q_0$ and such that when $p_t = p_M$, $q_t \leq q_V$. As in Case 2, and excluding strategies that are weakly dominated, player i will only optimally move to the safe option if his belief is p_M. This means that in states (p,q) such that $p \geq p_M$, $q \geq q_V$, player j essentially plays a single-player game, and he optimally switches to the safe option when $q = q_V$. Because the belief about player i's risky option is above the myopic player's exit belief, player i

finds it optimal to let player j occupy the safe option, and the equilibrium is

$$k^i(p,q)_{\text{CASE 3}} = \begin{cases} 1 & \text{if } p > p_M \\ 0 & \text{if } p \leq p_M \end{cases}, \quad k^j(p,q)_{\text{CASE 3}} = \begin{cases} 1 & \text{if } q > q_V \\ 0 & \text{if } q \leq q_V \end{cases}.$$

Arguing similarly for states $p_0 > q_0$, we complete the equilibrium strategies of both players, and establish the result of Theorem 1. $\square$

3.4.3 Proof of Theorem 2

We derive the MPE of the game with revocable exit. The proof will proceed as follows: we first derive an expression for the utility to player i from playing his risky option until some date τ at which player j exits. It is increasing in the continuation utility at date τ. We then show as a first step that to maximise this continuation utility agents will have incentives to preempt one another's exit for a set of states which we define in Section 3.4.3 below. As a second step we then fully characterise the agents' equilibrium best-response correspondences.

We will first derive an expression for the expected discounted utility of player i when both player i and j play their risky options from date $t = 0$ to date $t = \tau$ and player j exits at τ. Fix an arbitrary initial state (p_0, q_0) such that $p_0 \geq q_0$, $p_M \ll p_0 < 1$, $q_M \ll q_0 < 1$. Notice that for $\bar{k}^j = 1$, $L^R\mathsf{U}(p,q;1)$ solves the same differential equation as $L^R\mathsf{W}(p,q;1)$ in the previous appendix. We let $u(p,q)$ denote $L^R\mathsf{U}(p,q;1)$ and $\tilde{u}(s) := u(p(s), q(s))$. Replicating the solution from Appendix 3.4.2 we obtain:

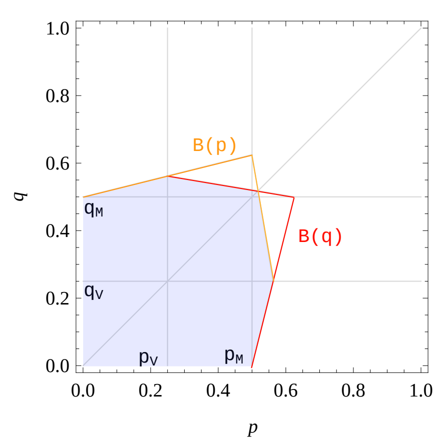
$$\tilde{u}(0;\tau) = e^{-\rho\tau}(p_0 e^{-\lambda\tau} + 1 - p_0)(q_0 e^{-\lambda\tau} + 1 - q_0) \left[\tilde{u}(\tau) - p_\tau \frac{\lambda}{\rho} - K\ p_\tau q_\tau \left(\frac{1-p_\tau}{p_\tau}\right)^{\frac{\lambda+\rho}{\lambda}}\right] + p_0 \frac{\lambda}{\rho} + K\ p_0 q_0 \left(\frac{1-p_0}{p_0}\right)^{\frac{\lambda+\rho}{\lambda}},$$

with $K = \frac{a-\lambda p_V}{p_V \rho}\left(\frac{p_V}{1-p_V}\right)^{\frac{\lambda+\rho}{\lambda}}$ and p_V denoting the single-player optimal exit belief.

Because we assumed that $p_M < p_0 < 1$, $q_M < q_0 < 1$, the utility to player i of staying on his risky option until player j switches to the safe option, $\tilde{u}(0,\tau)$, is a strictly increasing function of $\tilde{u}(\tau)$, the continuation utility at date τ.

Unraveling

As a first step, we compare continuation utilities to show that in states (p, q) such that $p < B(q)$ and $q < B(p)$ where B is defined in equation 3.5. Players will have incentives to preempt one another's exit and there will be unraveling of the exit decision. That set of states is depicted below.



We now compare continuation utilities to show that in states (p, q) such that $p < B(q)$ and $q < B(p)$ players will have incentives to preempt one another's exit, and there will be unraveling of the exit decision. If player j exits at date τ, then for some arbitrary $\Delta > 0$,

- if player i exits at date $\tau+\Delta$, then $\tilde{u}(\tau) = R_x(p(\tau), q(\tau))$ where x takes the value 0 or 1 when $q(\tau) \leq q_M$ or $\geq q_M$ respectively.
- if player i exits at date $\tau - \Delta$, then $\tilde{u}(\tau - \Delta) = S(p(\tau - \Delta), q(\tau - \Delta))$ and in the limit, as $\Delta \to 0$, $\tilde{u}(\tau) \to S(p(\tau), q(\tau))$,
- if player i exits a τ he faces a tie-break.

Similarly for player j when τ is player i's exit date. So a player is better-off anticipating his opponent's exit whenever $S(p(\tau), q(\tau)) > R_x(p(\tau), q(\tau))$. In the following, we drop the exit date τ and just concentrate on the states (p, q) to show that there will be unraveling of the exit decision.

In states $\{(p, q)|p \leq p_V\}$ switching to the safe option is (weakly) dominant for player i. This trivially follows from the single-player game. Similarly for player j in states $\{(p, q)|q \leq q_V\}$.

Consider the states $\{(p, q)|q \leq q_V, p \geq p_V\}$. If player i occupies his risky option, player j will occupy the safe option and stay on it forever, so the payoff to player i of occupying his risky option is $R_0(p, q) := p\frac{\lambda}{\rho}$. If player i occupies the safe option until player j's option produces a success, player i's payoff is $S(p, q) := \frac{a}{\rho} + q\ \frac{\lambda}{\lambda+\rho} \left[V(p) - \frac{a}{\rho}\right]$.

Player i's continuation utility is then maximised by also switching to the safe option as long as $S(p, q) > R_0(p, q) \ \Leftrightarrow\ p < B_0(q)$. Otherwise player i prefers being forced to experiment forever.

Consider the states $\{(p,q)|p_V \leq p \leq p_M, q_V \leq q \leq q_M\}$. Here the unraveling of the exit decision will start as players have incentives to preempt one another's exit: If player i switches to the safe option in state (p,q) with $p \searrow p_V$, player j's continuation payoff from staying on his risky option is $R_0(q,p)$ while his payoff from preempting player i's exit by some $\Delta > 0$ tends to $S(q,p)$ as $\Delta \to 0$ so that player j prefers preempting as long as $S(q,p) > R_0(q,p) \;\Leftrightarrow\; q < B_0(p)$. The converse argument holds for player i, establishing the unraveling in that set of states.

Consider the states $\{(p,q)|p_V \leq p \leq p_M, q_M \leq q \leq B_0(p_V)\}$. Player j has an incentive to preempt player i's exit as long as $S(q,p) > R_0(q,p) \;\Leftrightarrow q < B_0(p)$ even though, since $q \geq q_M$, player j will eventually return to his risky option if player i's belief falls too low. In that case player i has an incentive to preempt player j's exit as long as $S(p,q) > R_1(p,q) \;\Leftrightarrow\; p < B_1(q)$.

Similarly for player j in states $\{(p,q)|p_M \leq p \leq B_0(q_V), q_V \leq q \leq q_M\}$.

Finally consider the states $\{(p,q)|p_M \leq p \leq B_0(q_V), q_M \leq q \leq B_0(p_V)\}$. Here any player who occupies the safe option eventually leaves it if his opponent only produces unsuccessful trials. There is unraveling of the exit decision as long as $S(p,q) > R_1(p,q) \;\Leftrightarrow\; p < B_1(q)$ and $S(q,p) > R_1(q,p) \;\Leftrightarrow\; q < B_1(p)$.

Equilibrium

This series of steps in Section 3.4.3 establishes that there can be no equilibrium in which a player exits with certainty at date τ such that $(p(\tau), q(\tau))$

satisfy $p(\tau) < B(q(\tau))$ and $q(\tau) < B(p(\tau))$. We now argue that in equilibrium, there can be no exit at date τ such that $p(\tau) > \max(p_V, B_0(q(\tau)))$ and $q(\tau) > \max(q_V, B_0(p(\tau)))$: if player i exits at date τ satisfying the conditions above, then player j prefers letting i occupy the safe option than facing him in a tie-break or preempting him.

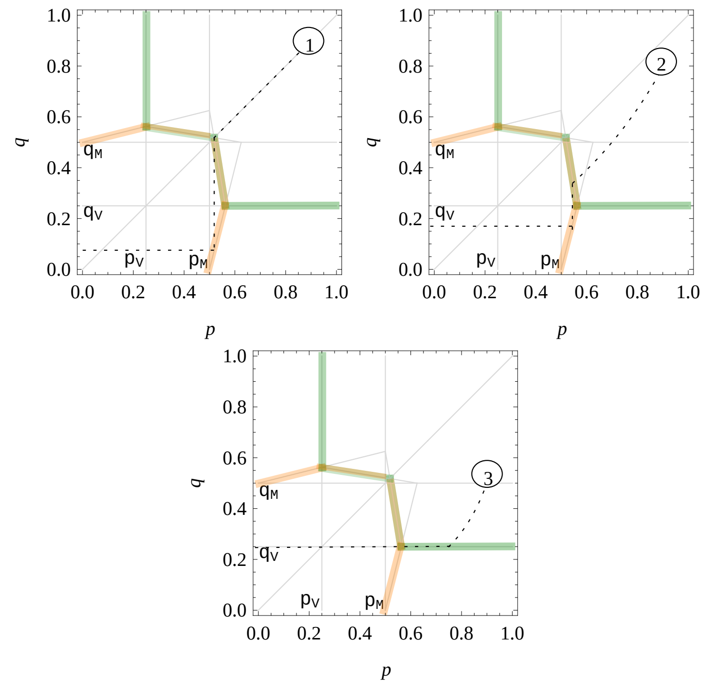
Notice furthermore that the term

$$e^{-\rho\tau}(p_0 e^{-\lambda\tau}+1-p_0)(q_0 e^{-\lambda\tau}+1-q_0)\left[S(p_\tau,q_\tau)-p_\tau\frac{\lambda}{\rho}-K\ p_\tau q_\tau\left(\frac{1-p_\tau}{p_\tau}\right)^{\frac{\lambda+\rho}{\lambda}}\right],$$

and therefore $\tilde{u}(0:\tau)$, are strictly increasing in τ for $p_\tau > p_V$ (they are maximised when $p_\tau = p_V$). Then because if player j were not preempting player i at date τ, player i would have an incentive to postpone his exit by some dt.

In fact, conditional on exiting before player j, player i aims to maximise his utility with respect to his exit date. If $p_\tau < p_V$, player i optimally switches to the safe option at date $\tau' < \tau$ such that $p_{\tau'} = p_V$. If on the other hand $p_\tau \geq p_V$, player i tries to exit as shortly as possible before player j. This maximisation only has a solution if player j's strategy is right-continuous in p, as explained in the previous appendix.

For the remainder of the argument we consider the three generic cases illustrated in the figures in Section 3.3.2, and reproduced here.



In all cases, following the argument in Section 3.4.3,

$$\bar{k}^i(p,q) = \begin{cases} 1 & \text{if } p \leq p_V \\ 1 & \text{if } q \leq q_V,\ p \leq B_0(q) \\ 0 & \text{if } q \leq q_V,\ p \geq B_0(q) \end{cases}, \quad \bar{k}^j(p,q) = \begin{cases} 1 & \text{if } q \leq q_V \\ 1 & \text{if } p \leq p_V,\ q \leq B_0(p) \\ 0 & \text{if } p \leq p_V,\ q \geq B_0(p) \end{cases}$$

Case 1: $p_0 = q_0$**.** Following the arguments above, the only equilibrium is for both players to play their risky option when $p > p_{\mathsf{U}}$ and to switch to the safe option in state $p = p_{\mathsf{U}}$.

$$\bar{k}^i(p,q)_{\text{Case 1}} = \begin{cases} 1 & \text{if } p > p_{\mathsf{U}} \\ 0 & \text{if } p \leq p_{\mathsf{U}} \end{cases}, \quad \bar{k}^j(p,q)_{\text{Case 1}} = \begin{cases} 1 & \text{if } q > q_{\mathsf{U}} \\ 0 & \text{if } q \leq q_{\mathsf{U}} \end{cases}$$

They then face a tie-break in which either player is allocated the safe option with positive probability. If player i gains access to the safe option, the belief about his risky option remains p_{U}. If player j's experimenting produces a success, player i immediately reverts to the single-player optimal strategy and achieves utility $V(p_{\mathsf{U}})$. If player j's experimenting remains unsuccessful after $\sigma_{p_{\mathsf{U}}}^{q_{\mathsf{U}}}$ periods player i prefers returning to his risky option, thus freeing the safe option of player j who then occupies it forever.

Case 2: $p_0 > q_0$ are such that at $t > 0$ satisfying $q_t = B_1(p_t)$ we have $q_t > q_V$. Following the arguments above, player i moves to the safe option in a state such that player j is indifferent between facing him in a tie-break or staying on his risky option and being forced to experiment temporarily. As noted above, for player i to have a best response, player j's strategy must be right-continuous in p. We therefore have that

$$\bar{k}^i(p,q)_{\text{Case 2}} = \begin{cases} 1 & \text{if } p > B_1(q) \\ 0 & \text{if } p \leq B_1(q) \end{cases} \quad , \quad \bar{k}^j(p,q)_{\text{Case 2}} = \begin{cases} 1 & \text{if } p \geq B_1(q) \\ 0 & \text{if } p < B_1(q) \end{cases}$$

In that case, player i, who is more optimistic than player j, is allocated the safe option with certainty. Then the game proceeds as in Case 1.

Case 3: $p_0 > q_0$ are such that at $t > 0$ satisfying $q_t = q_V$ we have $q_t \geq B_1(p_t)$. Here as long as $q_t \geq q_V$, $p_t \geq B(q_t)$ and player i has no incentive to occupy the safe option. Player j then essentially plays a single-player game and he optimally switches to the safe option when $q = q_V$ and player i finds it optimal to let player j occupy the safe option. The equilibrium, excluding weakly dominated strategies, requires

$$\bar{k}^i(p,q)_{\textsc{Case 3}} = \begin{cases} 1 & \text{if } p > B_0(q) \\ 0 & \text{if } p \leq B_0(q) \end{cases}, \quad \bar{k}^j(p,q)_{\textsc{Case 3}} = \begin{cases} 1 & \text{if } q > q_V \\ 0 & \text{if } q \leq q_V \end{cases}.$$

Then player j occupies the safe option with certainty and never switches back to his risky option. Arguing similarly for states $p_0 > q_0$, we complete the equilibrium strategies of both players, and establish the result of Theorem 2. $\square$

Chapter 4

Privately Observed Payoffs and Other Extensions

4.1 Introduction

The extension considered extensively in this chapter and the object of immediate future research addresses the concern that the assumption of publicly observed payoffs may be too strong. The following sections present some preliminary results on this issue. We concentrate on the stopping game ("irrevocable exit"), and hope to solve the game in which exit can be revoked in further research. In relaxing the assumption that payoffs are publicly observable, we need to re-define the Markov state to take into account that players now hold beliefs about whether the opponent's experimentation has as yet resulted in a success.

At the beginning of the game, neither player has had a chance to experiment yet, and the prior probabilities of either Poisson process having a positive arrival rate are common knowledge. As the game proceeds, a players may learn that his Poisson process has a positive arrival rate. He then has a dominant strategy which amounts to the single-player optimal behaviour described in Chapter 2. We attempt to solve for the equilibrium behaviour of a player who has yet to observe a Poisson event.

For any priors the players hold about the quality of their Poisson process, there only exists Markov-perfect equilibrium in pure strategies when their priors are so different that players never compete for the safe option, in the sense that playing his Poisson process is strictly dominant for one player when the other players switches to the safe option. When the priors are closer, and there is more intense competition for the safe option, we find that there is no pure strategy equilibrium.

We then set up the conditions for the existence of a mixed strategy equilibrium assuming that the randomising takes place over a connected support. We consider the case of players with equal priors about their Poisson processes and derive preliminary results about the form of the equilibrium. Fully characterising the equilibrium of the game remains the subject of ongoing research on our part. Let us point out that there is little existing work on experimentation with private monitoring. A few notable exceptions include Murto and Välimäki (2011) or Bonatti and Hörner (2011).

4.2 Private Payoffs

Consider the game with irrevocable exit from Chapter 3. Assume now that while players observe their opponent's actions, the outcome of their experimenting is private. In the previous sections, we assumed that this information was public. This meant that players held common posterior beliefs about the quality of both Poisson processes. Now, while a player sees whether his opponent is activating his Poisson process or not, he does not know whether this the Poisson process has already produced a success. So each player either faces an opponent whose Poisson process has already produced a Poisson event (we call this the "informed" type of opponent) or an opponent whose Poisson process has not yet produced a Poisson even (the "uninformed" type of opponent).

Each player now holds three beliefs:

1. A belief about the likelihood of his own Poisson process having a positive arrival rate. Let p and q denote player i and j's beliefs respectively.

2. A belief about the likelihood of his opponent's Poisson process having a positive arrival rate, conditional on the opponent not having observed a Poisson event yet. Let $\hat{q}$ and $\hat{p}$ denote player i and j's beliefs respectively.

3. A belief about the likelihood of his opponent's experimentation already having produced a Poisson event. Let θ and π denote player i and j's beliefs respectively. From the point of view of player i, player j's belief about his own Poisson process is 1 with probability θ and $\hat{p}$ with probability $1 - \theta$.

The vector of beliefs held by players i and j respectively at any date t

are $(p_t, \hat{q}_t, \theta_t)$ and $(q_t, \hat{p}_t, \pi_t)$. The vector of beliefs characterising the Markov state at any date t is

$$(p_t, \hat{q}_t, \theta_t; q_t, \hat{p}_t, \pi_t).$$

Recall that a denotes the flow payoff from playing the safe option, and each Poisson process has arrival rate $\lambda > 0$ or 0. Since we have assumed that $0 < a < \lambda$, the informed type has a dominant strategy, which is to always play his Poisson process. To complete the characterisation of the Markov Perfect Equilibrium in this game, we need to derive the equilibrium strategy of the uninformed type. Since in equilibrium the true belief p of the uninformed type is the same as his opponent's belief $\hat{p}$, the vector

$$(p, \theta; q, \pi)$$

with $p < 1$ and $q < 1$ is a sufficient summary statistic for the state when describing the uninformed player's problem. Without loss of generality we can think of a state as proceeding from an initial state $(p_0, \theta_0; q_0, \pi_0)$. We limit our analysis to Markov states proceeding from some initial state such that $\theta_0 = \pi_0 = 0$ and the vector of priors (p_0, q_0) is common knowledge.

A Markovian strategy μ^i for player i maps the beliefs (p, q) about the Poisson processes conditional on neither of them having produced a Poisson event, as well as his belief θ about the type of opponent he is facing, into $\mu^i(p, q, \theta)$, the probability that player i switches to the safe option in state $(p, \theta; q, \pi)$. Similarly for player j's strategy $\mu^j(q, p, \pi)$. The pair of strategies (μ^i, μ^j) for the uninformed types, together with the dominant strategies of the informed types, constitute a MPE if and only if the strategy of each uninformed type maximises his expected discounted utility in each state, given the strategies of both types of opponent.

Let $\mathcal{V}_i(.)$ denote player i's value function in this game. It solves the dynamic problem:

$$\begin{aligned}\mathcal{V}_i(p,q,\theta;\mu^j) = \max_{\mu^i(p,q,\theta)\in[0,1]} \big\{ &(1-\mu^i(p,q,\theta))L^R\mathcal{V}_i(p,q,\theta;\mu^j) \\ &+\mu^i(p,q,\theta)L^S\mathcal{V}_i(p,q,\theta;\mu^j)\big\},\end{aligned}$$

where

$$\begin{aligned}L^S\mathcal{V}_i(p,q,\theta;\mu^j) = \;& (1-\theta)\mu^j(q,p,\pi)\left(\iota\tfrac{a}{\rho} + (1-\iota)p\tfrac{\lambda}{\rho}\right) \\ & + (1-(1-\theta)\mu^j(q,p,\pi))\,\tfrac{a}{\rho}\end{aligned}$$

$$\begin{aligned}L^R\mathcal{V}_i(p,q,\theta;\mu^j) = \;& p\lambda dt\left(1+(1-\rho dt)\tfrac{\lambda}{\rho}\right) \\ & +(1-p\lambda dt)(1-\rho dt)\;(1-\theta)\mu^j(q,p,\pi)\;p'\tfrac{\lambda}{\rho} \\ & +(1-p\lambda dt)(1-\rho dt)\;(1-(1-\theta)\mu^j(q,p,\pi))\,\mathcal{V}_i(p',q',\theta';\mu^j).\end{aligned}$$

subject to $p' = p + dp$ and $\theta' = \theta + d\theta$, and where ι denotes the probability with which a tie is broken in favour of player i. Player j's value function is defined similarly.

4.2.1 Evolution of Beliefs

The state evolves in the following way:

1. The belief of the uninformed type of player i about his Poisson process having a positive arrival rate follows the law of motion:

   $$p_t + dp_t = p_t - p_t(1-p_t)\lambda dt \tag{4.1}$$

   Similarly for q_t.

2. Let θ_t be player i's belief at t about his opponent being the informed type conditional on player j never having switched to the safe option.

After a short time interval dt, player i updates this belief through two channels: he observes his opponent's action and draws inference based on his opponent's strategy μ^j which prescribes that player j exits with probability μ_t^j at date t. Moreover, he knows that his opponent could have observed a Poisson event during the short interval dt. By exiting the opponent reveals that he is uninformed and θ_{t+dt} jumps to zero. As long as he does not exit, θ_t follows the law of motion:

$$\theta_t + d\theta_t = 1 - \frac{(1-\theta_t)(1-\mu_t^j dt)}{1-(1-\theta_t)\mu_t^j dt}(1-q_t\lambda dt). \tag{4.2}$$

Solving for $d\theta_t$ and eliminating terms in $\mathcal{O}(dt^2)$ we obtain

$$d\theta_t = (1-\theta_t)\ [\theta\mu_t + q\lambda]\ dt. \tag{4.3}$$

Notice that if over the time interval $[t, t+s)$ the strategy μ_j prescribes that player j's uninformed type never exits, then

$$\theta_{t+s} = q_t(1-e^{-\lambda s}) = \frac{q_t - q_{t+s}}{1-q_{t+s}}.$$

4.2.2 No Pure Strategy Equilibrium

In this section we show that when there is sufficient competition between players (as defined in a precise sense) this game admits no pure strategy equilibrium. First we show that outside certain threshold beliefs, players have strictly dominant strategies (Lemma 7). We define initial states such that in all following states, playing his Poisson process is a strictly dominant strategy for one player, at least until switching to the safe option becomes strictly dominant for his opponent. In these states, there is a unique Markov Perfect Equilibrium in which the informed type of the player whose Poisson process is ex-ante least likely to have a positive arrival rate adopts the single-player optimal behaviour.

We further show that outside these states, players cannot exit one after the other in equilibrium (Lemma 8): one player always has strict incentives to deviate. The player meant to exit first can increase his payoff by inducing his opponent to believe that he is the informed type, simply by postponing his exit. The player meant to exit second may increase his payoff by preempting his opponent's exit. We finally show that there cannot be a pure strategy equilibrium in which players exit simultaneously either (Lemma 10): the safe option is then always allocated in a tie-break (with the unique tie-break rule defined in Lemma 9), and both players have an incentive to preempt their opponent's exit.

If in state $(p, \theta; q, \pi)$, the uninformed type of player j switches to the safe option, player i's continuation payoff from staying on his Poisson process in response is

$$L^R\mathcal{V}_i(p, q, \theta; \mu_t^j = 1) = \theta\ V(p) + (1-\theta)\ p\ \frac{\lambda}{\rho}. \tag{4.4}$$

If instead player i also switches to the safe option, he faces player j in a tie, and the tie-break rule allocates the option to player i with probability $\iota \in [0, 1]$. His payoff is

$$L^S\mathcal{V}_i(p, q, \theta; \mu_t^j = 1) = \theta\ \frac{a}{\rho} + (1-\theta)\left(\iota\ \frac{a}{\rho} + (1-\iota)\ p\ \frac{\lambda}{\rho}\right). \tag{4.5}$$

Finally, player i could preempt player j's switch, in which case, in the limit, his continuation payoff is $\frac{a}{\rho}$.

Denote by $p^*(\theta, \iota)$ the belief solving

$$p \geq p^*(\theta, \iota) \;\Leftrightarrow\; L^R \mathcal{V}_i(p, q, \theta; \mu_t^j = 1) \geq L^S \mathcal{V}_i(p, q, \theta; \mu_t^j = 1).$$

This means that for $p > p^*(\theta, \iota)$, switching to the safe option is a strictly dominated strategy for player i, when the tie-break rule is ι. We notice that $p^*(0, \iota) = p_M$, and $p^*(1, \iota) = p_V$. For $\theta > 0$, $p^*(\theta, \iota)$ solves

$$\left[1 + \iota\, \frac{1-\theta}{\theta}\right] \frac{a - p\lambda}{p\rho} \left(\frac{p}{1-p}\right)^{\frac{\lambda+\rho}{\lambda}} = \frac{a - p_V\lambda}{p_V\rho} \left(\frac{p_V}{1-p_V}\right)^{\frac{\lambda+\rho}{\lambda}}. \tag{4.6}$$

The right-hand side of Equation (4.6) is a constant. The left-hand side is decreasing in p on $[p_V, 1)$, increasing in ι and decreasing in θ. Therefore, $p^*(\theta, \iota)$ is unique, $\frac{\partial p^*}{\partial \iota} > 0$ and $\frac{\partial p^*}{\partial \theta} < 0$.

Assume that player j switches to the safe option in state $(p, \theta; q, \pi)$. For beliefs strictly above $p^*(\theta, \iota)$, player i strictly prefers letting the opponent occupy the safe option, while for beliefs strictly below this cutoff player i strictly prefers facing his opponent in a tie. The higher θ, the more optimistic player i is about player j being the informed type, and the lower his chances of losing access to the safe option if he decides not to switch when the uninformed type of player j does: the relative value of experimenting increases with θ. Similarly, the relative value of switching to the safe option and facing his opponent in a tie increases with ι for player i.

We define $q^*(\pi, \iota)$ in a similar fashion and notice that $q^*(0, \iota) = q_M$, $q^*(1, \iota) = q_V$, $\frac{\partial q^*}{\partial \pi} < 0$ and $\frac{\partial q^*}{\partial \iota} < 0$ (recall that ι is the probability that player i wins a tie).

We now show that for beliefs outside $[p^*(1, \iota), p^*(\theta, 1)]$, player i has a strictly dominant strategy.

Lemma 7. *In any state such that $p > p^*(\theta, 1)$, player i has a dominant strategy which is to play his Poisson process. In any state such that $p < p_V$, player i has a dominant strategy which is to switch to the safe option.*

Proof: In every state $(p, \theta; q, \pi)$, player i's continuation payoff is bounded from below by $\mathcal{V}_i(p, q, \theta; \mu_t^j = 1)$ (a player always prefers having access to the safe option to being forced to experiment by his opponent), and from above by $V(p)$, the single-player value-function. When $p > p^*(\theta, 1)$, $L^R\mathcal{V}_i(p, q, \theta; \mu_t^j = 1) > L^S\mathcal{V}_i(p, q, \theta; \mu_t^j)$ for all μ_t^j, even when $L^S\mathcal{V}_i(p, q, \theta; \mu_t^j)$ is maximised by the tie-break rule allocating the safe option to player i with certainty ($\iota = 1$). This establishes the first statement.

The second statement follows from the definition of p_V, the single-player optimal threshold belief introduced in Chapter 2. For beliefs $p < p_V$, the player has become too pessimistic about his Poisson process and sees no more value in experimenting.

The equivalent statements for player j are that: in any state such that $q > q^*(\pi, 0)$, player i has a dominant strategy which is to play his Poisson process, and in any state such that $q < q_V$, player j has a dominant strategy which is to switch to the safe option. $\square$

We conclude that in states $(p_t, \theta_t; q_t, \pi_t)$ such that $q_t = q_V$ and $p_t > p^*(\theta_t, 1)$, player j switching to the safe option and player i continuing to activate his Poisson process are best-responses to one another. Furthermore, neither player has an incentive to deviate by switching at an earlier date since $\forall s < t$, $p_s > p_t > p^*(\theta_t, 1)$ and $q_s > q_t - q_V$.

We would like to describe in terms of the initial state $(p_0, 0; q_0, 0)$ the states for which this strategy profile constitutes an equilibrium. To this end, notice that prior to the switching date t of player j, both players continuously activate their Poisson process. Therefore, by Equation (4.3) describing the law of motion of θ,

$$\theta_t = q_0(1 - e^{-\lambda t}) = \frac{q_0}{p_0}\frac{p_0 - p_t}{1 - p_t}.$$

Using this expression for θ in Equation (4.6), we can express the threshold belief at which player i is indifferent between facing his opponent in a tie and keep experimenting, in terms of (p_0, q_0) and denote it $\tilde{p}^*(p_0, q_0, \iota)$. It solves

$$\left[1 - \iota\left(1 - \frac{po}{qo}\,\frac{1-p}{po - p}\right)\right]\frac{a - p\lambda}{p\rho}\left(\frac{p}{1-p}\right)^{\frac{\lambda+\rho}{\lambda}} = \frac{a - p_V\lambda}{p_V\rho}\left(\frac{p_V}{1-p_V}\right)^{\frac{\lambda+\rho}{\lambda}},$$

For a given vector (p_0, q_0, ι), $\tilde{p}^*(p_0, q_0, \iota)$ is unique, increasing in ι and decreasing in p_0 and q_0 . We define $\tilde{q}^*(p_0, q_0, \iota)$ in a similar fashion.

Finally notice that as long as p and q follow the law of motion described in Equation (4.1) the ratio of their likelihood ratios is constant. Let the function $\Omega(x) = \frac{x}{1-x}$ for $x \in (0,1)$ denote the likelihood ratio. We are now ready to define the set of states for which the equilibrium described above exists.

Corrolary 1. *Given an initial state* $(p_0, 0; q_0, 0)$ *such that*

$$\frac{\Omega(p_0)}{\Omega(q_0)} \geq \frac{\Omega(\tilde{p}^*(p_0, q_0, 1))}{\Omega(q_V)}, \tag{4.7}$$

the strategy profile

$$\mu^i(p, q, \theta) = \begin{cases} 0 & \textit{if } p > p^*(\theta, 1), \\ 1 & \textit{else.} \end{cases}, \quad \mu^j(q, p, \pi) = \begin{cases} 0 & \textit{if } q > q_V, \\ 1 & \textit{else.} \end{cases}$$

constitutes the unique Markov Perfect equilibrium (up to variations in weakly dominated strategies). Similarly for states proceeding from an initial state $(p_0, 0; q_0, 0)$ *such that*

$$\frac{\Omega(p_0)}{\Omega(q_0)} \leq \frac{\Omega(p_V)}{\Omega(\tilde{q}^*(q_0, p_0, 0))}, \tag{4.8}$$

and the strategy profile

$$\mu^i(p, q, \theta) = \begin{cases} 0 & \text{if } p > p_V, \\ 1 & \text{else.} \end{cases}, \quad \mu^j(q, p, \pi) = \begin{cases} 0 & \text{if } q > q^*(\pi, 0), \\ 1 & \text{else.} \end{cases}$$

For the remainder of this section we concentrate on states proceeding from any initial state $(p_0, 0; q_0, 0)$ such that:

$$\frac{\Omega(p_0)}{\Omega(q_0)} \in \left(\frac{\Omega(p_V)}{\Omega(\tilde{q}^*(p_0, q_0, 0)}, \frac{\Omega(\tilde{p}^*(p_0, q_0, 1)}{\Omega(q_V)} \right) =: \Xi(p_0, q_0),$$

i.e. conditions (4.7) and (4.8) are not satisfied. Intuitively, these are states in which there is sufficient competition in the sense that the priors p_0 and q_0 are relatively close. We will show that in these states there is no equilibrium in pure strategies.

Lemma 8. *Given an initial state* $(p_0, 0; q_0, 0)$ *such that* $\frac{\Omega(p_0)}{\Omega(q_0)} \in \Xi(p_0, q_0)$, *there exists no equilibrium in pure strategies such that one player exits after the other.*

Proof: Assume there exists an equilibrium such that the uninformed type of player i exits at date t, and the uninformed type of player j exits at date $s > t$. The continuation payoff at date t to player i is $\frac{a}{\rho}$. If the informed type of player i deviates and does not exit at date t, then by equation (4.3) player j concludes that he is facing the informed type of player i, and reverts to playing the single-player optimal strategy: he will continue experimenting until his belief hits q_V, the single-player threshold. Player i can then just preempt that switch, and his payoff from this deviation is

$$p_t \frac{\lambda}{\rho}\left(1 - e^{-(\lambda+\rho)x}\right) + \left(1 - p_t(1 - e^{-\lambda x})\right) e^{-\rho x}\frac{a}{\rho},$$

where x solves $\frac{q_t \ e^{-\lambda x}}{q_t \ e^{-\lambda x} + 1 - q_t} = q_V$. As long as $x > 0$ this deviation is strictly profitable. This requires that $q_s > q_V$. By Lemma 7 q_s must belong to the interval $[q_V, q_M]$. If $q_s = q_V$ then player i's best response is to exit at t just prior to s, so that $q_t \searrow q_V$. $\square$

So our candidate equilibrium strategies are such that players exit simultaneously. We show that for every initial state, there exists a unique tie-break rule such that players are simultaneously indifferent between switching to the safe option or not. We then show that players always have profitable deviations, and conclude that there are no equilibria in pure strategies.

Lemma 9. *Given an initial state* $(p_0, 0; q_0, 0)$ *there exists a unique tie-break rule* $\iota^*(p_0, q_0)$ *such that both players are simultaneously indifferent.*

Proof: Fix an initial state $(p_0, 0; q_0, 0)$. For every $\iota \in (0, 1)$ there exists a unique pair $(\tilde{p}^*(p_0, q_0, \iota), \tilde{q}^*(q_0, p_0, \iota))$ satisfying the indifference conditions of both players. For these beliefs to be held by both players *simultaneously*, the ratio of their likelihood ratios must equal that of the priors. (Recall that the ratio of likelihood ratios remains constant over time.) Since $\frac{\partial}{\partial\iota}\tilde{p}^*(p_0, q_0, \iota) > 0$ and $\frac{\partial}{\partial\iota}\tilde{q}^*(p_0, q_0, \iota) < 0$, the ratio $\frac{\Omega(\tilde{p}^*(p_0,q_0,\iota))}{\Omega(\tilde{q}^*(p_0,q_0,\iota))}$ is a strictly increasing function of ι on $\left(\frac{\Omega(\tilde{p}^*(p_0,q_0,0))}{\Omega(\tilde{q}^*(p_0,q_0,0))}, \frac{\Omega(\tilde{p}^*(p_0,q_0,1))}{\Omega(\tilde{q}^*(p_0,q_0,1))}\right) = \Xi(p_0, q_0)$. We conclude that there exists a unique $\iota^*(p_0, q_0) \in (0, 1)$ such that

$$\frac{\Omega(\tilde{p}^*(p_0, q_0, \iota^*(p_0, q_0)))}{\Omega(\tilde{q}^*(p_0, q_0, \iota^*(p_0, q_0)))} = \frac{\Omega(p_0)}{\Omega(q_0)}. \quad \square$$

The existence of a unique tie-break rule such that both players are simultaneously indifferent between switching or not is evocative of Simon and Zame's inclusion of the tie-break rule into the equilibrium in games with payoff-discontinuities (Simon and Zame (1990)) - such as Bertand or Hotelling games. Whereas with public payoffs (cf. previous chapters) we could solve this difficulty by letting one player have a left-continuous, the other a right-continuous strategy, here this does not work because of discontinuities between payoffs when players exit together, or when players exit one after the other. Indeed, the proposed pure strategy equilibrium breaks down altogether.

Lemma 10. *Preempting opponent's switch is profitable deviation from switching simultaneously and facing tie-break.*

Proof: For $\iota^*(p_0, q_0) \in (0, 1)$, $p^*(p_0, q_0, \iota^*(p_0, q_0)) < p^*(p_0, q_0, 1)$ and so

$$LS\mathcal{V}_i(p^*(p_0, q_0, \iota^*(p_0, q_0)), q^*(p_0, q_0, \iota^*(p_0, q_0)), \mu_t^j = 1) < \frac{a}{\rho}.$$

The payoff from exiting simultaneously with his opponent is lower than the payoff from anticipating the opponent's switch by some small time interval $dt \to 0$. $\square$

We conclude that for states proceeding from initial states $(p_0, 0; q_0, 0)$ such that $\frac{\Omega(p_0)}{\Omega(q_0}} \in \Xi(p_0, q_0)$, there exists no pure strategy equilibrium.

4.2.3 Mixed Strategy

As a first step in exploring the existence of mixed strategy equilibria, consider games in which players have the same prior probability of having a Poisson process with positive arrival rate. The initial state $(p_0, 0; p_0, 0)$ is then summarised by p_0 and given an initial state and a strategy for his opponent, a player's belief about his opponent's type is given by Equation (4.3), so that the posterior belief $p < 1$ is a sufficient summary statistic for each state in which neither player has observed a Poisson event yet.

Consider the symmetric strategy profile $\mu_i(p) = \mu_j(p)$ where $\mu_i(p)$ denotes the probability with which player i's uninformed type switches to the safe option in state p. Player i's value function $\mathcal{V}(p; \mu_j)$ then solves the following dynamic problem :

$$\mathcal{V}(p; \mu_j) = \max_{\mu_i(p)\in[0,1]} \left\{(1-\mu_i(p))L^R\mathcal{V}(p;\mu_j) + \mu_i(p)L^S\mathcal{V}(p;\mu_j)\right\}$$

$$\begin{aligned}
L^S\mathcal{V}(p;\mu_j(p)) &= (1-\theta)\mu_j(p)\left(\iota\tfrac{a}{\rho} + (1-\iota)p\tfrac{\lambda}{\rho}\right) + (1-(1-\theta)\mu_j(p))\,\tfrac{a}{\rho} \\
L^R\mathcal{V}(p;\mu_j(p)) &= p\lambda dt\left(1 + (1-\rho dt)\tfrac{\lambda}{\rho}\right) \\
&\quad + (1-p\lambda dt)(1-\rho dt)\ (1-\theta)\mu_j(p)\ p'\tfrac{\lambda}{\rho} \\
&\quad + (1-p\lambda dt)(1-\rho dt)\ \ (1-(1-\theta)\mu_j(p))\,\mathcal{V}(p';\mu(p')),
\end{aligned}$$

where $p' = -p\lambda(1-p)dt$.

Conjecture 1. *There is a symmetric equilibrium in mixed strategies in which players exit with positive probability over some time interval* $[\underline{s}, \bar{s}]$ *such that* $p_V \le p_{\bar{s}} < p_{\underline{S}} \le p^*(\theta, 1)$.

Let $h(p_t)$ denote the hazard rate at which an uninformed player switches to the safe option over the time interval $[t, t+dt)$ conditional on not having switched yet. If we assume that $h(p_t)$ is continuous over the support $[\underline{p}, \bar{p}] = [p_{\bar{s}}, p_{\underline{S}}] \subseteq [p_V, p^*(\theta, 1)]$, then ties occur with zero probability. Hence $h(p_t)dt \equiv \mu(p_t)$ solves

$$\begin{aligned}\tfrac{a}{\rho} &= p_t\lambda dt\left(1 + (1-\rho dt)\tfrac{\lambda}{\rho}\right)\\ &+(1-p_t\lambda dt)(1-\rho dt)\ (1-\theta(p_t))h(p_t)dt\ p_{t+dt}\tfrac{\lambda}{\rho}\\ &+(1-p_t\lambda dt)(1-\rho dt)\ \ (1-(1-\theta(p_t))h(p_t)dt)\ \tfrac{a}{\rho}.\end{aligned}$$

Eliminating terms $\in \mathcal{O}(dt)$ and simplifying, we obtain that a player is indifferent between switching to the safe option and activating his Poisson process over the short time interval $[t, t+dt)$ whenever the expected hazard rate of exit by his opponent in state $p(t)$ satisfies

$$(1-\theta(p))h(p) = \tfrac{p\lambda(\lambda+\rho-a)-a\rho}{a-p\lambda} = (\lambda+\rho-a)\tfrac{p-p_V}{p_M-p}. \tag{4.9}$$

Expressing the law of motion of θ_t described in Equation (4.3) as a function of the state p rather than of time, we obtain the following ODE for the belief θ:

$$-p\lambda(1-p)\frac{d\theta(p)}{dp} = \theta(p)\ [(1-\theta(p))h(p) - p\lambda] + p\lambda\ . \tag{4.10}$$

Combining the two ODEs gives us the trajectory of beliefs when players randomise:

$$-p\lambda(1-p)\frac{d\theta(p)}{dp} = \theta(p)\ [\frac{p\lambda(\lambda+\rho-a)-a\rho}{a-p\lambda} - p\lambda] + p\lambda\ . \tag{4.11}$$

Solving this ODE gives a functional form for $\theta(p)$. Given this functional form, we could use Equation (4.9) and solve explicitly for $h(p)$, the hazard rate at which an uninformed player switches to the safe option.

4.3 Other Extensions

In this section we present other possible extensions to the model which we deem useful and of economic interest, though they are not analysed as part of this thesis. Although the results from Chapter 3, especially for the case where exit is revocable, can seem intuitive, it may be difficult to think of a great number of 'real-world' illustrations of the equilibrium behaviour. While examples of preemption abound, we may have to engage in loose interpretation when trying to think of situations in which the player currently forcing his opponent to experiment returns to his own experimentation and leaves the safe option for his opponent to take, after having exhausted the strategic option-value from occupying the safe option.

Certainly the assumption that the contested option is "safe" exacerbates the exciting result that preemption need not be irreversible. We could understand the notion of "occupying the safe option" as the protracted contract negotiation of a job-candidate aware that his competitor, who is himself hoping for an offer from that employer, might seek a job elsewhere. Recognising that there may be a *strategic* option-value associated with occupying an option also raises a question of *identification*.

Consider the situation along the equilibrium path in which one player occupies the safe option and forces his opponent to experiment. The player occupying the safe option does so in the hope of his opponent's experimentation producing a success, and even though his belief is close to the the myopic threshold: absent the opponent, the player would rather experiment with his own Poisson process. Similarly for the player being forced to experiment, whose belief may be below the myopic threshold: he prefers the safe option.

An outside observer who is not aware that the players are competing for access to the safe option and who does not know the players' beliefs might wrongly conclude that the player occupying the safe option prefers it to his Poisson process, and that the player occupying his Poisson process prefers it to the safe option. The outside observer would erroneously ignore the strategic option-value attached to occupying the safe option, and miss the possibility that the player occupying the safe option is currently forcing the opponent to experiment.

With view to approaching a more "realistic" setup, several extensions might be worth pursuing. Allowing for the options to be priced is one such example. Notice that the Planner Solution when exit is revocable could be implemented by the planner renting out access to the safe option over some short time-interval $dt \to 0$ by means of a second-price auction. Players then bid the difference in their continuation utilities with or without the option for that time interval, and the safe option would always be allocated to the player with the lowest expected arrival-rate.

Thinking of the options as employers setting wages at which they hire workers, it is also clear that a success by his opponent makes a player the monopolist vis-à-vis the safe option. How such a sudden shift in bargaining-power would affect the worker's wage is not clear: the firm could wait until the player has experimented and risk him never wanting the safe job should his experimentation result in a success. If however the player's Poisson process never produces a Poisson event, the firm can now hire the worker very cheaply.

Even without prices, we can think of options as strategic players deriving utility from being activated by a player and so get closer to a two-sided matching setup. It is clear from the results in Chapter 3 that an option benefits from competition, since players have incentives to occupy an option that they would not occupy in the absence of competition, so as to block the other player. From the point of view of an option, it might therefore be optimal to be the "second choice" for many players rather than the "first choice" for a few players.

With view to deriving the model's implications in a larger economy, one further extension one could consider would be to increase the number of players. In this case, even though players still benefit from an opponent being forced to experiment, each player would like another player to block the safe option, rather than occupying the safe option himself: there is an incentive to free-ride. At the same time, occupying the single safe option forces several players to experiment, making the strategy more appealing. Depending on how the two effects trade each other off, the set of beliefs for which players switch to the safe option in equilibrium could increase or decrease, although we conjecture that the first effect will dominate. Even if a player forces several players to experiment, the probability of them *all* having a success, and the player achieving the single-player value, is minute. The equilibrium behaviour in the game with revocable exit would gradually evolve towards the equilibrium behaviour in the game with irrevocable exit, with the relevant threshold beliefs gradually approaching the myopic threshold.

However, in letting the number of players increase, one must beware not to run into a motivational conflict: recall that we are assuming that payoffs are public, and players are aware of their competition. When the number of players increases, these assumptions may become indefensible.

Chapter 5

N-dimensional Blotto Game with Asymmetric Battlefield Values

5.1 Introduction

Budget-constrained multidimensional allocation problems were amongst the very first ones considered in game theory. The first version can be found in Borel and Ville (1938). This problem and similar ones later came to be known as "Colonel Blotto" games, after Gross and Wagner (1950)'s approach to the allocation problem.

In the simplest version of the Colonel Blotto game, two generals want to capture three equally valued battlefields. Each general disposes of one divisible unit of military resources. The generals have to simultaneously allocate these resources among the three battlefields. A battlefield is captured by a general if he allocates more resources there than his opponent. The goal of each general is to maximise the number of captured battlefields.

In that game, a pure strategy for player i is a 3-dimensional allocation vector $\mathbf{x}_i = (x^i_1, x^i_2, x^i_3)$ where x^i_k is the amount of resources allocated to the kth battlefield. The set of all pure strategies is the 2-dimensional simplex Δ^2. A mixed strategy is a trivariate distributions function $F : \Delta^2 \to [0, 1]$.

This version of the game was considered in Borel's course on probability Borel and Ville (1938) at the university of Paris in 1936-37. The solutions given by Borel reappear in Gross's and Wagner's unpublished research memorandum (1950).

They state that a mixed strategy F constitutes a symmetric equilibrium of the game if all one-dimensional margins of F are uniform over $[0, \frac{2}{3}]$. One geometrical approach to building such a distribution F consist of projecting a sphere, together with a uniform generic point belonging to its surface, onto the disc inscribed in an equilateral triangle.

Gross and Wagner conjecture that this geometrical method of generating the equilibrium distribution extends to Colonel Blotto games with more than three equally valued battlefields. This extension is formalised in Laslier and Picard (2002). It is worth noting that Weinstein (2005) presents a different geometric approach for case of $n \geq 3$ equally valued battlefields.

Roberson (2006) addresses the question of whether the univariate marginal distributions of the equilibrium strategies (n-variate distributions) are necessarily uniform for symmetric battlefield weights but possibly asymmetric budgets, and finds that they indeed have to be. That paper does not, however, solve the Blotto Game with asymmetric battlefield values. Another related paper is Kvasov (2007). It looks at a variation of the Blotto Game in which the allocation of resources is costly, and there too, battlefields are symmetric.

The present chapter generalises Gross and Wagner's geometric approach to construct equilibrium distributions of the n-dimensional Colonel Blotto game with asymmetric battlefield weights. The difficulty lies in inscribing a circle within an irregular n-gon. The necessary and sufficient conditions for this relate to the integer partitioning problem, a well-known problem of combinatorial optimisation.

The next section describes the model, then generalises the proofs of the existing literature to describe known equilibria of this game. Section 5.3 presents geometrical methods of constructing equilibrium distributions. It describes Borel's solutions as formulated in Gross and Wagner (1950), then Laslier and Picard's geometric construction method. Section 5.4 constitutes the main contribution of this chapter. It shows how, and under which conditions, this method can be extended to asymmetric n-dimensional cases. The conditions are related to a constrained version of the NP-complete "integer partitioning problem".

We end this chapter (Section 5.5) by illustrating the construction method using the example of US presidential elections. We argue that given the motivations of presidential candidates, the Colonel Blotto game is an apt model. Moreover, it turns out that in that example and given the actual distributions of electoral votes across US states, the construction method suggested performs very well at generating equilibrium distributions. The final section concludes.

5.2 Model and Equilibrium

Two players with identical budget normalised to one decide how to allocate their resources across n battlefields indexed by $k \in \{1, ..., n\}$. The absolute value of battlefield k is the positive integer[1] E_k. For all k, denote $e_k = E_k / \sum_{k=1}^n E_k$ the relative value of battlefield k and note that $\sum_{k=1}^n e_k = 1$. To make the game non-trivial, assume that $0 < e_k < 1/2$, or equivalently that $0 < E_k < \sum_{j \neq k} E_j$, for all $k = 1, ..., n$.

Player $i \in \{1, 2\}$ chooses a nonnegative vector of allocations $\mathbf{x}_i = (x_1^i, ..., x_n^i)$ where x_k^i is the amount of resources allocated to battlefield k. Player i wins in battlefield k if his resources in that battlefield, x_k^i, exceeds the resources x_k^j of the other player. Ties are resolved by flipping a coin. Both players are budget-constrained so the sum of a player's resources allocated across all battlefields cannot exceed that player's budget of 1.

A pure strategy of player i is an n-dimensional vector $\mathbf{x}^i$ satisfying the budget constraint. Denote $\mathcal{S}^i$ the set of pure strategies of player i:

$$\mathcal{S}^i = \Big\{\mathbf{x} \in [0, 1]^n : \sum_{k=1}^n x_k \leq 1\Big\}$$

Both players seek to maximise the aggregate value of captured battlefields. The function $g : \mathcal{S}^i \times \mathcal{S}^j \to \mathbb{R}$ measures the excess aggregate value of bat-

[1] Notice that we could choose battlefield weights in $\mathbb{R}$. But because we will relate this problem to the integer partitioning problem (and for real numbers, condition (P1E') introduced in Section 5.4.3 holds with zero probability), we restrict attention to integers from the outset.

tlefields captured by player i if he plays the pure strategy $\mathbf{x}^i$ while player j plays $\mathbf{x}^j$:

$$g(\mathbf{x}^i, \mathbf{x}^j) = \sum_{k=1}^{n} e_k \; sgn(x_k^i - x_k^j),$$

with $sgn(u) = 1$ if $u > 0$, 0 if $u = 0$ and -1 if $u < 0$.

A mixed strategy of player i is an n-variate joint distribution function $F^i : S^i \to [0, 1]$. Denote F_k^i the kth one-dimensional margin of F^i, i.e. the unconditional distribution of x_k^i. For each $k = 1, ..., n$, F_k^i maps $[0, 1]$ into itself. Define the payoff to a mixed strategy as the mathematical expectation of $g(\mathbf{x}^i, \mathbf{x}^j)$ with respect to the strategy profile (F^i, F^j).

The following proposition generalises existing results on the form of equilibria in Blotto games to the case of asymmetric battlefield weights. The proof is relegated to Appendix 5.7.1.

Theorem 3. *Consider the Colonel Blotto Game with asymmetric battlefield weights.*

(i) This game has no pure strategy Nash equilibrium

(ii) Both players meet their resource constraint in equilibrium.

(iii) Let F^* be a probability distribution of $\boldsymbol{x} \in \Delta^{n-1}$ such that each vector coordinate x_k ($k = 1, ..., n$) is uniformly distributed on $[0, 2e_k]$. Then (F^*, F^*) constitutes a symmetric Nash equilibrium.

The first point implies that an equilibrium, if it exists, must be in mixed strategies. The second point guarantees that the support of any equilibrium strategy is the $(n-1)$-dimensional simplex. Point three states that having univariate margins that are uniform on $[0, 2e_k]$ is a sufficient condition for a mixed strategy with support Δ^{n-1} to constitute a symmetric Nash equilibrium. Roberson (2006) shows that for homogeneous battlefield values ($\forall k\ e_k = 1/n$) uniform univariate margins are also a *necessary* condition for equilibrium.

Is it always possible to build a joint distribution satisfying the properties of F^*? We cannot answer this question in general, but we provide one method for building these equilibrium distributions. We then present conditions under which this method works, and address the question of when these conditions are likely to be satisfied. To this end we note a parallel to the constrained integer partitioning problem.

The following section describes the geometric construction method of Gross and Wagner, and later Laslier and Picard, while section 5.4 generalises it to accommodate asymmetric battlefield values. We obtain conditions under which this construction method always produces a joint distribution satisfying the properties of F^*.

5.3 Multivariate Distributions - Known Cases

The aim is to construct a n-variate distribution function F^* from given one-dimensional margins and given the equilibrium restrictions on the support of F^*. Indeed, in equilibrium candidates only use strategies in the $(n-1)$-dimensional simplex, Δ^{n-1}, which does not include the whole of $\times_{k=1}^{n}[0, 2e_k]$. Were it otherwise, it would be possible to construct a joint distribution with any correlation properties.

So the restriction of the support of F^* given its margins limits the number of possible interactions between resource allocations to different battlefields. So far, I have not been able to fully characterise the set of possible correlations satisfying the restrictions on F^*.

This section presents a geometrical method of constructing F^* that we will refer to as the *generalised disk solution*, in reference to the *disk* solution presented in Gross and Wagner (1950) and later with some modifications in Laslier and Picard (2002).

Note that because this is not the only way to construct multivariate distributions satisfying the restrictions above, this method might not describe the entire set of F^*s even in cases where the method is applicable.

5.3.1 Triangle Solution - Gross and Wagner (1950)

First, consider the case presented in Gross and Wagner (1950) for $n = 3$ asymmetric battlefield weights. The following process generates three dimensional vectors $\mathbf{x} = (x_1, x_2, x_3)$ in the two dimensional simplex Δ^2 such that each x_k is distributed uniformly over $[0, 2e_k]$.

Think of the triangle of sides[2] e_1, e_2, e_3, as belonging to the plane with z-coordinate zero in the three-dimensional space (x, y, z). Inscribe a disk of centre O and radius r within that triangle. This disk is the projection (onto the plane $(x, y, 0)$) of the sphere $\mathcal{S}$ of centre O and radius r belonging to the three dimensional space (x, y, z). Finally, let R be a generic point that

[2]For simplicity we identify a side of the triangle with its length. So we use e_k to refer both to a segment and to its length. Note also that this triangle always exists since $e_k < 1/2\ \forall k$.

is uniformly distributed on the surface of the sphere $\mathcal{S}$, and let P be the orthogonal projection of R onto the plane.

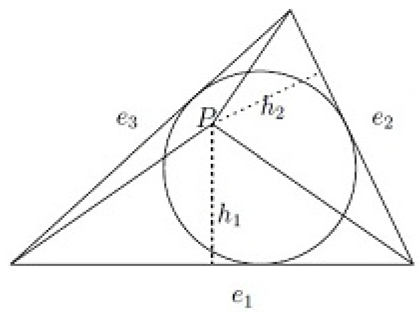


Figure 1: *The Triangle Solution.*

For all k, h_k is the distance of P from the side e_k. In the three-dimensional space, it is also the distance of R from $\mathcal{P}_k$, the vertical plane tangent to the sphere of centre O and which projects onto the side e_k.

If R is uniformly distributed on the surface of the sphere, what is the distribution of h_k? For all $t \in [0, 2r]$, the *spherical cap of height* t is the region of the sphere $\mathcal{S}$ that lies between the vertical plane $\mathcal{P}_k$, and the vertical plane parallel to $\mathcal{P}_k$ and at a distance t away from it. Then, for all $t \in [0, 2r]$, $\Pr(h_k < t) = \Pr(R \in \text{cap of height } t)$, and since R is uniformly distributed on the surface of the sphere, this probability equals the surface area of the cap[3] of height t, $t \in [0, 2r]$, divided by the total surface area of the sphere:

$$\Pr(h_k < t) = \frac{2\pi \int_o^t r\ dx}{2\pi \int_o^{2r} r\ dx} = \frac{t}{2r},$$

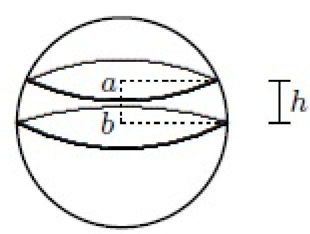
and so h_k is distributed uniformly on $[0, 2r]$.

Back in the two-dimensional plane, call A_k the area of the triangle of height h_k and side e_k subtended by P. For all k, $A_k = e_k h_k/2$. Since $h_k \sim U[0, 2r]$, it must be that $A_k \sim U[0, 2re_k/2] \equiv U[0, re_k]$.

Letting $A = A_1 + A_2 + A_3 = (e_1 + e_2 + e_3)r/2 = r/2$ be the total area of the triangle, we assimilate the fractions x_1, x_2, x_3, which are assumed to belong to the two dimensional simplex, to the fractions A_1/A, A_2/A, A_3/A, which belong to the two dimensional simplex by construction. So for all k, $x_k{=}A_k/A = 2A_k/r$. Then finally, since $A_k \sim U[0, re_k]$, it must be that $x_k \sim U[0, 2re_k/r]$, i.e. $x_k \sim U[0, 2e_k]$. $\square$

Note that this construction is unique as there is only one cyclical permutation of 3 objects, if we account for the orientation of the cycle (i.e. treat $\{x, y, z\}$ and $\{z, y, x\}$ as equivalent).

[3] Note that the result of this sub-section is largely driven by the following property of spheres: Consider the spherical segment of height h. Its surface (excluding the bases) is called a zone. Its mathematical expression is $2\pi \int_a^b r\ dx = 2\pi rh$. Note that this area is *independent of the vertical position* of the zone.



5.3.2 Regular n-gon - the *disk* solution - Laslier and Picard 2002

As n increases beyond three, note that different orderings of the e_k's create different supports for the equilibrium strategy. Moreover for $n \geq 4$ it is not possible to inscribe a circle in *any* n-gon. Irregular n-gons are the object of the next section.

Let us first consider the case of regular n-gons, which is the result presented in Laslier and Picard (2002). As supported by the *disk* solution, it is possible to construct a multivariate distribution F^* for the case in which all states carry the same value: $e_k = 1/n$ for all k. Then, regardless of n, it is possible to inscribe a circle within the n-gon; and following the same method as in the triangle case, the process generates n-dimensional vectors $\mathbf{x} = (x_1, ..., x_n)$ belonging to the $(n-1)$-dimensional simplex, such that each x_k is distributed uniformly over $[0, 2/n]$.

In the two-dimensional, oriented plane, consider the regular n-gon $\{P_0, ..., P_{n-1}\}$ centered at zero such that

$$P_k = \left(\rho \ cos \tfrac{(2k+1)\pi}{n}, \rho \ sin \tfrac{(2k+1)\pi}{n}\right) = \rho \ e^{i \frac{(2k-1)\pi}{n}}.$$

The disk that is inscribed within this n-gon is centered at zero and has radius r such that

$$\left| \frac{P_k + P_{k+1}}{2} \right| = \frac{\rho}{2} \sqrt{2(1 + cos \frac{4\pi}{n})} = r.$$

This disk is the projection onto the plane of the sphere centered at zero of radius r. To generate the n-dimensional vector $\mathbf{x}$, use the method corresponding to the three-dimensional case described above.

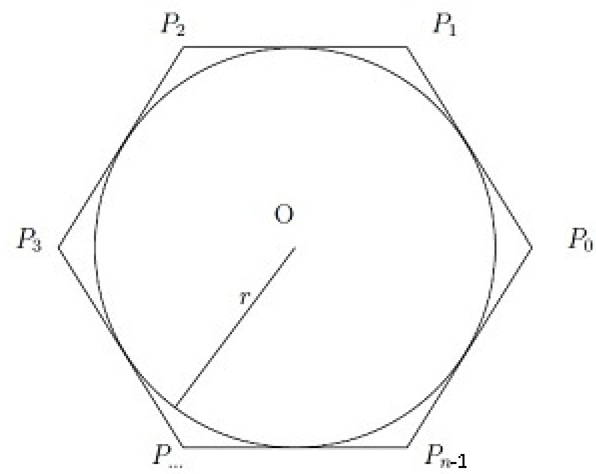


Figure 2: *Regular n-gon*

Note that there are as many *disk* solutions as there are ways to order n objects in a circle without taking into account the orientation of the circle, i.e. $(n-1)!/2$. Even though all sides have the same length, meaning that the n-gons $\{e_1, e_2, e_3, e_4\}$ and $\{e_3, e_2, e_1, e_4\}$ say, look identical, the correlations of vector coordinates deriving from the resulting joint distributions will be different.

5.4 Multivariate Distributions - Irregular n-gon

In this section, we present a novel construction method for the case where battlefield values differ. Note that if there exists an n-gon with sides of lengths corresponding to the battlefield values and that admits an inscribed circle, we can use the method for constructing F^* described above. But as noted in the previous section, for $n \geq 4$ it is not possible to inscribe a circle in *any* n-gon. Roughly, the figure needs to be sufficiently regular. Indeed, for some $\{e_k\}_{k=1}^n$, it may never be possible to inscribe a circle in an n-gon of sides e_k regardless of the ordering. This is the case for instance if one e_k is much larger than all the others.

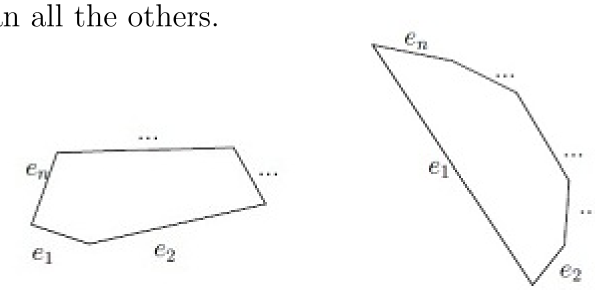


Figure 3: *Ill-behaved n-gons*

The next sub-section describes how to construct an irregular n-gon admitting an incircle, assuming this is possible. Then, sub-section 5.4.2 presents the necessary and sufficient conditions on battlefield weights guaranteeing it is possible to construct an irregular n-gon admitting an incircle.

5.4.1 Irregular n-gon - the *modified disk* solution

Consider the n-vector $\mathbf{e} = (e_1, ..., e_n)$ of battlefield weights, and define the n-vector $\gamma = (\gamma_1, ..., \gamma_n)$ to be a reordering of $\mathbf{e}$ satisfying conditions described in section 5.4.2. Let k, the index of the coordinates of γ, be congruent modulo n.

Given γ, consider the following method of constructing an irregular n-gon of ordered sides γ_1, γ_2, etc, such that a circle is inscribed in it.

For $k = 1, ..., n$, let Γ be a string of n connected segments $[P_{k-1}, P_k]$ of length γ_k with the following equidistance property: let T_k be a point of the segment $[P_{k-1}, P_k]$ such that, for each k, the distances $||T_k P_k||$ and $||P_k T_{k+1}||$ are the same, denoted t_k. The points T_k will be the tangency points between the n-gon and the circle inscribed in it.

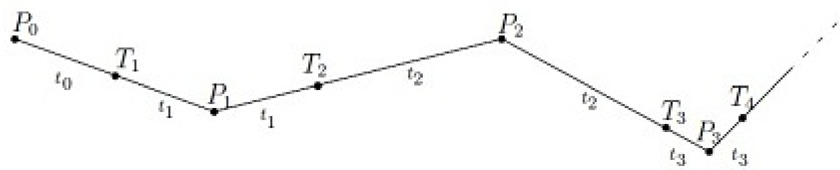


Figure 4: *The set* Γ.

Consider the disk (O, r) and two connected segments $[AB]$ and $[BC]$. Let both segments be tangent to the circle, and let K and L be the points of tangency of $[AB]$ and $[BC]$ respectively. It is a well known result that the distances $||K - B||$ and $||B - L||$ are then necessarily equal.

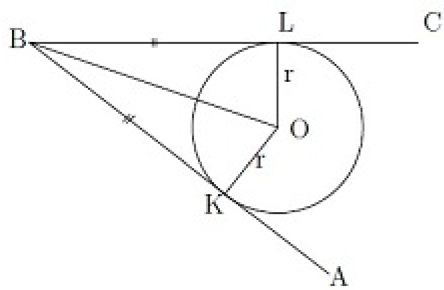


Figure 5: *Equidistance*

Accordingly, if a sequence of connected segments can be wrapped around a circle (regardless of the number of times the sequence goes around the circle) in such a way that all segments are tangent to that circle, then the points of tangency of two consecutive segments are equidistant from the point common to both segments.

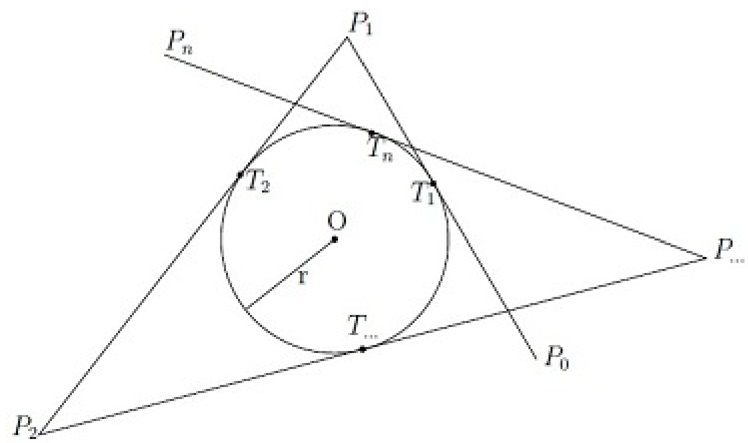


Figure 6: *Wrapping Gamma around a circle.*

This equidistance property is, by construction, satisfied by the set Γ. So Γ can be wrapped around *any* circle (O, r). The number of times we can wrap this set of connected segments around a circle depends on r. Theorem 4 states that there is only one value of r for which we can wrap a given Γ around a circle, such that $P_n = P_0$, closing the n-gon. Denote θ_k the angle $(P_{k-1}, 0, P_k)$.

Theorem 4. *For a given* Γ, $\sum_{k=1}^n \theta_k = f(r)$ *where* f *is a continuous, strictly monotone function. Therefore,* r^* *satisfying* $f(r^*) = 2\pi$ *is unique.*

Proof. Denote a_k the angle $(T_k, 0, P_k)$. Then $\sum_{k=1}^n \theta_k = 2\sum_{k=1}^n a_k$. The function $\sin^{-1}$ is defined (and monotonically increasing) on $[-1, 1]$, and since for all $x \in \mathbb{R}^{+*}$, $0 < x/\sqrt{x^2 + r^2} \leq x/\sqrt{x^2} = 1$, so

$$\sin a_k = \frac{t_k}{\sqrt{t_k^2 + r^2}} \quad \Leftrightarrow \quad a_k = \sin^{-1}\left[\frac{t_k}{\sqrt{t_k^2 + r^2}}\right],$$

and

$$\sum_{k=1}^n \theta_k = 2\sum_{k=1}^n \sin^{-1}\left[\frac{t_k}{\sqrt{t_k^2 + r^2}}\right] = f(r),$$

which is strictly decreasing, and hence invertible in r for all n. The proposition follows. □

Note that r^* depends on the particular choice of t_k so that any vector $\mathbf{e}$ may be associated with several r^*.

We now present the conditions on γ that are necessary and sufficient for the existence of a set Γ, and hence for the existence of an n-gon of sides given by γ and admitting an inscribed circle.

5.4.2 Necessary and sufficient conditions

When the n-gon is regular, it is always possible to inscribe a circle within it. As we deviate from the regular n-gon, what are sufficient conditions on $\{e_k\}_{k=1}^n$ and on the ordering of the sides of the irregular n-gon that need to be satisfied to ensure that a circle can be inscribed within it?

First note that the restriction $e_k < 1/2\ \forall k$ guarantees that a convex n-gon with sides of lengths given by $\{e_k\}_{k=1}^n$ exists.

This section describes conditions for reordering the coordinates of the n-vector $\mathbf{e} = (e_1, ..., e_n)$ to form the n-vector $\gamma = (\gamma_1, ..., \gamma_n)$. Recall that k, the index of the coordinates of γ, is congruent modulo n. The conditions are necessary and sufficient to be able to inscribe a circle in the irregular convex n-gon with ordered sides given by γ, and from there, to build an equilibrium strategy F^*.

It will be shown that some vectors $\mathbf{e}$ will not admit any reordering γ satisfying these conditions so that it will not be possible to build a distribution with the properties of F^* using the geometric method.

To be able to build such a set Γ, the vector γ needs to satisfy the following restrictions **(P1)** and **(P2)**, that are divided in sub-cases depending on whether n, the number of battlefields, is odd or even.

(P1E) If n is even, then

$$\sum_{i=1}^{n}(-1)^i\gamma_{k+i} \;=\; 0.$$

(P2E) If n is even, then for any k, there exists a constant $c > 0$ such that for $\nu = 1, 2, ..., \frac{n}{2}$,

$$\max_{\nu}\{\sum_{i=0}^{2\nu+1}(-1)^i\gamma_{k+i}\} \;<\; c \;<\; \min_{\nu}\{\sum_{i=0}^{2\nu}(-1)^i\gamma_{k+i}\}$$

(P1O) If n is odd, then for any k,

$$t_k = \frac{1}{2}\;\sum_{i=0}^{n-1}(-1)^{i+1}\gamma_{k+i}.$$

(P2O) If n is odd, then for any k,

$$\gamma_k > \left\|\sum_{i=1}^{n-1}(-1)^{i+1}\gamma_{k+i}\right\|.$$

These restrictions are all derived from the fact that by definition, $\gamma_k = t_k + t_{k+1}$, and from the two following requirements:

1)	Congruence	$\forall k,\ t_{k+n} = t_k.$
2)	Fit	$\forall k,\ 0 < t_k < \gamma_k.$

(P1) and **(P2)** hold if and only if congruence and fit are satisfied. The details can be found in appendix 5.7.2.

Congruence and fit are necessary and sufficient conditions for γ to generate a set Γ as defined in section 5.4.1. It follows that these properties of γ are necessary and sufficient for the resulting Γ to generate at least one n-gon admitting an incircle. Of course, they are all satisfied when all the coordinates of γ are the same - corresponding to the case of Laslier's and Picard's regular n-gon. The following theorem is the main result of this section:

Theorem 5. *If for a vector* $\boldsymbol{e}$ *of battlefield weights we can find a reordering* γ *satisfying **(P1)** and **(P2)**, then we can construct an irregular* n*-gon with an inscribed circle of radius* r^**.*

The radius r^* is defined in theorem 4. In the remainder of this section, we provide some insight into these properties and in particular (section 5.4.3), ask how easy they are to satisfy.

Conditions **(P2E)** and **(P1O)**, relate to the tangency points of the inscribed circle with the n-gon. They ensure that if t_k belongs to the interval $(0, \gamma_k)$, then t_{k+1}, which is equal to $\gamma_k - t_k$, belongs to the following interval, (γ_k, γ_{k+1}). We can see that while for n odd, the conditions on the length t_k are very strict (equality), for n even it will be sufficient for t_k to belong to the interval defined in **(P2E)**:

(P2E)' If n is even, then for all k,

$$t_k \in \Big(\max_{\nu}\{\sum_{i=0}^{2\nu+1}(-1)^i\gamma_{k+i}\}\ , \min_{\nu}\{\sum_{i=0}^{2\nu}(-1)^i\gamma_{k+i}\}\Big)$$

So for a given γ, if n is even, it is possible to build an infinity of sets Γ as long as **(P2E)'** is satisfied, while for n odd, there exists a unique Γ with distances t_k satisfying **(P1O)**.

The remaining two conditions, **(P1E)** and **(P2O)**, are discussed in the next sub-section.

5.4.3 The constrained integer partitioning problem

It is clear that while some vectors $\mathbf{e}$ may admit several corresponding vectors γ, others may admit none. Indeed, the properties are all regularity restrictions on the ordering of the coordinates of γ and impose some balance. Notice that **(P1E)** can be rewritten as:

(P1E)' If n is even, then

$$\sum_{i=1}^{\frac{n}{2}} \gamma_{(k+2i)} = \sum_{i=1}^{\frac{n}{2}} \gamma_{(k+2i-1)} = \frac{1}{2},$$

and that **(P2O)** can be rewritten as:

(P2O)' If n is odd, then for any k,

$$\gamma_k > \left\| \sum_{i=1}^{\frac{n-1}{2}} \gamma_{(k+2i)} - \sum_{i=1}^{\frac{n-1}{2}} \gamma_{(k+2i-1)} \right\|.$$

So the two conditions are similar in requiring that the n-gon generated by γ is balanced in the sense that the summed length of odd sides and the summed length of even sides are equal (for n even) or close in a precise sense (for n odd).

As a brief digression, note that they also can be interpreted as the requirement that there exists a coalition of states such that each state in that coalition and each state in the complement coalition is pivotal. Pivotality is not a very apt concept here, as players are maximising their plurality. It would be more fitting in a context where players maximise their probability of winning.

More can be said about when **(P1E)** and **(P2O)** may be satisfied by noting that these conditions are related to the *constrained integer partitioning problem*, a classic problem of combinatorial optimisation. The exercise consists in partitioning n integers into two subsets of given cardinalities such that the *discrepancy*, the absolute value of the difference of their sums, is minimized.

(P1E) corresponds to the constrained partitioning problem in which the cardinality of the two resulting subsets is $n/2$ and the discrepancy is equal to zero. A partition with a discrepancy of zero is called a *perfect partition.* **(P2O)** corresponds to n instances of a more relaxed version of the constrained partitioning problem just described: for each $k = 1, ..., n$, the aim is to partition $n - 1$ integers into two subsets of equal cardinality, such that the discrepancy is less than γ_k.

These are computationally difficult problems. The unconstrained partitioning problem is NP-complete, and while some algorithms deliver good approximations of the optimal partition (the partition with the lowest possible discrepancy), the brute force algorithms that compares the discrepancies of all possible partitions is still the best known solution to the problem.

Borgs et al. (2003) identify two *phases* of the constrained problem depending on its computational difficulty. They study the typical behaviour of the optimal partition when the n integers are i.i.d. random variables chosen uniformly from the set $\{1, ..., 2^m\}$ for some integer m.

They find that, for m and n tending to infinity keeping the ratio m/n constant, with probability tending to one there exists a perfect partition when $m/n < 1$. They call this the *perfect phase* of the problem. In the *hard* *phase* of the problem, for $m/n > 1$, the probability of a perfect partition tends to zero and the optimal partition is unique, making computation of the optimal partition more difficult there. Still, the minimum discrepancy, i.e. the discrepancy of the optimal partition, can be bounded from above and below.

While in the limiting case, the phase transition is sharp at 1, in finite cases, the phase transition happens within a specified interval containing 1, and it is not clear whether the transition is sharp. Finally, the number of perfect partitions in the perfect phase is lower than in the limiting case by about twenty percent for a given ratio m/n.

For the purpose of this chapter, the results of Borgs et al. allow the conclusion that **(P1E)** and **(P2O)** are likely to be more easily satisfied for $m/n < 1$ than for $m/n > 1$, and that while **(P2O)** may be satisfied for $m/n > 1$, **(P1E)** never is.

	$m/n < 1$	$m/n > 1$
(P1E)	easy	impossible
(P2O)	easy	hard

Finally note the importance of the assumption that battlefield values are integers. Indeed, were battlefield values drawn from $\mathbb{R}$, the condition for n even would hold with probability zero.

5.5 Application

One compelling illustration of this model is the election of US presidents by electoral college: first, during primaries, two candidates, one Democrat, the other Republican, are chosen to represent their party in the general election, which is then held simultaneously in all 51 US states (50 + D.C). Each state is allocated a number of electoral votes depending on its population[4]. There are 538 electoral votes in total. A candidate gains all electoral votes of a given state if he receives more than half the votes cast in that state. To win the election, a candidate must win at least 270 electoral votes.

This situation can be modeled as an asymmetric Colonel Blotto game under the following three assumptions: (i) presidential candidates face identical budget constraints, (ii) the probability of winning the election in a given state increases with campaigning resource allocated to that state, and (iii) candidates wish to maximise their plurality, rather than the probability of winning the election.

The first two assumptions are the least controversial. In fact assumption (i) is trivially satisfied if we think of the campaigning resource as time spent campaigning in each state.

What if we think of money as the resource? In practice, candidates can choose whether to self-finance their general election campaign, or (since 1976) can accept public funding[5]. To be eligible to receive the public funds, a candidate must limit spending to the donation[6]. So if both candidates are

[4]For details, see Appendix 5.7.3

[5]For information on the Public Matching Fund scheme, visit the Federal Election Commission at http://www.fec.gov/.

[6]In essence. More precisely, the candidate may not accept private contributions for the campaign. Private contributions may, however, be accepted for a special account maintained exclusively to pay for legal and accounting expenses associated with complying with the campaign finance law. These legal and accounting expenses are not subject to the expenditure limit. For more detail, see the FEC brochure for Public Funding of

publicly funded, it makes sense to assume that they both face the same resource constraint.

The assumption of equal budgets becomes more trying if at least one of the candidates is self-funded. Indeed, there is considerable evidence that in these cases, budgets differ, as seen in the latest US presidential elections.

The positive relationship between campaign effort and votes is well documented, whether campaigning effort is understood to be time spent campaigning in a state (Herr (2008)) or financial campaign expenditures in that state (Chapman and Palda (1984)). So assumption (ii) is also pretty unproblematic.

This is not so for the last assumption. In general, one would assume that candidates maximise the probability of their winning the election. Nevertheless one could argue that because presidential elections coincide with Senate and House of representative elections, presidential candidates do campaign so as to maximise the plurality of votes in they favour, not only so as to win the presidential election. This is more believable in cases where one candidate already expects to win with a significant plurality, but surely not when elections are close. Either way, it is fair to say that maximising the plurality in his favour is at least a candidate's secondary objective.

One strong argument supporting the claim that candidates care at least a little about plurality is that they do indeed campaign in all states, while ignoring small states (states with few electoral votes, that have little chance of being pivotal) would be consistent with the strategy of a candidate solely trying to maximise his probability of winning the election.

Presidential Elections at http://www.fec.gov/pages/brochures/pubfund.shtml.

So we can think of the US general election game as a Colonel Blotto game. In both cases candidates choose how to allocate a fixed amount of resources across states. Strategic considerations arise because of the positive relationship between campaign effort and votes. By spending more in a state than his opponent, a candidate increases his chances of winning that state.

In this section we look for a solution to a Colonel Blotto game in which each state has a value corresponding to its relative number of electoral votes. The distribution of electoral votes across states is shown in Appendix 5.7.3.

Two candidates with budgets $X_A = X_B = X$ decide how to allocate their campaigning funds across $n = 51$ states indexed by $k \in 1, ..., n$. The value of state k is e_k which corresponds to the number of electoral votes allocated to state k as a fraction of the total number of electoral votes, 538. For instance, the state of Alabama has 9 electoral votes, so for that state, $e = 9/538$. Accordingly $e_k < 1$ for all k and $\sum_{k=1}^{n} e_k = 1$.

Candidate i's plurality, i.e. the number of electoral votes won minus the number of electoral votes lost, is measured by the function $g_i : \mathcal{S}_i \times \mathcal{S}_i \to \mathbb{R}$ defined in section 5.2.

Since this matches the setup of section 5.2, the results of all following sections hold, including the existence of one equilibrium distribution. Indeed, consider the vector γ_n presented in Appendix 5.7.4. It is such that each e_k corresponds to the number of electoral votes allocated to state k as a fraction of the total number of electoral votes, 538. For clarity, we multiply all numbers back by 538. Note that this solution uses the current distribution of electoral votes (i.e. the third column in table 5.7.3), but that the construction method works equally well for the other two distributions.

This vector satisfies the conditions **(P1)** and **(P2)** for n odd ($n = 51$). Note that within the framework of section 5.4.3, the 51 partitioning problems corresponding to this exercise are in the perfect phase. Here, the greatest of the $n = 51$ integers is 55, the number of electoral votes for the state of California. So we can treat the electoral votes as n i.i.d integers chosen uniformly from the set $\{1, ..., 2^m\}$ with $m = 6$, in which case $m/n = 6/51 \ll 1$ (perfect phase) so that the partitioning problem should be relatively easy to solve. Indeed, a solution can be easily found heuristically, as shown in Appendix 5.7.4. This illustrates one possible equilibrium of the US general elections game.

5.6 Conclusion and Open Questions

This chapter describes a geometrical method for constructing equilibrium distribution in the Colonel Blotto game with asymmetric battlefield values. The appeal of geometrical methods for constructing n-dimensional distributions subject to restrictions on their support and their margins lies in the relative simplicity with which they describe complicated multi-dimensional objects. The drawback is that they may fail to generate the full set of distributions satisfying given restrictions on support and margins. This downside is limited when that set is well defined, as it is here, so that the exercise becomes to generate instances of these well-defined objects.

The method presented in this chapter generalises to the n-dimensional case a construction method first proposed by Gross and Wagner. It does particularly well in instances of the Colonel Blotto game in which the battlefield weights satisfy some clearly defined regularity conditions (Section 5.4.2). Though these conditions constrain the set of games in which this method reliably generates equilibrium strategies, they are less restrictive than the condition of symmetry across all battlefields (Laslier and Picard). Moreover, their implications suggest directions for further research.

Noticing that the conditions on the reordering γ can be interpreted as the requirement that there exists a coalition such that every battlefield is pivotal suggests a parallel between behaviour of candidates seeking to maximise plurality and candidates seeking to maximise probability of victory, though this work leaves the exact relationship between these games an open question.

Finally, the restrictions on the support of equilibrium distributions limit the number of possible correlations across x_k's. This captures the idea that even though it is intuitive that more resources are likely to be allocated to battlefields with greater weight, the solution suggests that allocations to different battlefields interact in a particular way. Looking more carefully at possible correlations across x_k's could be interesting from the empirical point of view.

5.7 Appendix Chapter 5

5.7.1 Proof of Proposition 1

Proof of (i) and (ii): Straightforward.

Proof of (iii): To prove this point, it is sufficient to show that the payoff to any pure strategy $\mathbf{y} \in \mathcal{S}^i$ against F^* is non-positive. First we show that the expected payoff to player i from playing F^* against F^* is zero. Let $\mathbf{x}^i = (x_1^i, ..., x_n^i)$ and $\mathbf{x}^j = (x_1^j, ..., x_n^j)$ be generated by F^*. Accordingly, for all $k = 1, ..., n$, x_k^i and x_k^j are drawn from the uniform distribution over $[0, 2e_k]$ and $Pr(x_k^j < x_k^i) = F_k^*(x_k^i) = \frac{x_k^i}{2e_k}$. So given $\mathbf{x}^i$, for all $k = 1, ..., n$,

$$E[sgn(x_k^i - x_k^j)|\mathbf{x}_i] = 2F^*(x_k^i) - 1 = \frac{x_k^i}{e_k} - 1.$$

And hence, for all $k = 1, ..., n$,

$$\begin{aligned} E[sgn(x_k^i - x_k^j)] &= \int_0^{2e_k} \left(\frac{t}{e_k} - 1 \right) \, dF_k^*(t) \\ &= \frac{1}{2e_k} \int_0^{2e_k} \left(\frac{t}{e_k} - 1 \right) \, dt \end{aligned}$$

which is zero for all $k = 1, ..., n$ so that:

$$E[g(F^*, F^*)] = \sum_{k=1}^{n} e_k \cdot E[sgn(x_k^i - x_k^j)] = 0.$$

Now consider the payoff to player i of playing an arbitrary pure strategy $\mathbf{y} \in \mathcal{S}^i = \Delta^{n-1}$ against F^*. Since for all $k = 1, ..., n$, $e_k < \frac{1}{2}$ and F_k^* is the uniform distribution on $[0, 2e_k]$, $F_k^*(y_k) = y_k/2e_k$ if $y_k \in [0, 2e_k]$ and $F_k^*(y_k) = 1$ if $y_k > 2e_k$. So

$$E[sgn(y_k - x_k^j)|\mathbf{y}] = 2F_k^*(y_k) - 1$$

$$= 2 \min\left\{1, \frac{y_k}{2e_k}\right\} - 1.$$

Hence:

$$E[g(\mathbf{y}, F^*)] = \sum_{k=1}^{n} e_k \ \min\left\{1, \frac{y_k}{e_k} - 1\right\}$$
$$\leq \sum_{k=1}^{n} e_k \Big(\frac{y_k}{e_k} - 1\Big)$$

The last term equals $\sum_{k=1}^{n} y_k - \sum_{k=1}^{n} e_k$ which is zero since $\mathbf{y} \in \Delta^{n-1}$ and $\sum_{k=1}^{n} e_k = 1$ by construction. So $g(\mathbf{y}, F^*) \leq 0 = g(F^*, F^*)$ for all $\mathbf{y} \in \mathcal{S}_i$. ■

5.7.2 Restrictions on γ, the reordering of e

In this appendix, I illustrate how to derive the conditions **(P1)** and **(P2)** from the property $t_k + t_{k+1} = \gamma_k$, and the requirements:

1)	Congruence	$\forall k,\ t_{k+n} = t_k$
2)	Fit	$\forall k,\ 0 < t_k < \gamma_k$

First, let's develop the first requirement. For n even:

$$t_{k+n} = t_k$$
$$\Leftrightarrow \quad t_k = \gamma_{k+n-1} - \gamma_{k+n-2} + \gamma_{k+n-3} - ... - \gamma_k + t_k$$
$$\Leftrightarrow \quad \textstyle\sum_{i=1}^{n} (-1)^i \gamma_{k+i} = 0$$
$$\Leftrightarrow \quad \textbf{(P1E)}$$

For n odd:

$$\begin{aligned}
& t_{k+n} = t_k \\
\Leftrightarrow\quad & t_k = \gamma_{k+n-1} - \gamma_{k+n-2} + \gamma_{k+n-3} - ... + \gamma_k - t_k \\
\Leftrightarrow\quad & 2t_k = \gamma_{k+n-1} - \gamma_{k+n-2} + \gamma_{k+n-3} - ... + \gamma_k \\
\Leftrightarrow\quad & 2t_k = \textstyle\sum_{i=1}^{n} (-1)^{i+1} \gamma_{k+i} \\
\Leftrightarrow\quad & \mathbf{(P1O)}
\end{aligned}$$

Now, let's develop the second requirement.

For n odd, from (P1O) we know that $t_k = \frac{1}{2}(\gamma_k - \gamma_{k+1} + \gamma_{k+2} - ... + \gamma_{k+n-1})$. So

$$0 < t_k < \gamma_k$$

$$\Leftrightarrow \quad -\gamma_k < -\gamma_{k+1} + \gamma_{k+2} - ... + \gamma_{k+n-1} < \gamma_k$$

$$\Leftrightarrow \quad \gamma_k > \left\| \sum_{i=1}^{n-1} (-1)^{i+1} \gamma_{k+i} \right\|$$

$$\Leftrightarrow \mathbf{(P2O)}$$

For n even, the fit requirement, $\forall k$, $0 < t_k < \gamma_k$ gives us n restrictions:

$$\begin{array}{lccccc}
(1) & 0 & < & t_k & < & \gamma_k \\
(2) & 0 & < & t_{k+1} & < & \gamma_{k+1} \\
(3) & 0 & < & t_{k+2} & < & \gamma_{k+2} \\
 & & & \vdots & & \\
(n) & 0 & < & t_{k+n-1} & < & \gamma_{k+n-1}
\end{array}$$

They can all be simplified to n restrictions on t_k:

$$\begin{array}{lrclcl}
(1) & 0 & < & t_k & < & \gamma_k \\
(2) & \gamma_k - \gamma_{k+1} & < & t_k & < & \gamma_k \\
(3) & \gamma_k - \gamma_{k+1} & < & t_k & < & \gamma_k - \gamma_{k+1} + \gamma_{k+2} \\
 & & & \vdots & & \\
(n) & \gamma_k - \gamma_{k+1} + ... + \gamma_{k+n-2} - \gamma_{k+n-1} & < & t_k & < & \gamma_k - \gamma_{k+1} + ... + \gamma_{k+n-2}
\end{array}$$

Notice that t_k faces $n/2$ upper bounds and $n/2$ lower bounds. All n conditions are satisfied if:

$$\max_{\nu}\{\sum_{i=0}^{2\nu+1}(-1)^i\gamma_{k+i}\} \ < \ t_k \ < \ \min_{\nu}\{\sum_{i=0}^{2\nu}(-1)^i\gamma_{k+i}\}$$

and for this to be possible, γ needs to satisfy **(P2E)**.

5.7.3 Distribution of Electoral Votes (Source: FEC www.fec.gov)

State	1981-1990	1991-2000	2001-2010	State	1981-1990	1991-2000	2001-2010
Alabama	9	9	9	Missouri	11	11	11
Alaska	3	3	3	Montana	4	3	3
Arizona	7	8	10	Nebraska	5	5	5
Arkansas	6	6	6	Nevada	4	4	5
California	47	54	55	New Hampshire	4	4	4
Colorado	8	8	9	New Jersey	16	15	15
Connecticut	8	8	7	New Mexico	5	5	5
Delaware	3	3	3	New York	36	33	31
D.C	3	3	3	North Carolina	13	14	15
Florida	21	25	27	North Dakota	3	3	3
Georgia	12	13	15	Ohio	23	21	20
Hawaii	4	4	4	Oklahoma	8	8	7
Idaho	4	4	4	Oregon	7	7	7
Illinois	24	22	21	Pennsylvania	25	23	21
Indiana	12	12	11	Rhode Island	4	4	4
Iowa	8	7	7	South Carolina	8	8	8
Kansas	7	6	6	South Dakota	3	3	3
Kentucky	9	8	8	Tennessee	11	11	11
Louisiana	10	9	9	Texas	29	32	34
Maine	4	4	4	Utah	5	5	5
Maryland	10	10	10	Vermont	3	3	3
Massachusetts	13	12	12	Virginia	12	13	13
Michigan	20	18	17	Washington	10	11	11
Minnesota	10	10	10	West Virginia	6	5	5
Mississippi	7	7	6	Wisconsin	11	11	10
				Wyoming	3	3	3

5.7.4 One possible support of the ***modified disk*** solution applied to US data.

For clarity, all numbers are multiplied by 538.

k	c_k	l_k	l_{k+1}	k	c_k	l_k	l_{k+1}
1	31	2	29	26	3	1	2
2	8	6	2	27	3	2	1
3	9	3	6	28	3	1	2
4	10	7	3	29	4	3	1
5	11	4	7	30	4	1	3
6	17	13	4	31	4	3	1
7	20	7	13	32	4	1	3
8	21	14	7	33	5	4	1
9	27	13	14	34	5	1	4
10	21	8	13	35	5	4	1
11	15	7	8	36	6	2	4
12	15	8	7	37	6	4	2
13	15	7	8	38	7	3	4
14	10	3	7	39	7	4	3
15	7	4	3	40	8	4	4
16	7	3	4	41	9	5	4
17	6	3	3	42	9	4	5
18	5	2	3	43	10	6	4
19	5	3	2	44	10	4	6
20	4	1	3	45	11	7	4
21	3	2	1	46	11	4	7
22	3	1	2	47	11	7	4
23	3	2	1	48	12	5	7
24	3	1	2	49	13	8	5
25	3	2	1	50	34	26	8
-	-	-	-	51	55	29	26

Bibliography

Asawa, M. and D. Teneketzis (1996). Multi-armed bandits with switching penalties. *IEEE transactions on automatic control 41*(3), 328–348.

Banks, J. and R. Sundaram (1994). Switching costs and the Gittins index. *Econometrica*, 687–694.

Bellman, R. (1957). *Dynamic programming*. Princeton University Press.

Bertsekas, D. (2007). *Dynamic programming and optimal control, vol. II.* Athena Scientific.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy 100*(5), 992.

Bolton, P. and C. Harris (1999). Strategic experimentation. *Econmetrica*, 349–374.

Bonatti, A. and J. Hörner (2011). Collaborating. *The American Economic Review 101*(2), 632–663.

Borel, E. and J. Ville (1938). Application de la theorie des probabilites aux jeux de hasard, original edition by Gauthier-Villars, Paris, 1938; reprinted at the end of Theorie mathematique du bridge a la portee de tous, by E. *Borel & A. Cheron, Editions Jacques Gabay, Paris*.

Borgs, C., J. Chayes, S. Mertens, and B. Pittel (2003). Phase diagram for the constrained integer partitioning problem. *Arxiv preprint cond-mat/0302536*.

Callander, S. and J. Hörner (2006). The wisdom of the minority. *WP Northwestern University*.

Camargo, B. and E. Pastorino (2010). Learning-by-Employing: The value of commitment.

Chakrabarti, S. (1999). Markov Equilibria in Discounted Stochastic Games* 1. *Journal of economic theory 85*(2), 294–327.

Chapman, R. and K. Palda (1984). Assessing the influence of campaign expenditures on voting behavior with a comprehensive electoral market model. *Marketing Science*, 207–226.

Dayanik, S., W. Powell, and K. Yamazaki (2008). Index policies for discounted bandit problems with availability constraints. *Advances in Applied Probability 40*(2), 377–400.

Frostig, E. and G. Weiss (1999). Four proofs of Gittins' multiarmed bandit theorem. *Applied Probability Trust*, 1–20.

Fudenberg, D. and J. Tirole (1985). Preemption and rent equalization in the adoption of new technology. *Review of Economic Studies 52*(3), 383–401.

Gittins, J. and D. Jones (1974). A dynamic allocation index for the sequential design of experiments. *Progress in Statistics 241266*.

Glazebrook, K., P. Ansell, R. Dunn, and R. Lumley (2004). On the optimal allocation of service to impatient tasks. *Journal of Applied Probability*, 51–72.

Gross, O. and R. Wagner (1950). A continuous colonel blotto game. *RAND Research Memorandum*.

Herr, J. (2008). The impact of campaign appearances in the 1996 election. *The Journal of Politics 64*(03), 904–913.

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy 87*(5), 972.

Keller, G. and S. Rady (1999). Optimal experimentation in a changing environment. *Review of Economic Studies*, 475–507.

Keller, G., S. Rady, and M. Cripps (2005). Strategic experimentation with exponential bandits. *Econometrica*, 39–68.

Kvasov, D. (2007). Contests with limited resources. *Journal of Economic Theory 136*(1), 738–748.

Laslier, J. and N. Picard (2002). Distributive politics and electoral competition. *Journal of Economic Theory 103*(1), 106–130.

Moscarini, G., M. Ottaviani, and L. Smith (1998). Social learning in a changing world. *Economic Theory 11*(3), 657–665.

Murto, P. and J. Välimäki (2011). Learning and information aggregation in an exit game. *Review of Economic Studies forthcoming*.

Niederle, M. and A. Roth (2009). Market Culture: How Rules Governing Exploding Offers Affect Market Performance. *American Economic Journal: Microeconomics 1*(2), 199–219.

Niño-Mora, J. (2007). Characterization and computation of restless bandit marginal productivity indices. In *Proceedings of the 2nd international conference on Performance evaluation methodologies and tools.* ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) ICST, Brussels, Belgium.

Papadimitriou, C. and J. Tsitsiklis (1987). The complexity of Markov decision processes. *Mathematics of operations research*, 441–450.

Posner, R., C. Avery, C. Jolls, and A. Roth (2001). The Market for Federal Judicial Law Clerks. *University of Chicago Law Review 68*(3), 793–902.

Roberson, B. (2006). The colonel blotto game. *Economic Theory 29*(1), 1–24.

Roth, A. (1984). The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of Political Economy 92*(6), 991–1016.

Rothschild, M. (1974). A two-armed bandit theory of market pricing. *Journal of Economic Theory 9*(2), 185–202.

Simon, L. and W. Zame (1990). Discontinuous games and endogenous sharing rules. *Econometrica 58*(4), 861–872.

Smith, L. and P. Sorensen (1997). Informational herding and optimal experimentation. *Working paper, Massachusetts Institute of Technology, Department of Economics*.

Toxvaerd, F. (2008). Strategic merger waves: A theory of musical chairs. *Journal of Economic Theory 140*(1), 1–26.

Weinstein, J. (2005). Two notes on the Blotto game. *Manuscript, Northwestern Univ. http://www. kellogg. northwestern. edu/faculty/weinstein/htm/blotto141. pdf*.

Weitzman, M. (1979). Optimal search for the best alternative. *Econometrica*, 641–654.

Whittle, P. (1981). Arm-acquiring bandits. *The Annals of Probability*, 284–292.

Whittle, P. (1988). Restless bandits: Activity allocation in a changing world. *Journal of Applied Probability*, 287–298.