Experimentation with Congestion[*]

Caroline D. Thomas[†]

April 10, 2011

The latest version of this paper is available at:
http://www.ucl.ac.uk/∼uctpcdt/

Abstract

We consider a 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 a time. Firstly, when players cannot reverse their decision to switch to the safe option, the socially optimal policy makes them experiment for longer than they would if they played alone. The equilibrium in the two-player game is in this case always inefficient and involves too little experimentation. As the competition intensifies, the inefficiency increases until the players behave myopically and entirely disregard the option-value associated with experimenting on the risky option. Secondly, when the decision to switch to the safe option is revocable, the player whose risky option is most likely to pay off will interrupt his own experimenting and, with view to easing the opponent's pressure on the common option, force him to experiment more intensely. Even if this does not succeed, the first player will eventually resume his own experimenting and leave the common option for the opponent to take. This result is striking and at odds with intuitions from standard bandit models.

JEL Classification: C72, C73, D83, J41
Keywords: Learning, Strategic Experimentation, Multi-Armed Bandit, Poisson Process, Job Assignment

[*] I thank Martin Cripps and Guy Laroque for supervision and helpful comments. This paper has greatly benefited from discussions with V. Bhaskar, Antonio Guarino, Philippe Jehiel, Lars Nesheim, Balazs Szentes, and Andreas Uthemann.

[†] Department of Economics, University College London, caroline.thomas@ucl.ac.uk

1 Introduction

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".

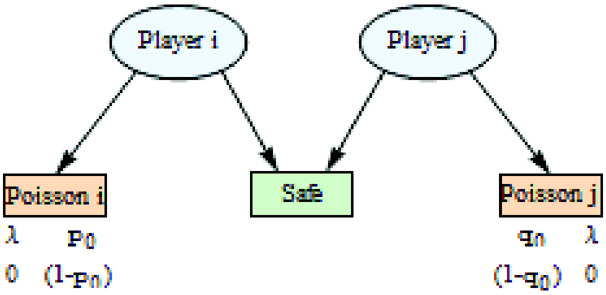
These sorts of considerations are common in all kinds of strategic situations. In the race towards developing a new technology a firm's incentive to invest in a new research project rather than relying on some averred method depends on the strategy of the competitor if the market can only support one 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: because most stores only allow customers to take x items into the dressing room, 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 aim of this paper is to propose a simple model of experimentation with congestion, in which to analyse the trade-offs from strategic interaction. 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*.

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 by assuming that learning takes place in teams, as in Keller, Rady, and Cripps (2005), or by assuming that the qualities of different arms are correlated, as in Murto and Valimaki (2008). In particular, they 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 paper 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 with irrevocable exit, 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 paper is organised as follows: In Section 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 sections. We then present a set of the efficient benchmarks in Section 3. When there is no congestion (3.1), 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 paper. 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 (3.2) and then when that decision is revocable (3.3). Finally, 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. Section 5 concludes and suggests directions for further research.

2 Model

In this section, we define the basic elements of the model on which all remaining sections of this paper 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. If 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 $[t, t+\Delta)$, $\Delta > 0$, a player, say i, activates his risky option without it producing a success, the belief about player i's option at $t + \Delta$ is

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

This is decreasing in Δ: the longer the player experiments without a success, the less optimistic he becomes about his risky option being good. By Bayes' rule 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

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

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 $\iota \in (0, 1)$.

3 Efficient Benchmarks

In this section we present a series of planner problems intended as efficient benchmarks for the models of strategic interaction in Section 4. First we consider the situation in which there is no congestion on the safe option (Section 3.1). Each player then faces an identical two-armed bandit problem with one risky and one safe arm. 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 increasingly 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.

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 allocated to that option switch to the safe option. If neither option produces a success, the planner will eventually allocate one player to the safe option. In Section 3.3, we assume that the planner can do this without restrictions. In Section 3.2 however, 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.

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. 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 his 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 $\Delta \to 0$.

3.1 No congestion - Single player model

First assume that there are two safe options. The planner maximises the join 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.

Formally, the single agent solves the following dynamic problem: at each 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 1.

We solve the agent's problem in Appendix A 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 paper, 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 A.

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

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

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 is increasing 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.

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

$$\mathcal{W}(p, q) = \max_{\kappa} \left\{\kappa\ L^{RR}\mathcal{W}(p, q) + (1 - \kappa)\ L^{RS}\mathcal{W}(p, q)\right\} \tag{2}$$

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:

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

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

Solving the planner's problem (cf. Appendix B), 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 $p \leq q$.

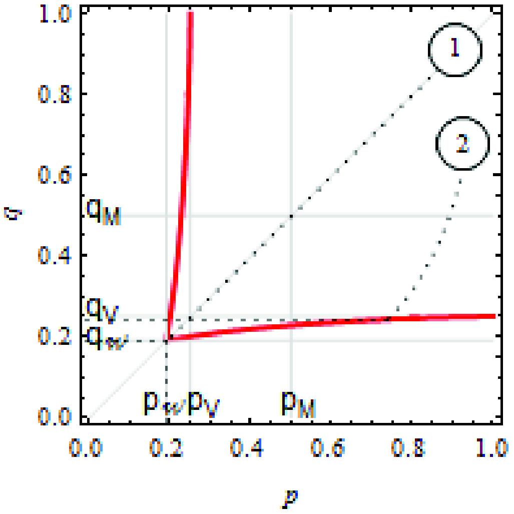


Figure 1: *Threshold beliefs in the planner problem with irrevocable exit.*

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.

If $p_0 = q_0$ (cf. 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

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

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 $p_0 > q_0$ (cf. 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 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.

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 $\Delta \to 0$. 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:

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

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

[2] Because it involves the planner alternating between two options, the existence of that solution is problematic in continuous time. Being aware of the existence issues in the limit, we concentrate on the discrete-time approximation and will use the intuition from a discrete-time problem.

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 $(p,q) \in [0,1)^2$,

$$\mathcal{U}(p,q) = \max\{L^{RR}\mathcal{U}(p,q), L^{RS}\mathcal{U}(p,q)\}. \tag{3}$$

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[3], generating the payoff $\mathcal{A}(p)$ as described in appendix C. 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.

[3] This policy does not make sense in continuous time, but recall that we are reasoning as though the model were in discrete time and let time intervals tend to zero.

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' is defined in Equation 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, 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 below.*

Proof: Cf. Appendix D.

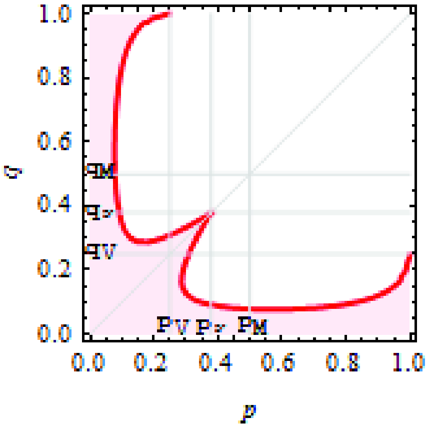


Figure 2: *Set of states in which policy RS is optimal (shaded area) and threshold beliefs in the planner problem with revocable exit.*

Consider the portion of the graph for $p \geq q$. 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 D.

4 Two-Player Game

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 4.1 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 4.2, we will relax this assumption. Then a player can decide to temporarily occupy the safe option, before returning to his risky option.

In Section 4.1 we consider the strategic situation for which the planner problem analysed in Section 3.2 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, 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 4.2), 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 of 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 thus forcing 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 for 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 the safe option for the pessimist. This result may be surprising in light of intuitions from the standard multi-armed bandit problem, in which a player never returns to any option he has rejected in the past.

4.1 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 section 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 section 3.2. 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)[4] 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 strictly dominated for a player whose risky option has produced a success. If only one risky option has produced a success, the other player follow 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

[4] 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.

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:

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

where

$$\begin{aligned}
L^S\mathsf{W}(p, q; k^j) :=&\ \ k^j(p, q)\ \tfrac{a}{\rho} + (1 - k^j(p, q))\ T^i(p, q), \\
L^R\mathsf{W}(p, q; k^j) :=&\ \ p\lambda dt \left(1 + e^{-\rho dt}\tfrac{\lambda}{\rho}\right) \\
&+ (1 - p\lambda dt) \Big((1 - k^j(p, q))\ e^{-\rho dt} p' \tfrac{\lambda}{\rho} \\
&\qquad + k^j(p, q)\ e^{-\rho dt}\left[q\lambda dt V(p') + (1 - q\lambda dt)\mathsf{W}(p', q'; k^j)\right]\Big),
\end{aligned} \tag{4}$$

and with p', q' as defined in Equation 1. 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[5] 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 section 3.2.

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 E. □

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

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 $p = p_M$, 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 F 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 anticipate 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.

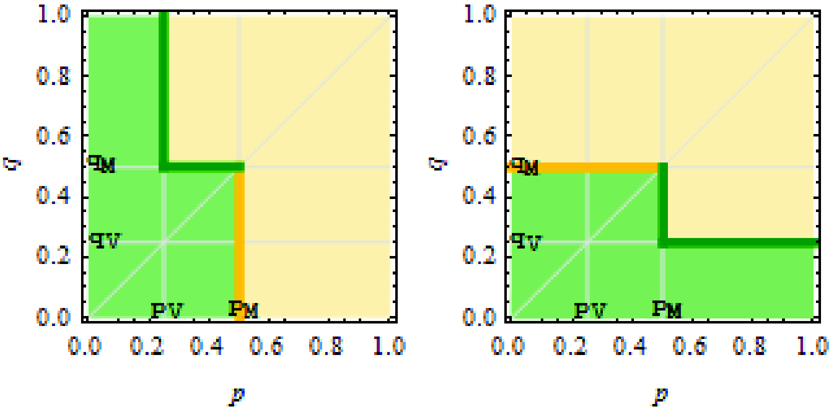


Figure 3: *Equilibrium strategies of player i and player j when exit is irrevocable. If a state* (p, q) *is in the green (dark) area, the player plays the safe option, if it is in the orange (light) area, the player plays his risky option.*

Theorem 1. *Consider the strategy profile illustrated above:*

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

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: Cf. Appendix F. 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.

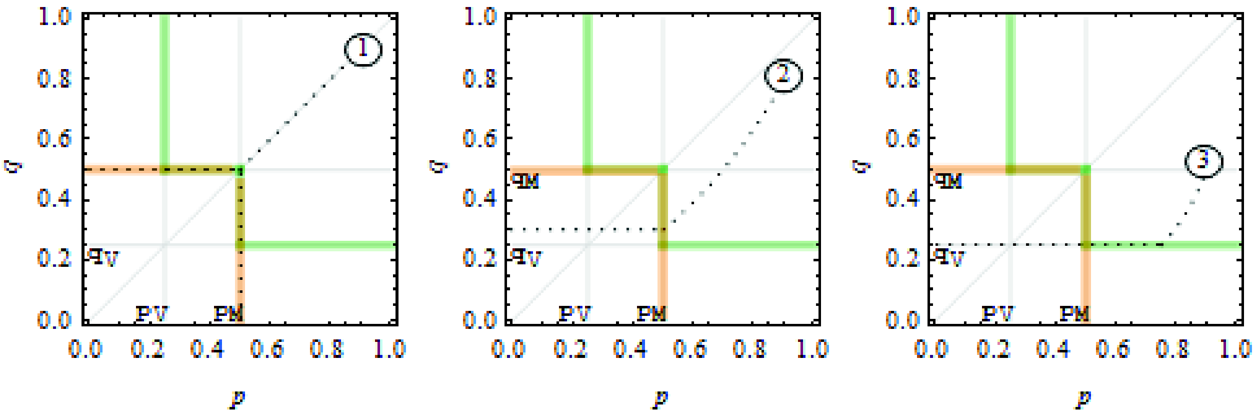


Figure 4: *In Case 1, the prior is* $p_0 = q_0 > \frac{a}{\rho}$. In Case 2, the priors $p_0 > q_0$ are such that at date $t > 0$ satisfying $p_t = p_M$ we have that $q_t > q_V$, while in Case 3 we have that $q_t \leq q_V$.

Case 1: If the prior is $p_0 = q_0 > \frac{a}{\rho}$, then in equilibrium both players switch to the safe option when beliefs reach the single-player myopic threshold belief, $p_M = \frac{a}{\lambda}$. 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[6] 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

[6] Relate to preemption games.

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.

Player j then effectively plays a single-player game and switches to the safe option when $q = q_V$. Then 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.

4.2 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. When exit is irrevocable, the more pessimistic player, say player j, is the first to switch to the safe option in equilibrium and the other player 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 of Section 3.1. 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 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:

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

where
(5)

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

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

and with p', q' as defined in Equation 1. 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 mechanics of the equilibrium, drawing parallels with the equilibrium in Section 4.1 when pertinent. To this end, we first define some notation (4.2.1) that will then serve to define the equilibrium strategies and illustrate the equilibrium dynamics (4.2.2).

4.2.1 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 case, the belief about the quality of the risky option which is not being activated remains constant over time.

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 $(x,y) \in [0,1]^2$,

$$x \leq B_0(y) \quad \Leftrightarrow \quad R_0(x,y) \leq S(x,y). \tag{6}$$

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 G.1 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.

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 $\{(p,q) : S(p,q) \geq R_0(p,q)\}$. 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.

This is driven by the fact 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 $S(p, B_0^{-1}(p))$.

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 $(x, y) \in [0, 1] \times [y_M, B_0(x_M)]$,

$$x \leq B_1(y) \quad \Leftrightarrow \quad R_1(x, y, \sigma_y^x) \leq S(x, y). \tag{7}$$

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. Finally we note that

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

$$B(y) = \begin{cases} B_0(y) & \text{if } 0 \leq x \leq x_M, \\ B_1(y) & \text{if } x_M \leq x \leq B_0(y_V). \end{cases} \tag{8}$$

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.

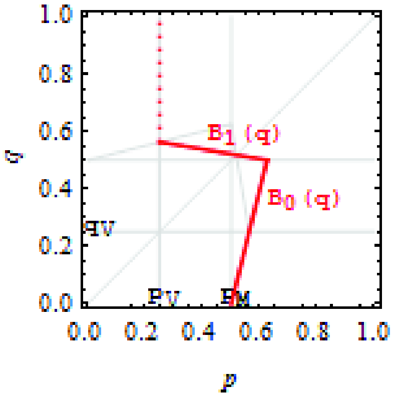


Figure 5: *Illustration of the boundary $B(q)$:*
For $q < q_M$, player i is better-off on the safe option than being forced to experiment until he produces a success whenever: $S(p,q) \geq R_0(p,q) \;\Leftrightarrow p \leq B_0(q)$.
For $q_M \leq q \leq B_0(p_V)$, player i is better-off on the safe option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

4.2.2 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 4.1 when pertinent.

As a first step, (cf. Appendix G.1) 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 8). 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.

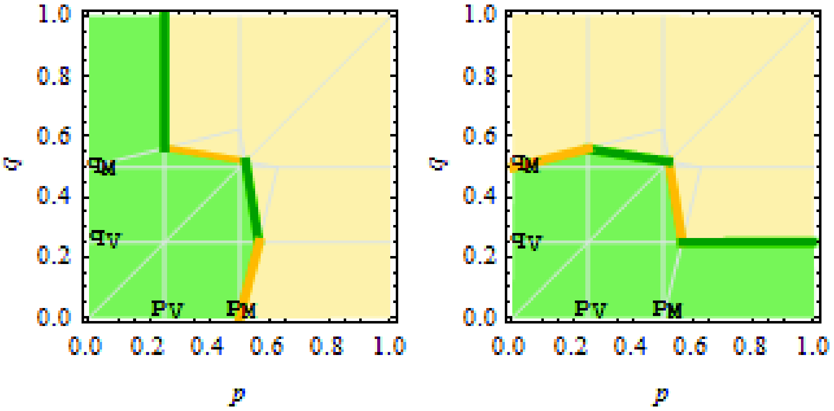


Figure 6: *Equilibrium strategies of player i and player j when exit is revocable. If a state (p, q) is in the green (dark) area, the player plays the safe option, if it is in the orange (light) area, the player plays his risky option.*

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

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

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

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: Cf. Appendix G.□

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.

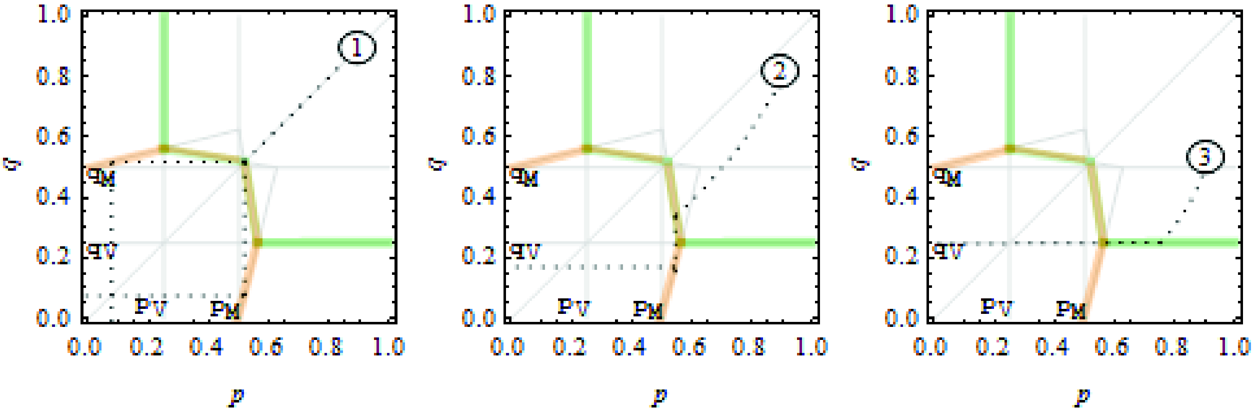


Figure 7: *In Case 1, the prior is* $p_0 = q_0 > \frac{a}{\rho}$. In Case 2, the priors $p_0 > q_0$ are such that at date $t > 0$ satisfying $q_t = B_1(p_t)$ we have $q_t > q_V$. In Case 3, the priors $p_0 > q_0$ are such that at date $t > 0$ satisfying $q_t = q_V$ we have $q_t \geq B_1(p_t)$.

Case 1: $p_0 = q_0 > \frac{a}{\rho}$**.** In equilibrium both players switch to the safe option when beliefs reach the state $p_{\mathrm{U}} = q_{\mathrm{U}}$ satisfying $B_1(p_{\mathrm{U}}) = B_1(q_{\mathrm{U}})$. 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 $p_{\mathrm{U}} > p_M$, and that the player who is not allocated the safe option in the tie-break 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.

5 Conclusion

Motivated by the question of how an agent optimally designs his strategic experimentation when other agents in the market constrain the set of available alternatives, we propose a simple model in which two agents will come to compete over the use of a common option. We have normalised this option to be safe, although we expect the main finding of this paper to hold even if it were risky: When competition for the common option is strong, a player will interrupt his own experimenting and, with the aim of easing the opponent's pressure on the common option, force him to experiment more intensely. Even if this does not succeed, the first player will eventually resume his own experimenting and leave the common option for the opponent to take. This behaviour would be inefficient in the absence of congestion.

The model we propose is simple enough to envisage various extensions. In this paper we study the cases of extreme congestion in that the safe option can only go to one player at a time, and no congestion at all (single-player game). We could extend it to allow for all intermediate levels of congestion by adding a second common option delivering a certain flow payoff $b \in [0, a]$ and varying its expected quality. Moreover, in this paper, we assume that players observe one another's actions and payoffs. Agents then always share their beliefs about the likely success of each player's experimenting. In future research it would be interesting to investigate how agents behave if only actions are observed. Another extension would be to allow the quality of the risky objects to be correlated across players.

6 Appendix

A Single Player - Value Function

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 plying the risky option is optimal.

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 $p \geq p_V$, playing the risky option is optimal and

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

while for $p \leq p_V$, 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 tot he safe option.

B 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 $\mathcal{W}(p,q) = L^{RR}\mathcal{W}(p,q) \geq L^{RS}\mathcal{W}(p,q)$. We obtain the following partial differential equation for the value function:

$$\begin{aligned}(p\lambda + q\lambda + \rho)\ \mathcal{W}(p,q) + p\lambda(1-p)\ \mathcal{W}_p(p,q) + q\lambda(1-q)\ \mathcal{W}_q(p,q) \\ = p\lambda\left[\tfrac{\lambda+\rho}{\rho} + V(q)\right] + q\lambda\left[\tfrac{\lambda+\rho}{\rho} + V(p)\right].\end{aligned} \tag{9}$$

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 $\tilde{\mathcal{W}}(s)$:

(10)

$$\tilde{\mathcal{W}}'(s) - (p(s)\lambda + q(s)\lambda + \rho)\ \tilde{\mathcal{W}}(s) = -p(s)\lambda\left[\frac{\lambda+\rho}{\rho} + V(q(s))\right] - q(s)\lambda\left[\frac{\lambda+\rho}{\rho} + V(p(s))\right]$$

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 $q(s_{\tilde{\mathcal{W}}}) \in [q_{\mathcal{W}}, q_V]$ 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 3.2 in the belief space (p,q) .

C 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:

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

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 $f(p){:=}\frac{2(p\lambda+\rho)}{p\lambda(1-p)}$, $g(p) := \frac{2a}{p\lambda(1-p)} + \frac{2}{(1-p)}\left[\frac{\lambda+\rho}{\rho} + V(p)\right]$ and noticing that $e^{\int f[x]\,dx}\big|_{x=0} = 0$.

Solving, we obtain the following expression for $\mathcal{A}$:

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

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. □

D 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 2The method resembles the one used in Appendix B to derive the set of threshold beliefs in the planner problem with irrevocable exit. For all (p, q), the Bellman equation (3) 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 (10), which is the ODE for $\mathcal{W}$ in the social planner problem with irrevocable exit. In Appendix B we have derived the family of solutions $\mathcal{W}_C$ to that ODE. We obtain the boundary in Figure 2by 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 (10). Solving for $s_{\mathcal{U}}$ we then obtain the boundary in (p, q) space depicted in Figure 2.

As is clear from that figure, for $\Delta > 0$ (where Δ is the time interval over which the planner allocated the options in our discrete-time approximation) the planner will alternate between policies RS and RR, for instance for priors $1 > p_0 \gg q_0 > 0$. But as $\Delta \to 0$, the planner will play so as to stay indifferent between RS and RR, and the state moves "along" the boundary in Figure 2.Our value-matching and smooth-pasting conditions are then in fact identifying an interval of dates over which the planner "alternates" 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

(11)

$$a(\lambda+\rho)(\lambda+2\rho) - (a-\lambda)\lambda^2 p_{\mathcal{U}}^2\left(\frac{1-p_{\mathcal{U}}}{p_{\mathcal{U}}}\frac{p_V}{1-p_V}\right)^{\frac{2(\lambda+\rho)}{\lambda}} =$$
$$p_{\mathcal{U}}\lambda\left(2(\lambda+\rho-a)(\lambda+\rho) + p_{\mathcal{U}}\lambda(a+\rho) + \ (\lambda+2\rho)2(a-\lambda p_V)\frac{p_{\mathcal{U}}}{p_V}\left(\frac{1-p_{\mathcal{U}}}{p_{\mathcal{U}}}\frac{p_V}{1-p_V}\right)^{\frac{(\lambda+\rho)}{\lambda}}\right).$$

E 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

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

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 $dt > 0$. □

F 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 ($k^i_t = k^j_t = 1$), 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 4, $w(p,q)$ 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

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

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:

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

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[7] 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.

[7] 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 by 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 $p_\tau = p_V$).

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[8] 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[9] 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 4.1.

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 $p = p_M$.

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

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

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

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

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

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

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

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

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

G 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 G.1 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 F we obtain:

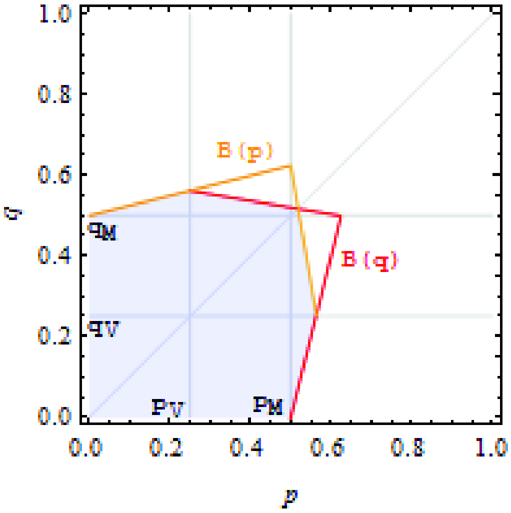
$$\begin{aligned}\tilde{u}(0;\tau) = \;& e^{-\rho\tau}(p_0 e^{-\lambda\tau} + 1 - p_0)(q_0 e^{-\lambda\tau} + 1 - q_0)\left[\tilde{u}(\tau) - p_\tau \frac{\lambda}{\rho} - K\ p_\tau q_\tau \left(\frac{1-p_\tau}{p_\tau}\right)^{\frac{\lambda+\rho}{\lambda}}\right] \\ & + p_0 \frac{\lambda}{\rho} + K\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

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 $p_M < p_0 < 1$, $q_M < q_0 < 1$, 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.

G.1 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 8. Players will have incentives to preempt one another's exit and there will be unraveling of the exit decision.



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 $S(q, p) > R_1(q, p) \ \Leftrightarrow\ q < B_1(p)$.

G.2 Equilibrium

This series of steps in Section G.1 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 j occupy the safe option than facing him in a tie-break or preempting him.

Notice furthermore that the term

$$e^{-\rho\tau}(p_0e^{-\lambda\tau} + 1 - p_0)(q_0e^{-\lambda\tau} + 1 - q_0)\left[S(p_\tau, q_\tau) - p_\tau\frac{\lambda}{\rho} - K\ p_\tau q_\tau\left(\frac{1-p_\tau}{p_\tau}\right)^{\frac{\lambda+\rho}{\lambda}}\right],$$

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 4.2.

In all cases, following the argument in Section G.1,

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

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

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

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

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

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. □

References

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

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.

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

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

Murto, P. and J. Valimaki (2008). Learning and Information Aggregation in an Exit Game. Technical report, working paper, Helsinki Center of Economic Research.

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

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.

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.