Experimentation and Disappointment

Barbara Luppi*
Department of Economics, University of Bologna
Piazza Scaravilli 2, 40126

November 29, 2003

Abstract

We depart from the classic setting of bandit problems by endowing the agent with a disappointment-elation utility function. The disutility of a loss is assumed to be greater than the elation associated with same-size gain, according to Kahneman-Tversky findings on the attitude of agents towards a change in wealth. We characterise the optimal experimentation strategy of an agent in a two-armed bandit problem setting with infinite horizon and we derive an existence theorem, specifying a condition on the disappointment aversion parameter. The model, solved in closed form in a one-armed bandit setting, shows that an agent who feels disappointment experiments more intensively than the agent characterised by the standard expected utility model, despite disappointment, but only if the degree of disappointment is under a certain threshold level. The threshold level depends both on the probability of rewards along the unknown projects relative to the expected number of trials and on the expected reward of the unknown project.

JEL C61, C73, D83

Key words: two and one-armed bandit, disappointment, learning

*E-mail address: bluppi@economia.unibo.it

Footfalls echo in the memory
Down the passage we did not take
T.S. Eliot

1 Introduction

In many areas of human activity, an agent has to choose from a number of actions, each with a cost and an uncertain reward. Some of these actions are highly likely to produce a short-term gain, while others, such as gathering information to eliminate some of the uncertainty, may result in a long-term benefit only.

The classic multi-armed bandit problem is a formalisation of such a situation: in each period the agent pays a cost to pull one of a fixed number of arms, different arms having different, unknown and possibly interdependent pay-off probabilities. The agent's problem is to maximise the expected discounted sum of pay-offs. In bandit problems currently in the economic literature projects are equated with arms (Weitzman,1981, Roberts and Weitzman, 1981).

The present work studies the bandit problem under alternative models of rational behaviour, providing a bridge between the standard optimising analysis of the bandit problem and the modern literature on alternatives to the expected utility model of behaviour under uncertainty, pioneered by Loomes and Sugden, among others.

The question addressed in the paper is the following: suppose the experimenter has psychological feelings that affect his rational choices under uncertainty: the individual receives not only the utility derived directly from the actual consequence of an uncertain prospect, but in addition he feels some degree of disappointment and elation.The agent forms a priori expectation about any uncertain prospect, when evaluating that prospect and, after uncertainty is resolved, the individual compares the actual consequence of the prospect with the a priori expectation: if the actual consequence turns out to be worse than the expectation, he feels disappointment. On the other hand, the individual experiences some degree of elation, if the actual consequence is better than the a priori expectation. Will this agent experiment more in the setting of a bandit problem?

We depart from the classic setting of bandit problems by endowing the agent with a disappointment-elation utility function (Sugden and Loomes, 1986): we assume the agent aims at maximising the expected utility of profits, instead of assuming the maximization of expected profits as an objective function. A particular specification of the disappointment-elation utility function is used to capture the spirit of Kahneman-Tversky findings (1979) on the attitude of agents towards a change in wealth. We assume that the disutility of a loss is greater than the elation associated with a gain of the same amount.

The decisional problem faced by the agent has the same characteristics of the standard one: facing the decision between an unknown arm and a safe one, the agent faces an optimal stopping problem, i.e. he has to decide how many trials to do along the unknown arm and when he finds optimal to switch to the safe one, so that once switched the agent won't find it optimal to choose the unknown arm again. Here the assumption about the agent's utility function modifies the objective function of the experimenter, i.e. how the agent evaluates his payoffs.

In the present paper, we consider an experimenter which is trying to choose between two projects. The agent who does not know with certainty the consequence of a particular project may choose it and observe the result. However such experimental procedure of choice is costly: in the short run, the agent bears the loss in case of a negative result. This loss must be traded off against the potential informational gain associated with experimentation, in terms of a more correct estimate of the probability of having a positive result when undertaking a certain project. In the standard framework, the agent's optimal strategy consists of finding the way of weighting these two opposite forces. Here, the agent must take into account an additional trade-off associated with experimentation: in the short run the agent faces the additional cost due to the sensation of disappointment he has in case the project yields a negative result. The experimenter trades off this cost against the gain in expected terms associated with the psychological feeling of rejoycing he will have by undertaking the project which the present information indicates is most profitable. This additional trade-off plays a key role in characterising the optimal strategy of the agent and deriving the main result of the paper.

In the first part of the paper, we consider two projects with unknown probability of a reward and an infinite horizon and we characterise the optimal strategy of the agent. We derive an existence theorem analogous to the one stated in Rothschild (1974), specifying an additional condition on the parameter capturing disappointment.

In the second part, the decision problem faced by the experimenter is examined using the mathematical tool of dynamic allocation indexes, which allow us to compare the sequential strategy chosen by the agent under expected utility framework and under disappointment theory in terms of experimentation intensity. In this setting we present the central result of the paper- that the agent who is characterised by the disappointment-elation utility model of Loomes and Sugden will choose to experiment more intensively than the consumer characterised by the standard expected utility model, under certain conditions regarding the degree of disappointment. This result appears to be counterintuitive, since one would expect that an agent who weighs more disappointment relative to elation will be less keen on experimenting than an agent characterised by the standard expected utility theory model. Here we show that an agent who feels disappointment experiments more, despite disappointment, but only if the degree of disappointment is under a threshold level. The threshold level depends both on the probability of rewards along the unknown projects relative to the expected number of trials and on the expected reward of the unknown project. The economic intuition driving the result lies in the trade-off explained above: an agent characterized by a disappointment-elation utility function will decide to experiment more, instead of switching to the known project, because he evaluates the possibility of having a sufficiently bad run of luck and the disappointment he could have in the case he sticks forever with the known project, paying the lower expected reward. It is shown that the result holds for n periods and for a generic distribution function F and that when the number of periods n increases, the threshold on the degree of disappointment increases, becoming a less tight constraint.

Finally, we characterise the decision procedure of the experimenter under a specific distributional assumption and within this framework we compare the standard expected utility theory case with the one where the experimenter is characterised by disappointment.

The paper is organised as follows. In Section 2 we overview the related literature, both in two-armed bandit problems and in the alternative frameworks to standard expected utility. In Section 3 we derive an existence theorem of the optimal experimentation strategy of the agent. In Section 4 the result on the intensity of experimentation of the disappointment-adverse agent is stated. In Section 5 we consider a special case of the decision procedure of the experimenter under a specific distributional assumption and a finite horizon time. Finally, Section 6 concludes.

2 Related literature

The departure point of our analysis is the seminal paper on experimentation of Rotschild (1974). He examines the pricing decision of a monopolist facing an unknown stochastic demand. The store can choose between two possible prices (arms), each with an unknown probability to make a sale; each period the monopolist selects the price to charge and he updates his beliefs on the basis of the resulting sales to customers. The monopolist faces the trade-off between the short-term benefit, represented by charging the price that maximises his payoff given his current information and the information gained on the demand at the other price, which has long-term value. Rotschild shows that with positive probability the monopolist will choose the inferior price, i.e. the price less likely to make the sale and therefore, inefficiency may arise in the long run. Rothschild's analysis relates to the multi-armed bandit literature, on which Gittins (1979) and others have worked to characterise the index policy. In the economics literature Weitzman (1981) and Roberts and Weitzman (1981) provide two important applications of bandit problems with independent arms to R&D settings.

Another literature branch relevant to the present paper is the one which examines alternative frameworks to the expected utility model of behaviour under uncertainty. Experimental research into choice under uncertainty has revealed that people behave in ways that systematically violate the set of basic axioms formulated by Von Neumann and Morgenstern (1947) and Savage (1954), upon which the conventional expected utility theory is built. In particular, the empirical evidence presented by Kahneman and Tversky (1979) and others show a number of patterns of choice that reveal behavioural regularities that sistematically contradict the predictions of conventional expected utility. After the discovery of Allais paradox and Ellsberg paradox, many attempts have been done to develop alternative frameworks for the analysis of choice under uncertainty, consistent with the observed behavioural regularities. A subset of them start from an attempt at a psychological explanation of Allais Paradox phenomena. One of the earliest was prospect theory (Kahneman and Tversky, 1979), generalized later by cumulative prospect theory (1992). Two more intuitive and parsimonious psychologically based theories are regret theory (Sugden and Loomes, 1982, 1987) and disappointment theory (Sugden and Loomes, 1986); both of them incorporate ex ante considerations of ex post psychological feelings: of regret, or rejoicing, in the former and of disappointment, or elation, in the latter. "The fundamental idea behind regret theory is that the psychological experience of "having x" can be influenced by comparison between x and y that one might have had, had one chosen differently. If, for example, I bet on a horse which fails to win, I may experience something more than a reduction in my wealth: I may also experience a painful sense of regret arising out of the comparison between my current state of wealth and the state that I would have enjoyed, had I not bet" (Sugden, 1991).

3 The model

Consider an experimenter which is facing the following decision problem. The agent has to decide which project (chosen from a fixed set) to undertake in each period. Here we assume only two projects are available to the agent. Each project pays out a fixed reward at an unknown probability. In the case he knows the true probability of the reward of each project i, Π_i, he would simply choose the project with the highest expected reward. Here we assume the agent does not know Π_i and he can learn the probability of the reward through experimentation.

The generality of the description lends itself to a number of possible interpretations: we can think the agent to be a researcher employed in an R&D department, which has been assigned to the task of finding a more efficient way to produce some commodity and he has to choose among two substitute technologies, each offering uncertain benefits until the development work is completed. Or consider the research work of a PhD student, who has to select the topic of his thesis from a pool of possible interesting areas, each with an uncertain probability to yield a publishable paper[1]. We can interpret the reward yielded by a project as a preliminary positive result in a simplified analytical framework he is working on or a significative t-test yielded by regressions with the new data at the basis of his research. On the other hand, the project can held no result, in the sense of no significative advances toward the goal of the research. The framework fits as well the case of a monopolistic store trying to price its commodity and learning the demand function through experimentation, as the one in Rotschild (1974), where a positive result indicates the event

[1] We obviously abstract from his skills!

of a customer buying the commodity, associated with a return equal to the price of the good net of the cost of production, and symmetrically the negative result represents the event the customer goes away the store without buying the good[2].

In mathematical and statistical literature this problem is known as the two armed bandit problem and it has been widely analysed[3]. Here, we assume that the agent's preferences are described by the following utility function (Sugden and Loomes, 1986):

$U(x_{is}) = x_{is} + D(x_{is} - \overline{x}_i)$

where x_{is} denotes the utility of project i under state of the world s and $\overline{x}_i$ denotes the expected utility of project i.

According to the utility function, the economic agent receives not only the utility derived directly from the actual consequence of an uncertain prospect, but in addition he feels some degree of disappointment and elation. When the agent evaluates a prospect, he forms a priori expectation about any uncertain prospect and after uncertainty is resolved he compares the actual consequence of the prospect with the a priori expectation. If the actual consequence turns out to be worse than the expectation, he feels disappointment. On the other hand, the individual experiences some degree of elation, if the actual consequence is better than the a priori expectation.

We assume x_{is} to be linear and exactly equal to the payoff in state s of project i and the disappointment-elation utility function $D(\cdot)$ to take the following functional form:

$$D(\xi) = \begin{matrix} \xi, & \xi \geq 0 \\ b\xi, & \xi < 0, b > 1 \end{matrix}$$

The utility function is defined on deviations from the reference point[4], i.e. the utility function is kinked at the origin. The parameter b captures the intensity of disappointment aversion and we require the utility function to be steeper for losses than for gains. This second characteristic reflects a salient feature of attitudes to changes in welfare, that is the disappointment that one experiences in losing a sum of money appears to be greater than the pleasure associated with gaining the same amount. This specification of the $D(\cdot)$ function captures the spirit of the Kahneman-Tversky descriptive theory of judgement under uncertainty (Kahneman and Tversky, 1979)[5][6]. The choice of this utility function allows us to investigate the effect of a psychological attitude, as disappointment

[2] Note that since we endow the agent with a utility function, not all economists would be satisfied with this intepretation

[3] See Gittins (1979), Berry and Fristedt (1985)

[4] The reference point is represented by the origin, since we do not endow the agent with any legacy or other forms of wealth inherited from the past.

[5] Here, for simplicity, we do not introduce Kahneman and Tversky's assumption according to which the utility function for changes of wealth is normally concave above the reference point ($D"(x) < 0$, for $x > 0$) and often convex below it ($D"(x) > 0$, for $x < 0$). The economic meaning of this assumption is that the marginal value of both gains and losses generally decreases with their magnitude.

[6] The parameter b is defined in an analogous way to the coefficient of loss aversion in Kahneman and Tversky (1979, 1992). Empirical studies conducted by Kahneman and Tversky indicate that the best estimate of the coefficient of loss aversion is 2,25.

aversion, on the experimentation strategy optimally chosen by the experimenter.

3.1 The analytical set-up

We consider two projects. The generic project i pays off 1 with probability Π_i and 0 with probability $(1 - \Pi_i)$. The experimenter does not know the true probability Π_i. He decides which project to undertake according to his prior beliefs about the parameter Π_i and selects the one with the highest probability of a reward. We assume the agent has an infinite horizon.

Let N_i be the number of trials along project i and s_i be the number of positive results along i, with $i = 1, 2$.

The information in the sample can be represented by using the statistics defined as follows:

$$\rho_i = \frac{1}{1 + N_i}$$

$$\mu_i = \frac{s_i}{1 + N_i}{}^{7}$$

When the agent chooses project i, ρ_i becomes $\dfrac{\rho_i}{1 + \rho_i}$. In case of a positive result on i, the statistics μ_i becomes:

$$s\left(\mu_i\right) = \frac{\mu_i + \rho_i}{1 + \rho_i}$$

while in case of a negative result, μ_i becomes:

$$f\left(\mu_i\right) = \frac{\mu_i}{1 + \rho_i}$$

Therefore, the information contained in the sample is given by $(\mu_1, \mu_2, \rho_1, \rho_2)$, which is a subset of R^4, constitued by a fourfold copy of the closed unit interval $[0, 1]$. Denote the domain of (μ, ρ) as Δ.

The experimenter's belief about the parameters π_i with $i = 1, 2$ are given by the prior density function $g\left(\pi_1, \pi_2\right)$. We do not assume the independence of the probabilities of success along the two projects, but we assume:

$$g\left(\pi_1, \pi_2\right) > 0 \quad \text{for all } (\pi_1, \pi_2) \in (0, 1) \times (0, 1)$$

excluding all nonextreme combinations of $\Pi_1$and Π_2.

The experimenter with experience (μ, ρ) updates his prior beliefs from $g\left(\pi_1, \pi_2\right)$ to $h\left(\pi_1, \pi_2, \mu, \rho\right)$. The probability density is proportional to

$$\pi_1^{\frac{\mu_1}{\rho_1}} \left(1 - \pi_1\right)^{\frac{[1-(\mu_1+\rho_1)]}{\rho_1}} \pi_2^{\frac{\mu_2}{\rho_2}} \left(1 - \pi_2\right)^{\frac{[1-(\mu_2+\rho_2)]}{\rho_2}} g\left(\pi_1, \pi_2\right)$$

The posterior mean of the experimenter's belief about the parameter Π_i given the sample information (μ, ρ) and the prior density function g is defined as follows:

$$\lambda_i\left(\mu, \rho\right) = \int_0^1 \int_0^1 \pi_i h\left(\pi_1, \pi_2, \mu, \rho\right) d\pi_1 d\pi_2{}^{8}$$

[7] As the number of trials increases, μ_i approaches the sample mean $\overline{\mu}_i = \dfrac{s_i}{N_i}$, i.e. $\lim_{\rho_i \to 0} \mu_i = \lim_{T_i \to \infty} \overline{\mu}_i$

[8] $\lambda_i\left(\mu, \rho\right)$ is defined and continuous $\forall\left(\mu, \rho\right)$ such that $\rho_i > 0$, $i = 1, 2$. It is possibly to show that $\lambda_i\left(\mu, \rho\right)$ can be extended by continuity to $[0, 1]^4$, since

3.2 Dynamic programming equations and properties

The experimenter maximizes the expected discounted utility of his rewards over the infinite horizon. The problem can be written in terms of dynamic programming equations, satisfying the following functional form[9]:

$$V(\mu,\rho) = \max_{\{i\}} W_i(\mu,\rho)$$

where

$$W_i(\mu,\rho) = \lambda_i(\mu,\rho)\left[1 + D\left(1 - \lambda_i(\mu,\rho)\right)\right] + \left(1 - \lambda_i(\mu,\rho)\right) D\left(-\lambda_i(\mu,\rho)\right) + \\ \delta\left[\lambda_i(\mu,\rho) V\left(\sigma_i(\mu,\rho)\right) + \left(1 - \lambda_i(\mu,\rho)\right) V\left(\psi_i(\mu,\rho)\right)\right]$$

where $0 < \delta < 1$ and $\sigma_i(\mu,\rho)$ and $\psi_i(\mu,\rho)$ are vectors indicating the state of information after respectively a positive or a negative result on i.

Define the functions $V^t(\mu,\rho)$ and $W_i^t(\mu,\rho)$ as follows:
$V^0(\mu,\rho) = 0$
$V^t(\mu,\rho) = \max_{\{i\}} W_i^t(\mu,\rho)$
where
$W_i^t(\mu,\rho) = \lambda_i(\mu,\rho)\left[1 + D\left(1 - \lambda_i(\mu,\rho)\right)\right] + \left(1 - \lambda_i(\mu,\rho)\right) D\left(-\lambda_i(\mu,\rho)\right) + \delta\left[\lambda_i(\mu,\rho) V^{t-1}\left(\sigma_i(\mu,\rho)\right) + \left(1 - \lambda_i(\mu,\rho)\right) V^{t-1}\left(\psi_i(\mu,\rho)\right)\right]$

Lemma 1 *The functions $V^t(\mu,\rho)$ and $W_i^t(\mu,\rho)$ are continuous*

Proof. The proof is given by induction. We know that
$V^0(\mu,\rho) = 0$
$V^1(\mu,\rho) = \max_{\{i\}} W_i^t(\mu,\rho)$
where:
$W_i^1(\mu,\rho) = \lambda_i(\mu,\rho)\left[1 + D\left(1 - \lambda_i(\mu,\rho)\right)\right] + \left(1 - \lambda_i(\mu,\rho)\right) D\left(-\lambda_i(\mu,\rho)\right)$
since $V^0(\mu,\rho) = 0$.

Given the continuity of $\lambda_i(\mu,\rho)$ and of the disappointment-utility function $D(\cdot)$, $W_i^1(\mu,\rho)$ is continuous.

Let us suppose that $V_i^{t-1}(\mu,\rho)$ and $W_i^{t-1}(\mu,\rho)$ are continuous. By the definition of $W_i^t(\mu,\rho)$, we have:
$W_i^t(\mu,\rho) = \lambda_i(\mu,\rho)\left[1 + D\left(1 - \lambda_i(\mu,\rho)\right)\right] + \left(1 - \lambda_i(\mu,\rho)\right) D\left(-\lambda_i(\mu,\rho)\right) + \delta\left[\lambda_i(\mu,\rho) V^{t-1}\left(\sigma_i(\mu,\rho)\right) + \left(1 - \lambda_i(\mu,\rho)\right) V^{t-1}\left(\psi_i(\mu,\rho)\right)\right]$

which is continuous since it is the sum of continuous functions. The same argument applies to $V_i^t(\mu,\rho)$, since:
$V_i^t(\mu,\rho) = \max_{\{i\}} W_i^t(\mu,\rho)$ ■

Lemma 2 *The functions $V^t(\mu,\rho)$ and $W_i^t(\mu,\rho)$ are monotonic*

$\lim_{\rho_i \to 0} \lambda_i(\mu,\rho) = \mu_i$
by the law of large numbers

[9] From now on we use the general specification of the disappointment-elation function $D(\cdot)$ and we consider the specific functional form assumed only when necessary to derive the result.

Proof. The proof is given by induction.

First, we show that $V^1 \geq V^0$ and $W_i^1 \geq W_i^0 = 0$.

We know that $V^0(\mu,\rho) = 0$. Therefore, we need to show that $V^1 \geq 0$. By simply applying the formula,

$W_i^1(\mu,\rho) = \lambda_i(\mu,\rho)\left[1 + D(1-\lambda_i(\mu,\rho))\right] + (1-\lambda_i(\mu,\rho))\, D(-\lambda_i(\mu,\rho)) \geq 0$

that is,

$W_i^1(\mu,\rho) = \lambda_i(\mu,\rho)\left[2-\lambda_i(\mu,\rho)\right] + (1-\lambda_i(\mu,\rho))(-b\lambda_i(\mu,\rho)) \geq 0$

iff $b \leq \dfrac{2-\lambda_i(\mu,\rho)}{1-\lambda_i(\mu,\rho)}$ since $\lambda_i(\mu,\rho) \in [0,1]$.

Let us assume that $V_i^{t-1}(\mu,\rho)$ is monotone increasing and show the monotonicity of $V_i^t(\mu,\rho)$.

By the definition of $W_i^t(\mu,\rho)$, we have:

$W_i^t(\mu,\rho) = \lambda_i(\mu,\rho)\left[1 + D(1-\lambda_i(\mu,\rho))\right] + (1-\lambda_i(\mu,\rho))\, D(-\lambda_i(\mu,\rho)) + \delta\left[\lambda_i(\mu,\rho)\, V^{t-1}(s_i(\mu,\rho)) + (1-\lambda_i(\mu,\rho))\, V^{t-1}(f_i(\mu,\rho))\right]$

which is monotone increasing, under the condition stated on the parameter of disappointment aversion b, since it is the sum of monotone increasing functions. The same argument applies to $V_i^t(\mu,\rho)$, since:

$V_i^t(\mu,\rho) = \max_{\{i\}} W_i^t(\mu,\rho)$ ■

Lemma 3 *$V^t(\mu,\rho)$ and $W_i^t(\mu,\rho)$ converge uniformly to $V(\mu,\rho)$ and $W_i(\mu,\rho)$ respectively*

Proof. In order to show the uniform convergence of $V^t(\mu,\rho)$ and $W_i^t(\mu,\rho)$, we need to show that exists a majorant M, greater than the maximum of $\lambda_i(\mu,\rho)\left[1 + D(1-\lambda_i(\mu,\rho))\right] + (1-\lambda_i(\mu,\rho))\, D(-\lambda_i(\mu,\rho))$. Note that M exists, since $\lambda_i(\mu,\rho) \in [0,1]$, $D(1-\lambda_i(\mu,\rho)) \in (0,1)$, $D(-\lambda_i(\mu,\rho)) \in R^- - \{-\infty\}$[10] and $\lim_{\rho_i \to 0} \lambda_i(\mu,\rho) = \mu_i$. Then, it has to be the case that $V_i^t(\mu,\rho) \leq M\sum_{\tau=0}^{t}\delta^\tau \leq M\sum_{\tau=0}^{\infty}\delta^\tau = \dfrac{M}{1-\delta}$

The sequences $V^t(\mu,\rho)$ and $W_i^t(\mu,\rho)$ are bounded above and converge respectively to $V(\mu,\rho)$ and $W_i(\mu,\rho)$, which are defined as follows:

$V(\mu,\rho) = \lim_{t\to\infty} V^t(\mu,\rho)$

$W_i(\mu,\rho) = \lim_{t\to\infty} W_i^t(\mu,\rho)$

Indicate with $\overline{V}^t(\mu,\rho)$ the present discounted value of the expected utility of the sum of rewards from the first t periods when following a policy described by the system of dynamic programming equations. Then, $V^t(\mu,\rho) \geq \overline{V}^t(\mu,\rho)$. Then, $V(\mu,\rho) \leq \overline{V}^t(\mu,\rho) + \delta^t\sum_{\tau=t}^{\infty}\delta^\tau M \leq V^t(\mu,\rho) + \delta^t\dfrac{M}{1-\delta}$, which implies $|V(\mu,\rho) - V^t(\mu,\rho)| \leq \delta^t\dfrac{M}{1-\delta}$. Since the previous inequality is independent of (μ,ρ), $V^t(\mu,\rho)$ converges uniformly to $V(\mu,\rho)$. A similar argument for uniform convergence applies to $W_i^t(\mu,\rho)$. ■

[10] given the condition on the parameter b stated in Lemma 2

Proposition 4 *$V(\mu,\rho)$ and $W_i(\mu,\rho)$ are continuous on Δ*

Proof. This follows from Lemma 1 and 3. ■

Define $A_i = \left\{(\mu,\rho) \in [0,1]^4 : W_i(\mu,\rho) > W_j(\mu,\rho)\right\}$ the information set such that project i is optimally chosen.

Lemma 5 *A_i is an open set*

Proof. This follows from Lemma 1. ■

3.3 Main results

Theorem 6 *If the true parameters satisfy $0 < \Pi_1 < \Pi_2 < 1$, and the disappointment aversion parameter satisfies $b \le \dfrac{2-\lambda_i(\mu,\rho)}{1-\lambda_i(\mu,\rho)}$, then an agent who follows an optimal strategy will with positive probability choose project 1 infinitely often and project 2 only a finite number of times.*

The proof of the theorem is given in an analogous way to the one stated in Rothschild (1974). Lemma 7 shows that the agent while experimenting will find optimal to undertake project 1 if he has a very bad run of trials along project 2, even if project 1 is the inferior one. The main difference in the proof lies in the additional condition on the disappointment aversion parameter b, characterizing the agent. This condition tells us that the agent will implement the same optimal experimentation strategy of a standard expected utility maximising agent under the condition that the intensity of disappointment aversion he feels is not "too strong". Note that the threshold value under which the agent is going to follow this experimentation strategy is endogenously determined by the beliefs on the likelihood of the generic project i to pay out the reward, $\lambda_i(\mu,\rho)$, updated in each period on the basis of the state of information (μ,ρ). The surprising finding relies in the fact that, for every value of $\lambda_i(\mu,\rho)$, the condition satisfies the asymmetry on the weights given to losses and gains, i.e. for every value of $\lambda_i(\mu,\rho)$, the agent is allowed to weight consistently more losses relative to gains and nonetheless he will follow the same experimentation strategy as a standard EUT maximiser.

Lemma 7 *Under the condition $b \le \dfrac{2-\lambda_i(\mu,\rho)}{1-\lambda_i(\mu,\rho)}$ and for every δ_1,δ_2, such that $\Pi_1 > \delta_1 > 0$ and $0 < \delta_2 < 1$, there exists $\epsilon > 0$ such that $W_1(\mu,\rho) > W_2(\mu,\rho)$ whenever $\mu_2 + \rho_2 < \epsilon$ and either $\mu_1 \ge \Pi_1 - \delta_1$ or $\rho_1 \ge \delta_2$.*

Proof. Consider the compact set:

$K = \{(\mu_1,\mu_2,\delta_1,\delta_2) \in \Delta \mid \mu_1 \ge \Pi_1 - \delta_1 > 0 \text{ or } \rho_1 \ge \delta_2 \text{ and } \mu_2 = \rho_2 = 0\}$

We want to show the conditions under which $K \subset A_1$. Consider:

$W_1(\mu,0,\rho,0) \ge \lambda_1(\mu,0,\rho,0)\left[1 + D(1-\lambda_1(\mu,0,\rho,0))\right] + (1-\lambda_1(\mu,0,\rho,0))\, D(-\lambda_1(\mu,0,\rho,0)) > 0$

From Lemma 2, the second inequality holds iff:

$b < \dfrac{2 - \lambda_i (\mu, 0, \rho, 0)}{1 - \lambda_i (\mu, 0, \rho, 0)}$
while if
$W_2 (\mu, 0, \rho, 0) \geq W_1 (\mu, 0, \rho, 0)$
then
$V (\mu, 0, \rho, 0) = W_2 (\mu, 0, \rho, 0) = 0 [1 + D (1)] + 1 [0 + D (0)] + \delta [0V (\mu, 0, \rho, 0) + 1V (\mu, 0, \rho, 0)]$
$\Longrightarrow V (\mu, 0, \rho, 0) = \delta V (\mu, 0, \rho, 0)$
This equality holds only if $V (\mu, 0, \rho, 0) = 0 < W_1 (\mu, 0, \rho, 0)$, a contradiction.

From Lemma 4, A_1 is an open set, centered at any (μ, ρ). Therefore, K is an open ball contained in A_1. Let us indicate with $B_r (x)$ the generic ball centered at point x with radius r. Then, there is a finite number J of balls covering the set A_1, such that $K \subset \bigcup_{j \in J} B_{r_j} (x_j)$.

Let us denote with $\epsilon = \min_{\{j \in J\}} r_j$. It has to be the case that any point $(\mu_1, \mu_2, \rho_1, \rho_2)$ such that $\mu_2 + \rho_2 < \epsilon$ and either $\mu_1 \geq \Pi_1 - \delta_1$ or $\rho_1 \geq \delta_2$ belongs to a open ball $B_{r_j} (x_j) \in A_1$. ■

For the second lemma needed to establish the theorem, refer to Rothschild (1974).

As in Rothschild, the proof of Theorem 1 is obtained without using the condition $\Pi_1 < \Pi_2$. This means that there exists a support in the probability distribution such that the experimenter will find optimal to choose to undertake project 1 only a finite number of times and project 2 infinitely often. This means that inefficiency may arise in the long run, since nothing guarantees even in the long term that the experimenter will end up choosing the arm more likely to pay out the reward. This result holds even when we characterise the preferences of the agent in terms of the disappointment-elation utility function.

In the following proposition we establish the limiting conditions on the parameter of disappointment aversion, b.

Proposition 8 *As the number of trials on project i increases, the condition on the parameter b becomes:*
$b \leq \dfrac{2 - \mu_i (\mu, \rho)}{1 - \mu_i (\mu, \rho)}$ *as* $\rho_i \to 0$
$b \leq 2$ *as* $\mu_i \to 0$ *and* $\rho_i \to 0$
$b \leq \infty$ *as* $\mu_i \to 1$ *and* $\rho_i \to 0$

Proof. The proof is straightforward ■

From Proposition 8 two are the main observations. First, b is allowed to assume values greater than 1, consistently with the Kahneman and Tversky psychological findings of a higher disutility associated with same size losses relative to gains. Secondly, and more surprisingly, even if the experience on project i is very poor, the condition on the disappointment parameter b never violates the requirement on the utility function.

4 The model with the Dynamic Allocation Index

In this section the problem of the optimal experimentation of the agent is handled by using the analytical tool of the dynamic allocation indexes. Gittins (1979) showed that the optimal policy in the framework of multi-armed bandit problems can be described in terms of the so called dynamic allocation index: if the arms are independent (that is pulling one arm is uninformative about other arms) then it is possible to attach to each arm an index, which depends only on the current state of information on that arm. According to the optimal strategy, the experimenter will find optimal to choose the arm with the highest index. This index acts therefore as a reservation value. The tractability of the problem with dynamic allocation indexes allows us to determine a closed form for the reservation value the agent assigns to each project, and in turns to make explicit quantitative comparison in terms of how much experimentation the agent will undertake when characterised by a disappointment-elation utility function, as defined in Section 3.

Within this framework we determine the condition on the parameter of disappointment aversion b under which the agent characterised by the disappointment-elation utility model of Loomes and Sugden will choose to experiment more intensively than the consumer characterised by the standard expected utility model.

4.1 The analytical set-up

Consider an agent that faces the decision between a project that pays off a reward with an uncertain probability and a "safe" one. Assume the first project X_1 has a binomial distribution with $p = \Pr(X_1 = 1)$ unknown but selected by a known priori distribution, F, while the second X_2 has a binomial distribution with a known parameter, q. In other terms, the experimenter knows that the second project pays off q with certainty, while with project 1 he gets a reward equal to 1 with probability $p = \Pr(X_1 = 1)$, which is drawn from a known (general) distribution $F(p)$. In statistical terms, the problem is reduced to a one armed bandit problem.

The experimenter is allowed to play n times and his objective is to determine the sequential strategy which will maximise the expected utility of rewards of the unsafe project (out of n independent observations). We consider n-horizon uniform discounting. The problem is therefore reduced to an optimal stopping one, in which the experimenter has to decide in each period whether to go on with experimentation along the unknown project or to undertake the safe one.

The agent's preferences are described by the disappointment-elation utility function (Sugden and Loomes, 1986), as defined in the above section.

We characterize the optimal experimentation strategy in terms of the dynamic allocation index or Gittins index, indicated as $Q(n, F)$, which is a function of n trials remaining and F the a priori distribution of p at that time.

According to the optimal strategy, the experimenter will choose the safe project if q[11] $> Q(n, F)$ and to undertake project 1, yielding an uncertain reward, in the opposite case. The optimal strategy followed by the experimenter takes the form of an optimal stopping rule.

According to Bradt, Johnson and Karlin (1956), within this analytical setup, the optimal strategy has the following form for the appropriate k_i:

1. Observe the result on the unknown project X_1 until a failure occurs.

2. There exists an integer $k_1 \geq 0$ such that if at least k_1 positive results preceded the first negative one, continue with X_1; otherwise switch to the safe project X_2 for the remaining trials.

3. There is an integer $k_2 \geq 0$ attached to the second negative result such that if at least $k_1 + k_2$ positive results with X_1 precede the second negative one of X_1, continue with X_1; otherwise switch to X_2 for the remaining trials.

4. In general, let S_r be the number of positive results that precede the r-th negative one of X_1. If $S_r \geq k_1 + k_2 + ... + k_r$, continue with X_1; otherwise switch to X_2 for the remaining trials.

Thus, any sequence $k = (k_1, k_2, ..., k_n)$ of integers, $0 \leq k_i \leq n$, corresponds to a strategy of the same form as the optimal.

Let E_k denote expectation given k and $W_n(F, q)$ denote the expected value of utility of the n observations against a priori distribution function F on project 1 and a given parameter q on project 2, pursuing an optimal strategy. In using any strategy for n trials, X_1 will be used a certain number, N_x, of times, and there will be used a certain number, S_x, of positive results with X_1; similarly for X_2.

In the following Proposition, a closed formula for the dynamic allocation index $Q(n, F)$ is determined. We use a specification of the disappointment-elation utility function, where $\overline{c}$ denotes the a priori expected reward yielded by project 1 calculated when n periods of experimentation are available to the agent and against a priori distribution function F. In calculating the explicit formula taken by the index in Proposition 9, we do not allow the agent to update $\overline{c}$ after each period of experimentation- this implies that the evaluation of the disappointment and elation is made against the a priori expected reward of project 1 and not with respect to the updated expected reward according to the new beliefs formed after each trial[12]. This simplifying assumption allows us to obtain a well-defined closed formula for $Q(n, F)$. Nonetheless, it does not play a key role in the following analysis, since it can be shown that the expression of $Q(n, F)$ given in the following proposition underestimates the expected utility associated with experimentation and can be roughly interpreted

[11] Note that along the safe project the agent feels neither elation nor disappointment

[12] In the previous model, with dynamic programming equations, the agent evaluates disappointment and elation by taking into account in each period the beliefs correctly formed after each trial

as a downside boundary to the expected utility associated with undertaking the unsafe project[13]. Intuitively, the agent who can update his current evaluation of the expected reward of project 1 according to the beliefs formed after each period of experimentation can implement a "better" strategy and therefore reach in expected terms higher payoffs level associated with each strategy. Allowing the agent to update $\overline{c}$ in each period according to his beliefs does not undermine the result stated in the next section on the parameter of disappointment aversion, b.

In the following Proposition, the explicit formula of the dynamic allocation index associated with the unsafe project is given.

Proposition 9 $Q(n,F) = \max\limits_k \left\{ \frac{E_k[S_x]}{E_k[N_x]}(1+D(1-\overline{c})) + \left(1-\frac{E_k[S_x]}{E_k[N_x]}\right) D(-\overline{c}) \right\}$

Proof. At the boundary, $q = Q(n,F)$. This implies that $nq = W_n(F,q)$. Since the optimal strategy is defined in terms of a sequence of k, it has to be the case that: $nq = \max\limits_k \{E_k[S_{x_1}](1+D(1-\overline{c})) + (E_k[N_{x_1}] - E_k[S_{x_1}]) D(-\overline{c}) + E_k[N_{x_2}]q\}$, where $\overline{c} = \int_0^1 pdF$. Note that neither $E_k[S_{x_1}]$ nor $E_k[N_{x_1}]$ depend on q. Moreover, neither $E_k[S_{x_1}]$ nor $E_k[N_{x_2}]$ depend on the way the agent evaluates his payoffs. Note $E_k[N_{x_1}] = n - E_k[N_{x_2}]$. Hence, $q = Q(n,F)$ implies $nq = \max\limits_k \{E_k[S_{x_1}](1+D(1-\overline{c})) + (E_k[N_{x_1}] - E_k[S_{x_1}]) D(-\overline{c}) + E_k[N_{x_2}]q\}$

$\Leftrightarrow q \geq \frac{E_k[S_{x_1}]}{E_k[N_{x_1}]}(1+D(1-\overline{c})) + \left(1-\frac{E_k[S_{x_1}]}{E_k[N_{x_1}]}\right) D(-\overline{c})$ for all k.

For some k, it holds with equality:

$q = \max\limits_k \left\{ \frac{E_k[S_{x_1}]}{E_k[N_{x_1}]}(1+D(1-\overline{c})) + \left(1-\frac{E_k[S_{x_1}]}{E_k[N_{x_1}]}\right) D(-\overline{c}) \right\}$ ■

According to Proposition 9, the agent attaches to the unsafe project an index $Q(n,F)$, which can be interpreted as the expected payoff yielded by the unsure project. Note that $Q(n,F)$ depends on the expected success to trial ratio and on the evaluation of project payoffs in case of a positive and a negative result.

4.2 Main results

In this section, we address the question whether an agent characterized by a psychological sensation of disappointment experiments more than a standard expected utility maximizer agent. According to the literature on the armed bandit problem, the natural benchmark is provided by the framework in which the agent is characterised by the standard expected utility paradigm. In the standard optimising literature on bandit problem, it is assumed that the agent is risk neutral and maximises expected profits; we consider, therefore, the case where the utility function of the agent is linear and exactly equal to the payoff[14].

[13] As it is shown explicitly in the following section under a specific distributional assumption

[14] This choice is motivated for homogeneity with the standard economic literature on bandit problems

Proposition 10 *An agent characterised by a disappointment-elation utility function experiments more than a standard EUT maximising agent when the following condition on the disappointment aversion parameter b is satisfied:*

$$b \leq \min\left[\frac{E_k\left[S_x\right](1-\overline{c})}{\overline{c}\left[E_k\left[N_x\right]-E_k\left[S_x\right]\right]}, k+\frac{E_k\left[S_x\right](1-\overline{c})}{\overline{c}\left[E_k\left[N_x\right]-E_k\left[S_x\right]\right]}\right]$$

Proof. The agent will decide to experiment more when characterized by disappointment when the following inequality is satisfied:

$Q^{D}(n,F) \geq Q^{EUT}(n,F)$

where:

$Q^{D}(n,F)$ is calculated according to Proposition 9.

$Q^{EUT}(n,F) = \max\limits_{k}\left\{\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}\right\}$[15]

Under the specific assumption of piecewise linear disappointment-elation utility function, this is equivalent to check under which conditions the following inequality holds:

$$\max_{k}\left\{\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}(2-\overline{c})+\left(1-\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}\right)-b\overline{c}\right\} \geq \max_{k}\left\{\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}\right\}$$

Two possible cases arises.

1. $\max\limits_{k}\left\{\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}\right\}$ coincide in both expressions. In this case, the previous inequality simplifies to $-b\overline{c}+\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}\left\{1-\overline{c}-(-b\overline{c})\right\} \geq 0 \Rightarrow b \leq \frac{E_k\left[S_x\right](1-\overline{c})}{\overline{c}\left[E_k\left[N_x\right]-E_k\left[S_x\right]\right]}$,

2. $\max\limits_{k}\left\{\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}\right\}$do not coincide in the expressions of $Q^{D}(n,F)$ and $Q^{EUT}(n,F)$

Let us indicate $\overline{\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}} = \max\limits_{k}\left\{\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}\right\}$ under EUT. In this case, the previous inequality becomes:

$\left[\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}-\overline{\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}}\right]+-b\overline{c}+\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}\left\{(1-\overline{c})-(-b\overline{c})\right\} \geq 0 \Rightarrow b \leq k+\frac{E_k\left[S_x\right](1-\overline{c})}{\overline{c}\left[E_k\left[N_x\right]-E_k\left[S_x\right]\right]}$

where $k = \frac{\left[\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}-\overline{\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}}\right]}{\overline{c}\left[E_k\left[N_x\right]-E_k\left[S_x\right]\right]}$ ■

In the standard framework of expected utility theory, the agent faces a well-known trade-off, according to which in the short run he bears the loss in case of a negative result. This loss must be traded off against the potential informational

[15] See Karlin, Bradt and Johnson (1956)

gain associated with experimentation, in terms of a more correct estimate of the probability of having a positive result when undertaking a certain project. The agent's optimal strategy consists of finding the way of weighing these two opposite effects associated with the experimental procedure.

When introducing disappointment, the agent must take into account an additional trade-off associated with experimentation: in the short run the agent faces the additional cost due to the sensation of disappointment he has in case the project yields a negative result. The experimenter trades off this cost against the gain in expected terms associated with the psychological feeling of rejoyicing he will have by undertaking the project which present information indicates is most profitable. This additional trade-off plays a key role in characterising the optimal strategy of the agent and determines the condition on the disappointment aversion parameter.

The condition stated on the parameter b can be written as follows[16] to enlight the existence of an additional trade-off for the agent:

$$-b\overline{c}+\frac{E_k\left[S_x\right]}{E_k\left[N_x\right]}\left\{1-\overline{c}-(-b\overline{c})\right\}\geq 0$$

where the first term represents the loss associated with the disappointment sensation in the short run caused by a negative result when undertaking project 1. The second term represents the gain associated with the elation sensation in the long run given by the "real" elation sensation of the agent because of a positive result and the elation sensation due to not feeling disappointment on that occasion. This overall elation sensation is weighted by the success to failure ratio (in expected terms).

The economic intuition driving the result lies in the trade-off explained above: an agent characterized by a disappointment-elation utility function will decide to experiment more, instead of switching to the safe project, because he evaluates the possibility of having a sufficiently bad run of luck and the disappointment he could have in the case he sticks forever with the known project, paying the lower expected reward.

In other terms, the agent will experiment more under disappointment if the intensity of disappointment aversion captured by the parameter b is lower than the ratio of the expected elation to the expected disappointment.

5 An example: the uniform distribution

Here we consider a special case of the decision procedure of the experimenter presented in the previous section. We fully characterize the decision rule followed by the experimenter under a specific distributional assumption of the parameter p. In particular, we assume p is distributed according to a uniform on the support $[0,1]$, i.e. $p=\Pr(X_1=1)\sim f=U(0,1)$. We allow the agent to experiment under a finite horizon time.

[16] The condition is stated referring to case 1, according to the Proof in Proposition 10, which is the more interesting from an economic point of view

We provide a full characterisation of his decision procedure both in the standard framework of expected utility and in the one with disappointment. We derive an explicit comparison between the two decision procedure in terms of the parameter of disappointment aversion, b.

5.1 The Expected Utility Theory framework

We consider an agent characterized by the standard expected utility function. In particular, we assume the utility function to be linear and equal to the payoff.

Suppose $n = 3$, i.e. the agent has three more periods of experimentation.

The individual will decide at $n = 3$ whether to undertake the project 1 or to choose the safe one. Each time he takes the decision whether to experiment or not, he faces the trade-off between a sure payoff equal to q yielded by the safe project and the cost of experimenting by investing in project 1, that is represented by the loss in the case of a negative result.

The individual will calculate the boundary at which he is indifferent between experimenting and not experimenting. This is analogous to determine the dynamic allocation index attached to project 1.

The case $n = 3$ is an interesting one, since it allows to show how the optimal strategy specified in terms of a sequence of k_i works. Since the horizon time is $n = 3$, we should distinguish between two different potential optimal strategies in terms of k_1 in the calculation of the boundary: k_1,the number of positive results realized before a a negative one required by the agent in order not to switch to the safe project, can either take the value 1 or 2 when $n = 3$. The agent chooses the strategy that maximises the value of the dynamic allocation index, i.e. he selects the experimentation strategy in terms of k_1 that maximises the expected payoff under experimentation.

Proposition 11 *Under $n = 3$ and $p \sim U(0, 1)$, the decision rule is summarized as follows:*

Undertake project 1 if and only if $q \leq \dfrac{13}{22}$

If it yields a negative result, undertake project 1 if and only if $q \leq \dfrac{3}{8}$

If it yields a negative result again, undertake project 1 if and only if $q \leq \dfrac{1}{4}$

The dynamic allocation index is determined by applying the following procedure. In order to be indifferent between experimenting and not, the agent must equate the expected payoffs under two different strategies: the strategy under which he chooses the sure project for all three periods and the strategy under which he undertakes project 1 in the first period and he switches to the safe one only in the case he experiences a negative result in the first run of trials under the specification of $k_1 = 2$. In analytical terms[17]:

[17] We leave the indication of the zero payoff to ease the comparison with the boundary in the case the agent evaluates his payoff according to the disappointment elation utility function.

$$3q =$$
$$1p+0\left(1-p\right)+1p^2+0p\left(1-p\right)+q\left(1-p\right)+1p^3+0p^2\left(1-p\right)+qp\left(1-p\right)+q\left(1-p\right)$$

i.e. the sum of the sure payoff of the safe project undertaken for three periods must be equal at the boundary to the sum of the expected payoffs in the case the agent undertakes project 1 in the first period and then follows the strategy specifying $k_1 = 2$. This means that in the second run of trial the individual sticks with the unsure project, in case of a positive result in the first period, while he switches to the safe project, with a sure payoff equal to q, in case of a negative result. Again, in the third period, he chooses to experiment by undertaking project 1 in the case of two consecutive positive results in the previous two trials, while he chooses to switch to the safe project otherwise in case of a negative result in the second period

Taking expectations and solving for q:

$$q = \underset{k_1=2}{Q\left(3, U\right)} = \frac{\int_0^1 \left(p + p^2 + p^3\right) dp}{\int_0^1 (1 + p + p^2) dp} = \frac{13}{22}$$

Similarly, for $n = 3$ and $k_1 = 1$:

$$3q = 1p + 0\left(1-p\right) + 1p^2 + 0p\left(1-p\right) + q\left(1-p\right) + 1p^3 + 0p^2\left(1-p\right) + \\ 1p^2\left(1-p\right) + 0p(1-p)^2 + q\left(1-p\right)$$

i.e. the only difference happens when the first negative result occurs in the second period, after one positive realized in the first trial: in this case, the individual chooses to experiment even in the third period when he follows the strategy specifying $k_1 = 1$.Taking expectations:

$$q = \underset{k_1=1}{Q\left(3, U\right)} = \frac{\int_0^1 \left(p + 2p^2\right) dp}{\int_0^1 (1 + 2p) dp} = \frac{7}{12}$$

The individual will choose the strategy in terms of k_1 in order to maximise the expected payoff from experimenting, i.e. the expression of the boundary $Q\left(3, U\right)$ will take the following form:

$$Q\left(3, U\right) = \max_{\{k_1=1; k_1=2\}} \left\{ \frac{\int_0^1 \left(p + p^2 + p^3\right) dp}{\int_0^1 (1 + p + p^2) dp}, \frac{\int_0^1 \left(p + 2p^2\right) dp}{\int_0^1 (1 + 2p) dp} \right\} = \frac{13}{22}$$

The individual will decide to experiment if and only if $q \leq Q\left(3, U\right)$ [18]. In case of a success, the player will find convenient to undertake again project 1, while in case of a negative result, he will update his beliefs on U and calculate again the dynamic allocation index attached to project 1, according to information obtained in the experimentation.

[18] By convention, the individual wil experiment when indifferent between the safe and the uncertain project, i.e. when at the boundary.

In case of a failure, the time horizon is $n = 2$ and the agent has to calculate $Q\left(2, U^f\right)$, where U^f indicates the posterior distribution updated after the realization of one negative result yielded by project 1, given the uniform as the prior distribution function. Now, the only possible case is $k_1 = 1$.

Under $k_1 = 1$ and $n = 2$, the dynamic allocation index takes the following form:

$$q = Q\left(2, U^f\right) = \frac{\int_0^1 \left(p + p^2\right)\, dU^f}{\int_0^1 (1 + p)\, dU^f} = \frac{3}{8}$$

Suppose the agent undertakes project 1 and it yields a negative result for the second time. The decision problem of the agent is simplified, since the horizon time shrinks to $n = 1$. Let us indicate the posterior distribution after two failures with $U^{ff}(p)$. The dynamic allocation index in this case is:

$$q = Q\left(1, U^{ff}\right) = 1\int_0^1 p dU^{ff} = \frac{1}{4}$$[19]

After two negative results, with $n = 1$, the individual will compare q with the updated probability of getting a positive result, given that he experienced a negative one in the first two runs of trial.

5.2 Introducing disappointment

In this section, we characterize the sequential strategy of the agent facing the same decisional problem described in the previous section, but here we assume the player experiences the psychological feelings of disappointment and elation and he takes them into account when evaluating the optimal strategy.

The agent's preferences are described by the disappointment-elation utility function, as defined in section 3.

Proposition 12 *Under $n = 3$ the decision rule is summarized as follows:*
Undertake project 1 if and only if
$q \leq \max\limits_{\{k_1, k_2\}} \left\{\frac{39}{44} - \frac{9}{44}b, \frac{21}{24} - \frac{5}{24}b\right\}$
If it yields a negative result, undertake project 1 if and only if $q \leq \frac{5}{8} - \frac{5}{24}b$
If it yields a negative result, undertake project 1 if and only if $q \leq \frac{7}{16} - \frac{3}{16}b$

In order to keep the agent indifferent between experimenting or not, it has to be verified the following equality:

[19] The general formula indicating the probability of getting a success given that the individual has experimented t times and has registered s successes, if the parameter $p = \Pr(X = 1)$ is drawn from a uniform distribution $U(0, 1)$, takes the following form

$$q \leq \frac{s+1}{t+2}$$

where s indicates the number of successes out of all trials and t the number of trials along the unknown arm.

In our simple case, $s = 0$, $t = 2$. It follows that the individual will experiment again along the unknown arm if $q \leq \dfrac{1}{4}$

$$\begin{gathered}3q = (1 + (1 - \bar{c}))\, p + -b\bar{c}(1-p) \\ + (1 + (1 - \bar{c}))\, p^2 - b\bar{c}p(1-p) + q(1-p) \\ + (1 + (1 - \bar{c}))\, p^3 - b\bar{c}p^2(1-p) + qp(1-p) + q(1-p)\end{gathered}$$

At the boundary, the agent equates the sum of the utility of the sure payoff associated with undertaking three times the safe project and the sum of the expected utility of payoffs in the case the agent decides to undertake project 1 inf the first period and then act according to the optimal strategy specifying $k_1 = 2$. The mechanism driven by the vector k plays a key role in understanding the functional form of dynamic allocation indexes, since now the agent experiences a negative level of utility in case of a negative result yielded by project 1, while he has zero utility level in case of standard expected utility theorem paradigm. According to the same procedure shown above, the dynamic allocation index has the following expression:

$$Q\left(3,U\right) = \max_{\{k_1,k_2\}} \left\{ \begin{array}{c} (1 + (1 - \bar{c}))\, \dfrac{\int_0^1 \left(p + p^2 + p^3\right) dp}{\int_0^1 \left(1 + p + p^2\right) dp} + (-b\bar{c})\, \dfrac{\int_0^1 (1-p)\, dp}{\int_0^1 \left(1 + p + p^2\right) dp}, \\ (1 + (1 - \bar{c}))\, \dfrac{\int_0^1 \left(p + 2p^2\right) dp}{\int_0^1 (1 + 2p)\, dp} + (-b\bar{c})\, \dfrac{\int_0^1 \left(1 + p - 2p^2\right) dp}{\int_0^1 (1 + 2p)\, dp} \end{array} \right\}$$

Under the distributional assumption of a uniform distribution, it reduces to: $Q^D\left(3,F\right) = \max\limits_{\{k_1,k_2\}} \left\{\frac{39}{44} - \frac{9}{44}b, \frac{21}{24} - \frac{5}{24}b\right\}$

The decision rule followed by the individual remains unchanged: the agent will decide to experiment if and only if $q \leq Q^D\left(3,F\right)$, i.e. the individual will decide to experiment if the utility associated with the payoff of the sure project is lower or equal to the expected utility associated with payoff obtained by following the optimal strategy on project 1. Assume $q \leq Q^D\left(3,F\right)$. Accordingly to the decision rule, the individual will find convenient to undertake project 1 If he experiments and gets a positive result, he will find convenient to experiment again. In the case he gets a negative result,the time horizon is $n = 2$ and the agent has to calculate $Q\left(2,U^f\right)$, where U^f indicates the posterior distribution updated after the realization of a negative result with project 1. When the time horizon is $n = 2$, the only possible case is $k_1 = 1$.

Under $k_1 = 1$ and $n = 2$, the dynamic allocation index takes the following form:

$$q = Q\left(2,U^f\right) = \left(1 + \left(1 - \bar{c}^f\right)\right) \frac{\int_0^1 \left(p + p^2\right) dU^f}{2 - \int_0^1 (1-p) dU^f} + \left(-b\bar{c}^f\right) \frac{\int_0^1 (1-p^2) dU^f}{2 - \int_0^1 (1-p) dU^f}$$

where $\bar{c}^f = \int_0^1 p dU^f = \frac{1}{3}$

It follows that if $q \leq Q^D\left(2,U^f\right)$ boundary calculated after a negative result, the individual will find convenient to undertake again project 1. In the opposite case, he will decide to switch to the safe project forever.

The optimal strategy predicts a "stay with a winner" rule, so that the agent needs to evaluate the convenience of his action only in case of a negative result..

Suppose the agent experiments along the unknown arm and a negative result occurs for the second time. The decision problem of the agent is simplified, since the horizon time shrinks to $n = 1$. and the individual will compare q with the updated expected utility of the reward of project 1, given that he experienced a negative result in the first two runs of trial. Therefore, the dynamic allocation index becomes:

$Q^D\left(1, F^{ff}\right) = \left(1 + \left(1 - \overline{c}^{ff}\right)\right) \int_0^1 pdF^{ff} + \left(-b\overline{c}^{ff}\right) \int_0^1 (1-p)\, dF^{ff}$

where $\overline{c}^{ff} = \int_0^1 pdF^{ff} = \frac{1}{4}$

5.3 Comparisons

Does the agent experiment more when he feels disappointment in case of a failure? The aim of this section is to provide an answer in the simple setting with a finite horizon examined above and we determine the conditions on the disappointment parameter under which the economic agent is willing to experiment more when characterized by a disappointment-elation utility function with respect to the standard EUT paradigm.

The agent will decide to experiment more when characterized by disappointment when the following inequality is satisfied:

$$Q^D\,(n, F) \geq Q^{EUT}\,(n, F).$$

This table exhibits the values of the parameter that satisfy the above condition under the specification of the disappointment-elation function considered in Section 3 and 4 and under different assumption on the time horizon of experimentation. In the second last column we report the value of the parameter b in the case we do not allow the agent to update the distribution function and the expected payoff $\overline{c}$ of the unsafe project after each experimentation trial, while in the last column we report the value of parameter b when updating is allowed.

$N = 2$	$Q\,(2, U) = \frac{5}{9}$	$b \leq \frac{5}{4}$	$b \leq 1, 15$
	$Q\left(1, U^f\right) = \frac{1}{3}$	$b \leq 1$	$b \leq 1$
$N = 3$	$Q\,(3, U) = \frac{13}{22}$	$b \leq \frac{13}{9}$	$b \leq 1, 24$
	$Q\left(2, U^f\right) = \frac{3}{8}$	$b \leq \frac{6}{5}$	$b \leq 1, 16$
	$Q\left(1, U^{ff}\right) = \frac{1}{4}$	$b \leq 1$	$b \leq 1$
$N = 4$	$Q\,(4, U) = \frac{8}{13}$	$b \leq \frac{8}{5}$	$b \leq 1, 28$
	$Q\left(3, U^f\right) = \frac{4}{9}$	$b \leq \frac{4}{3}$	$b \leq 1, 2$
	$Q\left(2, U^{ff}\right) = \frac{7}{25}$	$b \leq \frac{7}{6}$	$b \leq 1, 15$
	$Q\left(1, U^{fff}\right) = \frac{1}{5}$	$b \leq 1$	$b \leq 1$

From the inspection of table I, it can be seen that the parameter b increases as N increases: i.e. as the number of trials of experimentation before the end of the game increases, the maximum intensity of disappointment aversion that the agent can bear, in order to have more experimentation with respect to the standard EUT, increases. Not surprisingly, the parameter b decreases after each failure: since the agent's utility is reduced after each failure and losses have a higher weight than gains, he is willing to bear a lower intensity level of disappointment after a failure. It is worth noting that the agent can weigh consistently more losses relative to gains in psychological terms and still decide to experiment more than a standard EUT maximiser- which again is counterintuitive.

More interestingly, the table shows that the result of Proposition 10 are not vulnerable to the simplifying assumption of an expected value of the payoff of the unsafe project not updated after each trial of experimentation. Note that when we allow the agent to update both the distribution function and the expected payoff, the range of values of the parameter b satisfying the condition on more experimentation shrinks but the agent can still weigh more losses to gains and decide to experiment more with respect to a standard expected utility maximiser agent.

6 References

1. Bradt, R. N., S. M. Johnson, and S. Karlin, (1956) "On Sequential Designs for Maximizing the Sum of n Observations", *Annals of Mathematical Statistics*, Volume 27, Issue 4 (Dec., 1956), 1060-1074.

2. Berry, D and Fristedt B., (1985) "Bandit Problems: Sequential Allocation of Experiments" Chapman&Hall

3. De Groot, M. H, (1970) "*Optimal Statistical Decisions*" McGraw Hill

4. Gittins, J., (1979) "Bandit processes and dynamic allocation indices",. *Journal of the Royal Statistical Society B*, Volume 41, 148-64

5. Hey, John D., Chapter 6, "Experiments and the economics of individual decision making"

6. Kahneman, D. and A. Tversky, (1979) "Prospect Theory: an analysis of decision under risk", *Econometrica*, Volume 47, 111-132.

7. Kahneman, D and A. Tversky (1991) "Loss Aversion nad Riskless Choice: A Reference Dependent Model", *Quarterly Journal of Economics*, Volume 104, 1039-61

8. Kahneman, D. and A. Tversky, (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty*, Volume 5, 297-323

9. Loomes, G. and R. Sugden, (1982) "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty", *The Economic Journal*, Volume 92, 805-824.

10. Loomes, G. and R. Sugden, (1986) "Disappointment and Dynamic Consistency in Choice under Uncertainty", *Review of Economic Studies*, Volume 53, 271-282.

11. Loomes, G. and R. Sugden, (1987) ”Some Implications of a more General Form of Regret Theory”, *Journal of Economic Theory*, Volume 41, 270-287.

12. Roberts K. and M. Weitzman, (1981), ”Funding criteria for research, development, and exploration projects” *Econometrica*, Volume 49, 1261-1288

13. Rothschild, M. (1974), ”A Two-armed Bandit Theory of Market Pricing”, *Journal of Economic Theory*, Volume 9, 185-202.

14. Savage, L.J. (1954), *The Foundations of Statistics*, New York, Wiley.

15. Sugden, R. (1991) ”Rational Choice: A Survey of Contributions from Economics and Phylosophy”, The *Economic Journal*, Volume 101, Issue 497, 751-785.

16. Von Neumann, J. and Morgenstern, O. (1947), *Theory of Games and Economic Behaviour* (2nd edition), Princeton, Princeton University Press.

17. Weitzman M., (1979) ”Optimal search for the best alternative” *Econometrica*, Volume 47, 641-654.