The Role of Risk Attitudes in Probabilistic Environments

Santiago Alonso Díaz[*]
Carlos Esteban Forero[**]

sadiaz@bcs.rochester.edu
ce.forero1129@uniandes.edu.co

[*] Professional in Finance and Foreign Affairs, Universidad Externado de Colombia. Specialist in applied mathematics, Universidad Sergio Arboleda. MSc in Neuroscience and Education, Columbia University. PhD(c), Brain and Cognitive Sciences, University of Rochester

[**] Economist, Universidad de los Andes.

Fecha de recepción: 2 de agosto de 2012
Fecha de aceptación: 4 de julio de 2013
Forma de citar:
Alonso-Díaz, S., & Forero, C. (2013). The Role of Risk Attitudes in Probabilistic Environments. *ODEON*, 7, 63-86.

1. Introduction

The aggregate of individual decisions determines the movement of markets. From this observation one can ask an important question: under what mechanisms do these individuals behave? For example, one classical model, the efficient-market hypothesis, states that prices carry all the necessary information of assets (Fama, 1970), and as a corollary, the behavior of individuals should be some function of their direction. Furthermore, agents can form rational expectations based on the dissemination and aggregation of information that markets, with appropriate institutions, allow (Plott & Sunder, 1988). One strong assumption of this proposal is that prices carry sufficient and complete information, but in cases of herding, defined as the imitation of prior actions of others and inefficient reliance on public information (Vives, 1997; Avery & Zemsky, 1998), prices fail to integrate private information that agents have because they decide on external events, causing a deviation from equilibrium (Banerjee, 1992).

In any case, rational expectations and herding behavior leave open the question of the factors that allow individuals to infer the relevance of the signals (e.g. prices or prior actions of others). To fill this gap, in a recent paper, Brugier, Quartz, & Bossaerts (2010) found that social circuits of the brain, which allow people to make inferences about others and their states, were active during trading sessions. Importantly, circuits that support logic and numerical reasoning were passive, and not even performance on math tests (similar to the ones used to recruit in Wall Street) were predictive of the ability to infer the direction of the market. In their study, it was high performance on social tests (i.e. those that measure theory of the mind; ToM: the ability to infer intentionality) that predicted how well subjects assessed the direction of the market.

The Brugier et al. (2010) study revealed that the social tools our brain has are used in making decisions, but additional cognitive factors should be relevant, in particular our capacity to learn statistical regularities of the environment. It is known that in environments with feedback, subjects manage to learn the intrinsic probabilities of success (or failure) that options have (Estes, Campbell, Hatsopoulos, & Hurwitz, 1989). For that reason we take the ability to capture, implicitly or explicitly, the statistical properties of objects in the environment (e.g. probability of success of a given Option A vs. Option B) as central in decision-making. For example, a person, e.g. a trader, who believes, given the available information, it is more probable that share 1 gives appropriate dividends than an alternative share 2, should select share 1. This depends on a myriad of factors, but if the trader is in a fast-paced environment, with limited time to make decisions, his selection should be a function of his personal abilities to implicitly assign probabilities to each share (assuming similar levels of dividends and prices).

This paper, therefore, is focused on the ability to learn statistical properties, especially probabilities, based on the history of the object/option/share/other. Specifically, we are interested in the details of probabilistic learning, with feedback (for a review of a task used to assess probabilistic learning see Meeter, Radics, Myers, Gluck, & Hopkins, 2008) and how risk attitudes should modulate it. Our hypothesis is as follows: First, risk seeking subjects should have better probabilistic learning abilities because they explore more and collect additional evidence on the available options. Second, risk-averse subjects should have lower learning rates because they prefer to be safe and to exploit the known option. This is inefficient if a more lucrative alternative exists and is vaguely explored. These predictions will be operationalized using a multi-armed bandit game, in which exploration and exploitation behaviors can be observed (Daw, O'Doherty, Dayan & Seymour, 2006) and, importantly, it is a game that mimicks actual financial decision-making (Payzan-LeNestour & Bossaerts, 2012). Risk attitudes will be measured using a typical framing effect task that captures risk behavior in gains and in losses (De Martino, Kumaran, Seymour, & Dolan, 2006). In addition, we conducted cognitive controls using Raven's Progressive Matrices to see if the effects (if any) depend on personal cognitive endowments.

2. Risk Preferences, learning and multi-armed bandit game

In principle, there are two general sources of learning: 1) Direct instruction and 2) Trial and error. The former is hardly dependent on risk attitudes because learning comes from a central figure which passes on knowledge, based on personal experiences or previous history of instruction to the learner, who does not take risks because almost everything is transmitted to him or her. The latter type of learning, on the other hand, is the one we are interested in because it does imply risk taking. For example, when an infant is learning characteristics of his/her environment, many experiences take the form of trial and error, which depends on risk taking. Learning that stoves are hot can be a painful experience that requires exploration. In that sense, trial and error requires subjects that take risks, in the form of exploratory behaviors (i.e. look for new and potential better results), or play it safe, in the form of exploitation (i.e. continuing with the same course of action that has shown the best positive results so far).

One possibility of how learning occurs in trial and error, which we will refer to from now on as feedback environments, is reinforced learning (Kaelbling, Littman, & Moore, 1996; Sutton & Barto, 1998). The general framework of reinforcement is updating: an initial estimation is updated according to external information. For example, if someone is asked how many balls are in a bowl and gives an overestimate, he/she can be informed via feedback. In the following turns, estimates will be corrected, according to the feedback, until the exact number is told. In this simple illustration, two elements are worth noting: 1) An initial estimate that is updated and 2) feedback. Importantly, sensitivity towards the feedback determines learning. This sensitivity takes the form of a learning parameter in reinforced learning models (formal details in the methodological section).

One classic game, suitable for reinforced learning frameworks, is the multi-armed bandit game (Robbins, 1952; Vermorel & Mohri, 2005). This game is a probabilistic environment in which learning occurs through feedback. In general, subjects face n levers, e.g. 4. In the simplest case, each lever has the same class of probability distribution (e.g. normal distribution) with different parameters (e.g. different means but equal standard deviations). The task of the subject is to maximize his/her reward in a given number of trials (e.g. 150 trials). He is naïve on the statistical properties of each lever and must learn through trial and error.

In addition to being a suitable environment for reinforced learning, the multi-armed bandit game allows for explorative and exploitative behaviors. For example, if after n trials a subject believes that lever number 3 is giving more reward than others, he could decide to exploit that lever. Similarly, he could also start to explore after some trials, to learn whether other levers are indeed less rewarding. For this reason, the multi-armed bandit game is an ideal environment to see if risk preferences and learning are related. In order to be successful, a balance between exploitation and exploration must exist. We relate the former to risk-averse behavior, while the latter is closer to risk-seeking behavior. For that reason, we hypothesize that risk profiles and learning are connected, via exploitative and explorative behaviors.

It is important to clarify that risk attitudes are dissociable in gain and loss frames (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981). That is, individuals tend to be risk seeking in losses and risk averse in gains. In their classic example, Tversky & Kahneman (1981) asked two groups of people to chose between two possible actions, one safe and the other risky. The risky option was identical in both groups (e.g. “there is 1/3 of probability that all will be saved and 2/3 that all will die"). It was the safe option that was framed differently across groups. The first group saw the safe option in terms of gains (e.g. "if this option is taken, 200 out of 600 people will be saved") and the second group saw it in terms of losses (e.g. "if this option is taken, 400 out of 600 people will die"). Note that both safe options were identical in logical structure; they varied only in the way they were written. Tversky & Kahneman (1981) found that the first group played it safe, while the second preferred the risky option. This dissociation of risk attitudes between gain and loss frames is robust (meta-analysis of framing effects in Kuhberger, 1998; Piñon & Gambara, 2005), so we decided to use a task that captured that fact (see more details in the section on methodology).

3. Methodology

3.1 Participants:

Students were recruited from a high school in Bogotá (10th grade, average age 17). We selected this population, instead of the usual undergraduates that most experimental economics studies use, because they are more likely to be naïve on the tasks; particularly the framing task (which undergrad students of economics might know from one of their courses). A total of 31 male students completed all tasks. We ran only male students so as to simplify/avoid gender analysis.

3.2 Procedure:

To avoid excessive effort and attention issues, two sessions, held on different days, were run. In the first one, students did the framing task; in the second one, they completed the multi-armed bandit game and the Raven's test. Students were run in groups of approximately 16 students. The best students in the behavioral tasks, from each group, received a reward of $20,000 COP (approximately US$10 dollars). This reward scheme was intended to motivate students to behave as well as possible.

3.3. Tasks:

Subjects completed 3 tasks:

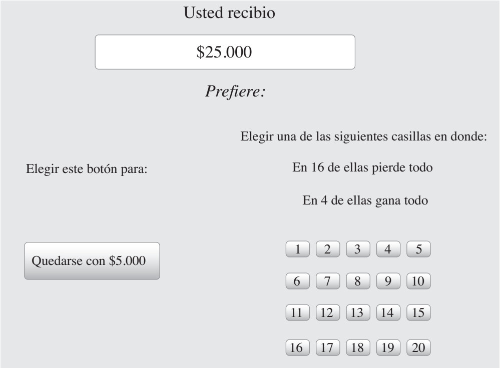
a) Framing Task (Figure 1)
b) Multi-Armed Bandit game -MAB (Figure 2)
c) Cognitive Control (Raven's Progressive Matrices)

3.3.1 Framing task:

We based this task on the De Martino et al. (2006) design. In each trial subjects were told to imagine that a monetary endowment was given to them. Two options then appeared. On the left side there was a safe option, written either as a gain or a loss in relation to the monetary endowment. On the right side, there was a risky lottery with 20 buttons. Some of them were win-all buttons, while others were lose-all buttons (Figure 1). The expected value of both the safe and the risky option was identical to elicit risk attitudes, instead of expected value computations. There were a total of 32 trials (4 endowments, 4 probabilities, 2 frames)[1] and 16 additional control trials in which either the safe or the risky option had higher expected value[2].

Risk attitudes were measured as the number of times each subject decided in favor of the risk option in gain and loss frames, separately. A total measure of risk attitudes was the sum of both. Additionally we computed a rationality index as follows: the proportion of trials where the risky option was chosen in gain frames was substracted from the same proportion in loss frames. A value of zero indicates low susceptibility to the framing manipulation i.e. being equally averse to (or seeking) risk in either frame. This difference, in absolute values, was linearly transformed so that 0 means least rational and 1 most rational (see more details in De Martino et al. 2006).

Figure 1: Framing Task



Usted recibio

$25.000

Prefiere:

Elegir una de las siguientes casillas en donde:

Elegir este botón para:

En 16 de ellas pierde todo

En 4 de ellas gana todo

[1] The 4 endowments were: $25,000, $50,000, $75,000 and $100,000 COP (exchange rate per dollar approximately $2,000 COP). The 4 probabilities were: 0.2, 0.4, 0.6, and 0.8. The two frames were: gain and loss.

[2] Control trials had weighted expected value to confirm that subjects were choosing non-randomly.

3.3.2 Multi-armed bandit game (MAB):

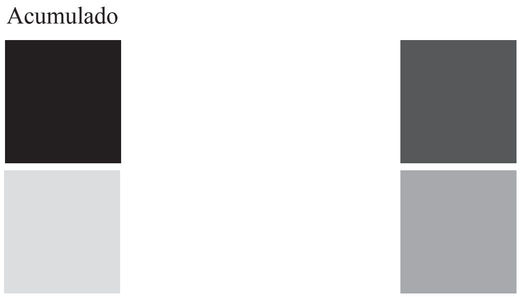
We followed the Daw et al. (2006) design, with some modifications. Students saw 4 buttons on screen. Each button was colored differently. Every time a button was pushed, a number appeared indicating how many units were won by pushing it. At the top of the screen, the student saw the total number of units won so far (Figure 2). Their task was to push the buttons freely and try to maximize the total number of units awarded. Every student completed 149 trials.

Each button gave units following normal distributions with $\sigma = 4$ but with different initial means: blue button 80, red button 60, yellow button 40 and green button 30. The maximum number units were limited to 100 and the minimum to 1. To avoid excessive exploitation, each time a button was pushed its mean diffused as follows:

$$\mu_{i,t+1} = \lambda\mu_{i,t} + (1 - \lambda)\Theta + \nu \qquad (1)$$

where λ=0.9836 is the decay parameter, Θ=50 is the decay center, ν is the diffusion noise - which is distributed normally with a mean 0 and $\sigma_d = 2.8$, i stands for each button and t is the current trial. Notice that if a subject overexploits one button, its mean will eventually go to the decay center. This design, therefore, has implicit incentives to explore. We changed the original design of Daw et al. (2006) slightly because in their study all buttons diffused every time a button (any button) was pushed. Our modification of only diffusing the pushed button implies that the button with the highest initial mean (the blue button) should attract attention faster and for a longer time, and in this sense it becomes a safe option. Because the existence of a strong button, and the fact that students were learning the statistical properties of the buttons, which have diffusing means, any exploration is a stronger indication of risk-seeking behavior than in the original design.

Figure 2: MAB (Multi-armed bandit)



Learning and exploitation parameters, for each subject, were estimated with a softmax rule, updated with a reinforced learning model as proposed in Daw et al., (2006), and computed with max-likelihood methods. The softmax rule was:

$$P_{i,t} = \frac{e^{\beta \hat{\mu}_{i,t}^{pre}}}{\sum_j e^{\beta \hat{\mu}_{j,t}^{pre}}} \qquad (2)$$

where $P_{i,t}$ is the probability of choosing button i at trial t, β is the exploration/exploitation parameter and $\hat{\mu}_{i,t}^{pre}$ is the estimated mean of button i at trial t. This formula is saying that the probability of choosing a button i in trial t is a non-linear function (i.e. exponential) of the estimated mean of that button in previous trials. High β indicates strong exploitative tendencies. In the denominator there is a standardization term, which is the sum of that non-linear function in all the different j buttons at trial t.

The updating of $\hat{\mu}_{i,t}^{pre}$ follows a reinforced learning model:

$$\hat{\mu}_{i,t}^{post} = \hat{\mu}_{i,t}^{pre} + k_t \delta_t \qquad (3)$$

where $\hat{\mu}_{i,t}^{post}$ is an updating factor of the mean of button i at trial t, $\hat{\mu}_{i,t}^{pre}$ is the estimated mean of button i at trial t, k_t is the learning rate and δ_t is the error prediction term, equal to the reward/feedback received in trial t for pushing button i minus $\hat{\mu}_{i,t}^{pre}$. Notice that k_t indicates how sensitive a subject is to deviations of his own personal estimate of rewards ($\hat{\mu}_{i,t}^{pre}$) of a particular button. In this model, the learning parameter k_t is a function of the variances of the buttons:

$$k_t = {\hat{\sigma}_{i,t}^{2\,pre}} \Big/ {(\hat{\sigma}_{i,t}^{2\,pre} + \hat{\sigma}_d^2)} \qquad (4)$$

Because buttons are diffusing, subjects update, for each button, the process described in (1). Therefore for trial t+1:

$$\hat{\mu}_{i,t+1}^{pre} = \hat{\lambda} \hat{\mu}_{i,t}^{post} + (1 - \hat{\lambda}) \hat{\theta}$$

Notice that $\hat{\mu}_{i,t+1}^{pre}$, which is the estimated mean of button i in the trial t+1, is a function of the updated mean $\hat{\mu}_{i,t}^{post}$ in (3). $\hat{\theta}$ and λ are as described in (1).

Variances are updated similarly with:

$$\hat{\sigma}_{i,t}^{2\,post} = (1 - k_t)\hat{\sigma}_{i,t}^{2\,pre} \qquad (5)$$

and

$$\hat{\sigma}_{i,t+1}^{2\,pre} = \hat{\lambda}^2\hat{\sigma}_{i,t}^{2\,post} + \hat{\sigma}_d^2 \qquad (6)$$

The max-likelihood procedure for (2) was done using Matlab® and the algorithm by Jee Hoon Yoo, of the University of Bristol[3]. With it, we computed the values for k_t, β and the number of times students decided to exploit (defined as decisions in favor of the button with the highest expected value predicted by the model). These values give a proxy of learning abilities and are the ones that are going to be related to risk attitudes.

It is important to observe that reinforced-learning modeling depends on the initial values of the mean $\hat{\mu}_{i,0}^{pre}$ and standard deviation $\hat{\sigma}_{i,0}^{2\,pre}$ because the updating procedure in (3) has to start with an initial estimate. It is not possible to know it for each subject. To solve this, the algorithm by Jee Hoon Yoo selects random starting points. This is an excellent solution, in regard to the impossibility of asking subjects their starting estimates (even if asked, they would also make a random guess), but it has a cost since parameters change every time the algorithm is run (precisely because a different starting point is used). To address this issue, we ran the model 100 times for each subject and we found that the algorithm was consistent: the parameters Kt, β and number of exploitations, between runs, were strongly correlated (all 14,850 correlations, i.e. 4950 per parameter, were significant at $p<0.05$; the average r_s between runs were: for $K_t = 0.60$, $\beta = 0.59$ and number of exploitations = 0.90; the average correlations between the median of each parameter and each run were: for $K_t = 0.74$, $\beta = 0.75$ and number of exploitations = 0.94) and Friedman's Tests (which is the non-parametric equivalent of a one-way ANOVA and checks if parameter computations between runs are different) failed significance (for Kt: $\chi^2(99)=97.42$, $p=0.52$; β: $\chi^2(99)=78.10$, $p=0.94$; number of exploitations: $\chi^2(99)=66.73$, $p=0.99$).

Therefore, even though the algorithm used different starting points, the computations were stable. Nonetheless, to acknowledge the variability of each run, and because they were not symmetrically distributed, we decided to use the median of the parameters Kt, β and number of exploitations of the 100 runs as the relevant measures of the multi-armed bandit game.

[3] The algorithm can be found at http://www.cs.bris.ac.uk/~rafal/rltoolbox/index.html

3.3.3. Cognitive control:

To see if any of the effects was due to cognitive abilities, students completed the Raven's progressive matrices test. This is a well known cognitive test and it has been used previously in economic research as cognitive control (Burks, Carpenter, Goette, & Rustichini, 2009). Basically, there are 60 trials. In each trial there is a set of images where one element is missing. Subjects have to choose out of 6 options the one that best fits the pattern of images. Each trial receives one point. In addition to absolute scores, we calculated percentiles using the norms in Cayssials, Albajari, Aldrey, Liporace, Naisberg & Scheinsohn (1993).

4. Results[4]

Table 1 shows that none of the learning parameters were significantly related (i.e. $p<0.05$) to risk attitudes. That is, risk seeking in loss, gain and in the total framing task were not strongly correlated with K_t (r(31)=0.01, p=0.92; r(31)=-0.12, p=0.50; r(31)=-0.06, p=0.71, respectively), β (r(31)=-0.10, p=0.57; r(31)=0.24, p=0.19; r(31)=0.08, p=0.66, respectively) or the number of times subjects decided to exploit ((r(31)=-0.04, p=0.79; r(31)=-0.15, p=0.40; r(31)=-0.08, p=0.66, respectively).

The correlations in Table 1 do not control for two important confounding variables: cognitive abilities and reaction times in the multi-armed bandit game. The former was related, negatively, with β (i.e. with Raven Score r(31)=-0.36, p=0.04; with Raven percentile r(31)=-0.34, p=0.06) and showed a positive trend with number of exploitations (r(31)=0.31, p=0.08), while the latter was strongly correlated, negatively, with number of exploitations (r(31)=-0.49, p<0.01).

The cognitive results indicate that subjects with high scores tend to follow an exploitative strategy. This is the best strategy given that in our design one of the buttons was clearly better (i.e. it had a high initial average; see methodological section). Nonetheless, neither Raven's score nor percentile were correlated with the number of times the highest paying button was pushed (r(31)=0.10, p=0.56; r(31)=0.10, p=0.56, respectively) or the amount of units won (r(31)=0.16, p=0.38; r(31)=0.16, p=0.37). Therefore, the only statement that follows from the data is that subjects with higher cognitive scores followed more exploitative than explorative behavior in our version of the multi-armed bandit game.

[4] All correlations reported in this paper are Spearman's correlations.

Table 1 has one supplemental finding regarding cognitive abilities worth noting. We found a positive trend between the Raven score (r(31)=0.32, p=0.07) and Raven percentile (r(31)=0.34, p=0.05) with the rationality index from the framing task (low scores indicate least rational and high scores most rational; details in De Martino, et al. 2006). Therefore, subjects with high cognitive skills were also more consistent, as measured by the index. This is important because it confirms that decision-making is not independent of cognitive processing, both in MAB and framing tasks (the framing results extend the findings in Burks, Carpenter, Goette, & Rustichini, 2009). However, at the same time it is clear that it is not about better performance: in MAB, cognitive scores were not related to any of the learning parameters or number of units won; in the framing tasks we used, being more consistent, as measured by the rationality index, is irrelevant because in each trial the expected value of the risky and safe option are equal. Our results only point out that cognitive abilities in the MAB game is related to exploitative strategies and in the framing task to the rationality index (that only measures how consistent a subject is across risk situations in gains and losses).

Table 1. Correlations (Spearman's)

	Kt	**β**	**# of Exploitations**	**RT MABł**	**Rationality Index**	**Risk in loss**	**Risk in gain**	**Risk Total**	**Raven Score**	**Raven %**
Kt	1									
β	-0.75**	1								
# of Exploitations	0.24	-0.21	1							
RT MABł	-0.27	0.25	-0.49**	1						
Rationality Index	0.22	0.05	0.07	0.04	1					
Risk in loss	0.01	-0.10	-0.04	-0.25	-0.04	1				
Risk in gain	-0.12	0.24	-0.15	0.05	0.43**	0.60**	1			
Risk Total	-0.06	0.08	-0.08	-0.11	0.24	0.83**	0.92**	1		
Raven Score	0.26	-0.36**	0.31*	-0.06	0.32*	-0.16	0.04	-0.01	1	
Raven %	0.27	-0.34*	0.24	-0.05	0.34*	-0.31*	-0.06	-0.14	0.94**	1

*p<0.01 □ Reaction Times Multi-armed Bandit

**p<0.05

As for reaction times, the negative correlation shows that faster subjects exploited. Interestingly, this means that once a subject decided to exploit, his decision was automatic; he just pushed a button repeatedly (ergo, the fast reaction times). This implies that exploitative strategies, in our design, were fast and as a corollary depend on intuitive systems, rather than reflective ones (Camerer, Loewenstein, & Prelec, 2005), or just follow a heuristic one (i.e. pushing any button fast).

To explore the relationships of risk attitudes and learning further, via model parameters in the MAB game, we controlled for these variables (cognitive abilities and reaction times in MAB) using separate linear regressions (Table 2). The only models that were highly explicative and significant were the ones that had number of exploitations as a dependent variable. In particular, in the model that only controlled for cognitive abilities, risk-seeking behavior in gains becomes a highly significant (and negative) predictor of number of exploitations. That is, after the cognitive control, risk seekers in gains tend to explore more. The results can also be interpreted alternatively: after controlling for risk attitudes, cognitive results strongly modulate number of exploitations. This last interpretation complements the finding of Table 1, in that exploitation is positively related to performance in the Raven's test, but it is stronger for risk averse in gains. Interestingly, the standardized coefficient for risk seeking in loss almost achieved a trend (i.e. p=0.105), but with positive sign. This means that, in general, risk-seeking attitudes consistent with Tversky & Kahneman (1981) proposal, of being risk seeking in losses and risk averse in gains, seems to elicit exploitative behavior (i.e. because of the opposite signs of the standardized coefficients).

The model that used RT in MAB as control was also significant and it made risk attitudes irrelevant in regard to number of exploitations (i.e. these predictors were not significant in the model), which could indicate that subjects followed a heuristic of pushing one button rapidly. But when both controls (Raven Percentile and RT in MAB) were introduced, risk attitudes in gains and percentile score were again significant (Table 2). This is notable because it confirms that behavior in the task was not only a heuristic, but was also modulated by risk attitudes (in gains) and cognitive processing.

Table 2. Standardized Coefficients/R^2
(Significance)

	Dependent Variables			
	Kt	**β**	**Number of Exploitations**	**Units Won**
	Cog. Control			
Risk Seeking in Loss	0.279 (0.211)	-0.305 (0.187)	0.301 (0.105)	0.020 (0.934)
Risk Seeking in Gains	-0.369 (0.102)	0.366 (0.116)	-0.505 (0.009)	-0.105 (0.658)
Raven Percentile	-0.398 (0.050)	-0.237 (0.247)	0.684 (0.000)	0.192 (0.363)
R^2	0.162 (0.182)	0.106 (0.380)	**0.431** **(0.001)**	0.036 (0.798)
	RT. Control			
Risk Seeking in Loss	0.064 (0.788)	-0.307 (0.191)	-0.217 (0.304)	0.013 (0.956)
Risk Seeking in Gains	-0.171 (0.453)	0.331 (0.140)	-0.071 (0.722)	-0.071 (0.756)
RT MAB	-0.108 (0.604)	-0.220 (0.281)	-0.510 (0.009)	0.160 (0.448)
R^2	0.041 (0.764)	0.100 (0.409)	**0.262** **(0.039)**	0.027 (0.861)
	RT & Cog. Control			
Risk Seeking in Loss	0.286 (0.268)	-0.492 (0.059)	0.104 (0.592)	0.159 (0.552)
Risk Seeking in Gains	-0.374 (0.129)	0.499 (0.043)	-0.364 (0.055)	-0.204 (0.421)
Raven Percentile	0.402 (0.063)	-0.334 (0.115)	0.582 (0.001)	0.265 (0.234)
RT MAB	0.012 (0.954)	-0.320 (0.128)	-0.336 (0.041)	0.239 (0.279)
R^2	0.162 (0.311)	0.183 (0.243)	**0.517** **(0.001)**	0.080 (0.692)

Bold indicates significant model.

This does not mean that having high cognitive scores or being risk-seeking in gains was better. Table 2 shows that the number of units won during the MAB task was independent of (at least not linearly related to) these predictors (i.e. none of the models in Table 2 that had “Units Won” as dependent variable achieved significance). In fact, and not surprisingly, “Units Won” was only related to the number of times the best button was pushed (i.e. the one with the highest initial mean; see methodological section) (r(31)=0.32, p=0.07), which in turn was positively related to the learning parameter K_t (r(31)=0.69, p<0.01). Because this learning parameter was always between 0 and 1 (range: 0.00007 – 1) it implies that subjects who won more units captured better the error signal from the reward/feedback received in each trial (i.e. K_t modulates the error signal δ_t; see explanations of formula (3)). In other words, their reward-updating mechanism, via reinforced learning, seemed more efficient.

5. Discussion

Our initial hypothesis stated that risk attitudes had to be connected to learning in probabilistic and feedback environments. Simple correlations (i.e. without any control) failed significance (Table 1) but when cognitive scores and reaction times in MAB were introduced as controls, risk seeking in gains emerged as an important negative modulator of exploitative strategies. In other words, subjects that tended to be risk seekers in gains explored more. Importantly, none of the risk profiles or cognitive scores was related to the number of units won in MAB (Table 2), which indicates that they influence strategies, not performance. In the following paragraphs we will try to discuss why risk attitudes (specifically in gains) and cognitive abilities are connected to explorative/exploitative strategies. We will also do some reflections on its relevance for financial decision making and conclude with some limitations of our study.

5.1 Risk attitudes, cognitive abilities and decision making

Being a risk-seeker in gain is indicative of having a stronger tendency towards risk, because the usual finding is that people are risk averse in gains (Tversky & Kahneman, 1981; Kuhberger, 1998; Druckman, 2001; Gonzalez, Dana, Koshino, & Just, 2005; De Martino, Kumaran, Seymour, & Dolan, 2006). This observation, coupled with our results, suggests that people who like risk prefer to explore. Three complementary reasons might account for this: 1) they are sensitive to (i.e. they like) new feedback; 2) they dislike leaving options behind; 3) they are less susceptible to "hot-stove" effects.

The first reason (sensitivity to new feedback) is supported by recent findings that show that prediction errors (which come in the form of feedback) elicit physiological outputs and the release of dopamine in particular. That is to say, it has been found that prediction errors (via feedback) activate dopaminergic circuits and neurons (Schultz, 1998). The fact that feedback can provoke dopamine release is interesting because it can be connected to actual behaviors. For example, Cohen et al. (2010) found that adolescents are hypersensitive to positive feedback (as measured by activity in the striatum, a brain region known to be sensitive to positive feedback), and they conclude that this increased signal might account for the risk-seeking behavior that many adolescents exhibit. An important characteristic of the dopamine signal is that once a result comes stable it diminishes in intensity, precisely because there is no prediction error in stable results. In the case of the MAB game, once a button is learned it seizes to produce error signals and dopamine release is consequently reduced, making risk-seeking subjects explore new buttons. This depends on an increased sensitivity to dopamine in risk-seeking individuals compared to risk-neutral or risk-averse ones. Frydman, Camerer, Bossaerts, & Rangel, (2010) indeed found that risk- seeking subjects tended to carry more of the allele MAOA-L. This gene codes for the enzime MAOA which regulates the catabolism of monoamines, including dopamine. Those with the allele MAOA-L produce less of this enzyme, and for this reason they decompose dopamine less efficiently (i.e. more dopamine is present in their synapse). All this suggests that feedback signals could be more potent in risk-seeking individuals, and when it ceases to appear it is more notable, making them explore new options (a hypothesis that has yet to be confirmed, but additional literature also links risk behaviors and dopamine: Kuhnen & Chiao, 2009; Dreber, et al., 2011).

The second reason (dislike of leaving options behind) is an interesting phenomenon that has been observed experimentally by Shin & Ariely (2004). They found that when subjects had to decide, repeatedly, between three doors that had different rewards distributions (similar to our button design), the threat of disappearing doors that were not sampled increased their attractiveness (i.e. they were selected more), even if they were of little interest. This is similar to another behavioral effect: foregone payoffs. When subjects are shown the outcomes of risky unselected options, these become more luring (i.e. in terms of selecting them more) (see details in Yechiam & Busemeyer, 2006). Therefore, our results can be explained because risk-seekers might be more susceptible to a dislike of leaving options behind and decide to explore, just in case.

The third reason (decreased susceptibility to hot-stove effects) refers to the hot-stove effect, in which learning increases risk aversion, especially if one (or more) of the options has positive outcomes (March, 1996) and is symmetrically distributed (Denrell, 2007). A hypothetical example can help to clarify and generate an intuition. Assume that an individual who goes to a new restaurant that has many good and bad dishes on the menu (according to his taste). After 5 or 10 visits he realizes that he really likes dish X (positive outcome). On subsequent visits he continues sampling more dishes, some of which are not at all tasty. The effect states that after learning he will eventually return, and frequently select, dish X, to the detriment of other options. His learning increased his preference for the safe dish X and made him avoid other risky options on the menu that might be good or bad. Notice that this depends on the fact that dish X is usually prepared the same way every time (symmetrically distributed). What could be happening in the MAB environment is that risk-seekers explore more because they are less susceptible to selecting the same button (dish), even after learning its positive outcomes.

The previous three possibilities are hypothetical because most studies do not control for risk attitudes of subjects, so it is not possible, for example, to know if the hot-stove effect is indeed weaker in risk-seeking populations. In general, however, they help to explain why risk-seekers (in gains at least) prefer to explore. More importantly, in our results, this depended on the cognitive control. The signs in the regressions of Table 2 indicate that low cognitive scores coupled with high risk-seeking in gains, decreased exploitation (or, similarly, increased exploration). We have already tried to explain the results of risk seeking, so now we turn to the potential reason why cognitive abilities, as measured by Raven's test, is connected to a strategic profile (i.e. exploitation/exploration).

Raven's test can be described as a pattern-recognition task: a matrix is missing one element and the subject has to pick, from a set of options, which one is consistent, pattern-wise, with the others. In this sense, our results have to be stated more precisely: pattern- recognition abilities increase exploitative behavior. This clarification is important because cognitive ability is not a unitary concept, and more importantly, it is not equivalent to intelligence, which is a complex construct (e.g. Gould, 1981; Sternberg, 1999) not derivable from one test (or many).

Some authors who work with cognitive performance and economic decisions interpret results as "facilitation" (Benjamin, Brown, & Shapiro, 2006; Burks et al., 2009). That is to say, subjects with higher cognitive abilities make better decisions because they process/perceive with greater ease/precision the relevant data and options. This is an interpretation that requires the existence of a general cognitive factor (call it *g*) that is useful for cognitive tasks as well as for economic tasks. The problem is that it requires extensive details on why and how the *g* factor uses the same tools employed in an inter-temporal task, or a risk-taking task; and suffice it to say that it has not been explained in sufficient detail (to the best of our knowledge).

Alternatively, others affirm that before adhering to a facilitation stance, which requires a clear definition of what is a better decision (i.e. deciding between normative or positive approaches), and accepts the existence of a general cognitive factor *g*, results do point to the fact that high performers in cognitive tasks make decisions *differently* (Frederick, 2005). Our results are in line with this position. We found that cognitive scores on Raven's test modulated the amount of exploitations positively. This was indeed the best strategy in our design, but only if the highest-mean button was exploited; interestingly cognitive scores were not correlated with the number of times this button was pushed (r(31)=0.10, p=0.56) or the number of units won (Table 2). Therefore, the only statement that is supported by the data is that subjects with high scores exploited more. But why should this be the case? Raven's test measures pattern recognition, which can be defined as an ability to capture regularities. We hypothesize that this notion of regularity lies behind exploitative behaviors. High performers in Raven's test tend to confirm regularities by over sampling buttons. This is speculative but it serves to explain why a cognitive score, such as Raven's, is connected to a strategic behavior.

5.2. Financial decisions

The similarity of the MAB game to financial decisions has been described by Payzan-LeNestour & Bossaerts (2012). To make their point, and among other arguments, they use over-the-counter (OTC) transactions as an example. In these transactions it is important to know what to invest in and also with whom to trade, and on many occasions this requires sampling of both trading partners and investments. Beyond OTC, many financial decisions require exploration/exploitation and learning, especially when facing assets with behavior that follows some sort of probability distribution.

Our results are relevant for the two main ways of considering behavior in finance: 1) Bayesian updating (which supports proposals such as the efficient-market hypothesis); and 2) Behavioral finance (which supports proposals such as herding behavior). In the first approach, prices are corrected efficiently, via arbitrage, because subjects update their estimates following rational rules, such as Bayes (for more on Bayes rule in finance see Pastor & Veronesi, 2009). In the second, equilibrium deviations are common because people are full of psychological limitations that make them feel overconfident, be averse to losses, and fall prey to a myriad of cognitive mishaps (more on behavioral finance in Barberis & Thaler, 2003). We now turn to brief reflections on how our findings address each proposal.

Bayesian updating depends on feedback. Just to recall, in the original Bayes formulation, the posterior probability is a function of what is called prior probability. The former can be taken as my current assessment of the probability of the behavior that I am interested in (e.g. prices), and the latter, my beliefs before any new information arrives (which is why it is called ‘prior’). Bayesian learning models affirm that when individuals receive new information they update their assessments (posterior) following a version of the original Bayes Theorem. Once the update is complete, the posterior becomes the new prior or belief for the next point of time when new information arrives (more in Griffiths, Kemp, & Tenenbaum, 2008).

For Bayes formulations to be efficient in markets, subjects also have to sample efficiently from the different options. If there is an excess of exploitation or explorations, learning is truncated, precisely because Bayesian learning requires appropriate updating. The implications should be clear, especially if there is a short time span for learning. For example, if there is overexploitation of one asset (or trading partner in OTC markets), to the detriment of others, learning will be sub-optimal. We found in our MAB design that exploration was related to risk-seeking behavior in gains. To the best of our knowledge, there is no population profile of risk preferences in market professionals, such as traders. Intuitively, it should be expected to find a bigger percentage of risk seekers in this type of population, in comparison to other professions. This could mean, for example, a tendency to over-explore in stock markets with high density of risk seekers.

In regard to behavioral finance, we consider that our results are relevant in the characterization of noise traders. This type of trader behaves against tenets of rational expectations, and can influence prices away from equilibrium. Noisy trading can occur for a variety of reasons, but biases and heuristics have been prominent in explaining it. For example, on average, people exhibit belief perseverance: once people form an opinion/belief, they cling to it (Anderson, 2007). For Barberis & Thaler (2003), this and other similar cognitive phenomena can explain financial puzzles, such as the equity premium puzzle, which is assumed to follow from noisy trading.

Our results indicate that noisy trading and deviations from equilibriums could be a function of strategic profiles. For example, assume that our MAB design is a market and that the number of times a button is pushed is a proxy for the number of transactions, and that they, in turn, determine the movement of prices. We should expect that if we divide our market into risk-seekers in gains and non-risk seekers in gains (while controlling for cognitive ability) to have an underpriced blue button (the one with the highest mean) in the first market (the one with risk seekers in gains) because they prefer to explore other options, and make less transactions for the blue one. This type of interpretation is not dependent on the usual cognitive illusions found in behavioral finance, but on strategic profiles. In other words, it is more a matter of how people prefer to behave (e.g. strategies used) than their biases (e.g. belief perseverance). In fact, the notion that strategic profiles characterize noisy traders goes in hand with heuristics programs that defend the idea that decision making is more about rules of behavior than information processing (Gigerenzer & Brighton, 2009).

5.3 Limitations

Our experiment has three limitations that must be addressed in future investigations:

1. Results cannot be extended beyond our participants, who were all male and young. Future research has to explore female populations, has well as other age groups
2. We used an arbitrary number of runs (100) to compute the median of the parameters. We consider 100 runs to be sufficient, but that does not imply that the random methodology, used in the algorithm to find starting points, is perfect. In fact, starting points is a methodological challenge for reinforced-learning modeling that deserves thorough attention.
3. We used a very high mean for one of the buttons and we did not diffuse all the buttons every time one was pushed. The reason was given in the methodological section, but this makes our learning environment very specific and further research should thus use different experimental designs to check if our results hold.

6. Conclusion

Risk seeking, while controlling for pattern-recognition abilities, does influence behavior in the MAB game via strategic tendencies (i.e. exploiting/exploring). This influence is not connected to performance, because none of the risk profiles was related to the number of units won in the multi-armed bandit game. Interestingly, neither did the cognitive control affect this variable. Therefore, even though we did find elements (risk attitudes in gains and pattern-recognition abilities) that explained decision-making in a probabilistic environment such as MAB, we failed to find a clear mark on what the high performers were doing correctly.

Implications for the field of financial decision-making are clear: strategic profiles in markets are important, and could affect prices. It is not only about information processing, but also about strategies, which are a function of both cognitive abilities and risk profiles.

7. References

Anderson, Craig A. (2007). Belief perseverance. In Baumeister, Roy F. & Vohs, Kathleen D., *Encyclopedia of Social Psychology*, pp. 109-110, Thousand Oaks: Sage.

Avery, Christopher & Zemsky, Peter. (1998). Multidimensional Uncertainty and Herd Behavior in Financial Markets. *The American Economic Review,* vol. 88, n.º 4, pp. 724-748.

Banerjee, Abhijit. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics,* vol. 107, n.º 3, pp. 797-817.

Barberis, Nicholas & Thaler, Richard. (2003). A survey of behavioral finance. In Constantinides, George M., Harris, Milton & Stulz, René M., *Handbook of the Economics of Finance: Volume 1B, Financial Markets and Asset Pricing*, pp. 1053–1128, North Holland: Elsevier.

Benjamín, Daniel J.; Brown, Sebastián A. & Shapiro, Jesse M. (2006). Who is "Behavioral"? Cognitive Ability and Anomalous Preferences. *Social Science Research Network*: http://dx.doi.org/10.2139/ssrn.675264, Retrieved July 25, 2012.

Bruguier, Antoine J.; Quartz, Steven R. & Bossaerts, Peter L. (2010). Exploring the nature of trader intuition. *The Journal of Finance,* vol. 65, n.º 5, pp. 1703-1723.

Burks, Stephen V.; Carpenter, Jeffrey P.; Goette, Lorenz & Rustichini, Aldo. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *PNAS* early edition: www.pnas.org_cgi_doi_10.1073_pnas.0812360106, Retrieved November 14, 2010.

Camerer, Colin; Loewenstein, George & Prelec, Drazen. (2005). Neuroeconomics: How Neuroscience can Inform Economics. *Journal of Economic Literature,* vol. 43, 9-64.

Cayssials, Alicia; Albajari, Veronica; Aldrey, Adriana; Liporace, Mercedes F.; Naisberg, Carina & Scheinsohn, Maria J. (1993). Normas de la ciudad de Buenos Aires Argentina. In Raven, John C., *Test de Matrices Progresivas,* pp. 11-17, Argentina: Paidos.

Cohen, Jessica R.; Asarnow, Robert F.; Sabb, Fred W.; Bilder, Robert M.; Bookheimer, Susan Y.; Knowlton, Barbara J.; Poldrack, Russell A. (2010). A unique adolescent response to reward prediction errors. *Nature Neuroscience*, doi:10.1038/nn.2558 (advanced online publication), Retrieved March 08, 2011.

Daw, Nathaniel D.; O'Doherty, John P.; Dayan, Peter & Seymour, Ben &. Dolan Raymond J. (2006). Cortical substrates for exploratory decisions in humans. *Nature,* Vol. 441, pp. 876-879.

De Martino, Benedetto; Kumaran, Dharshan; Seymour, Ben & Dolan, Raymond J. (2006). Frames, biases and rational decision making in the human brain. *Science,* Vol. 313, pp. 684-687

Denrell, Jerker. (2007). "Adaptive Learning and Risk Taking. *Psychological Review ,* Vol. 114, n.*º* 1, pp. 177-187.

Dreber, Anna; Rand, David G.; Wernerfelt, Nils.; Garcia, Justin R.; Vilar, Miguel G.; Lum, Koji & Zeckhauser, Richard. (2011). Dopamine and risk choices in different domains: Findings among serious tournament bridge players. *Journal of Risk and Uncertainty*, vol. 43, n.º 1, pp. 19-38.

Druckman, James N. (2001). Evaluating framing effects. *Journal of Economic Psychology*, vol. 22, pp. 91-101.

Estes, William; Campbell, Jane A.; Hatsopoulos, Nicholas & Hurwitz, Joshua B. (1989). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology*, vol. 15*,* n.º 4, pp. 556-571.

Fama, Eugene F. (1970). Efficient Capital Markets: A review of theory and empirical work. *The Journal of Finance,* vol. 25*,* n.º 2, pp. 383-417.

Frederick, Shane. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, vol. 19, n.º 4, pp. 25-42.

Frydman, Cary; Camerer, Colin; Bossaerts, Peter & Rangel, Antonio. (2010). MAOA-L carriers are better at making optimal financial decisions under risk. *Proceedings of The Royal Society*, doi:10.1098/rspb.2010.2304.

Gigerenzer, Gerd & Brighton, Henry. (2009). Homo Heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science,* vol. 1, pp. 107-143.

González, Cleotilde; Dana, Jason; Koshino, Hideya & Just, Marcel. (2005). The framing effect and risky decisions: Examining cognitive functions with fMRI. *Journal of Economic Psychology*, vol. 26, pp. 1-20.

Gould, Stephen J. (1981). *The Mismeasure of Man.* New York: W. W. Norton & Company.

Griffiths, Thomas L.; Kemp, Charles & Tenenbaum, Joshua B. (2008). Bayesian models of cognition. In Sun, Ron., *Cambridge Handbook of Computational Cognitive Modeling,* pp. 59-101, New York: Cambridge University Press.

Kaelbling, Leslie P.; Littman, Michael L. & Moore, Andrew W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, vol. 4, pp. 237-285.

Kahneman, Daniel & Tversky, Amos. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, vol. 47*,* n.º 2, pp. 263-292.

Kuhberger, Anton (1998). The Influence of Framing on Risky Decisions: A Meta-analysis. *Organizational Behavior and Human Decision Processes*, vol. 75*,* n.º 1, pp. 23-55.

Kuhnen, Camelia M. & Chiao, Joan Y. (2009). Genetic Determinants of Financial Risk Taking. *PLoS One*, vol. 4*,* n.º 2, pp. 1-4.

March, James G. (1996). Learning to be risk averse. *Psychological Review*, vol. 103, n.º 2, pp. 309-319.

Meeter, Martijn; Radics, G.; Myers, Catherine; Gluck, Mark & Hopkins, Ramona O. (2008). Probabilistic categorization: How do normal and amnesic patients do it? *Neuroscience and Biobehavioral Reviews*, vol. 32, pp. 237-248.

Pastor, Lubos & Veronesi, Pietro. (2009). Learning in financial markets. National Bureau of Economic Research: http://www.nber.org/papers/w14646.pdf, Retrieved July 27, 2012

Payzan-LeNestour, Elise & Bossaerts, Peter. (2012). Learning to Choose the Right Investment in an Unstable World. Social Science Research Network: http://dx.doi.org/10.2139/ssrn.2056927, Retrieved July 18, 2012

Piñon, Adelson & Gambara, Hilda. (2005). A meta-analytic review of framing effect: Risky, attribute and goal framing. *Psicothema*, vol. 17, n.º 2, pp. 325-331.

Plott, Charles R. & Sunder, Shyam. (1988). Rational expectatations and the aggregation of diverse information in laboratory security markets. *Econometrica*, vol. 56, n.º 5, pp. 1085-1118.

Robbins, Herbert (1952). Some aspects of the sequential design of experiments. *Bulletin of the AMS*, vol. 58, pp. 527-535.

Schultz, Wolfram (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, vol. 80, pp. 1-27.

Shin, Jiwoong & Ariely, Dan (2004). Keeping Doors Open: The Effect of Unavailability on Incentives to Keep Options Viable. *Management Science,* vol. 50, n.º 5, pp. 575-586.

Sternberg, Robert J. (1999). The theory of successful intelligence. *Review of General Psychology*, vol. 3, n.º 4, pp. 292-316.

Sutton, Richard S. & Barto, Andrew G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Tversky, Amos & Kahneman, Daniel. (1981). The framing of decisions and the psychology of choice. *Science*, vol. 211, pp. 453-458.

Vermorel, Joannes & Mohri, Mehryar. (2005). Multi-Armed Bandit Algorithms and Empirical Evaluation. *Machine Learning: ECML 2005 (Conference Proceedings),* pp. 437-448, Porto: Springer.

Vives, Xavier. (1997). Learning from others: A welfare analysis. *Games and Economic Behavior*, vol. 20, pp. 177-200.

Yechiam, Eldad & Busemeyer, Jerome R. (2006). The effect of forgone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making*, vol. 19, pp. 1-16.