UNIVERSIDAD CENTRAL

DOCUMENTOS DE INVESTIGACIÓN

ADMINISTRACIÓN DE EMPRESAS

Administración de Empresas

Probabilistic Categorization with Meaningful Information Using Traders and Non-Traders

Santiago Alonso Díaz

N.° 5

Abril de 2011

UNIVERSIDAD CENTRAL
FACULTAD DE CIENCIAS ADMINISTRATIVAS, ECONÓMICAS Y CONTABLES
Departamento de Administración de Empresas

DOCUMENTOS DE INVESTIGACIÓN

Administración de Empresas

Probabilistic Categorization with Meaningful Information Using Traders and Non-Traders

Santiago Alonso Díaz

N.º 5
Abril de 2011

Una publicación del Departamento de Administración de Empresas

Diego Otero Prada
Decano Facultad de Ciencias Administrativas, Económicas y Contables

Claudia Ramírez Méndez
Directora Departamento de Administración de Empresas

Documentos de investigación. Administración de Empresas, N.° 5.
Probabilistic Categorization with Meaningful Information Using Traders and Non-Traders
ISBN para PDF: 978-958-26-0266-6
Autor: Santiago Alonso Díaz
Primera edición: abril de 2011
Ediciones Universidad Central
Carrera 5 N.º 21-38. Bogotá, D. C., Colombia
Tel.: 334 49 97; 323 98 68, exts.: 2353 y 2356
editorial@ucentral.edu.co

Catalogación en la Publicación Universidad Central
Alonso Díaz, Santiago
Probabilistic categorization with meaningful information using traders and non-traders / Santiago Alonso Díaz ; editora Edna Rocío Rivera Penagos. -- Bogotá : Ediciones Universidad Central, 2011. -- (Documentos de investigación. Administración de Empresas ; no. 5)
35 p. ; 28 cm.

ISBN para PDF: 978-958-26-0266-6

Comerciantes – Estudio de caso 2. Inversiones – Toma de decisiones – Estudio de caso 3. Probabilidades – Estudio de caso 4. Planeación empresarial – Estudio de caso I. Rivera Penagos, Edna Rocío, ed.
II. Universidad Central. Facultad de Ciencias Administrativas, Económicas y Contables. Departamento de Administración de Empresas

658.15 –dc22 PTBUC/RVP

Producción Editorial

Departamento de Comunicación y Publicaciones
Dirección: Edna Rocío Rivera P.
Coordinación editorial: Héctor Sanabria R.
Diseño y diagramación: Jairo Iván Orozco A.
Diseño de carátula: Mauricio Ladino

Editado en Colombia - Published in Colombia

Content

Probabilistic categorization with meaningful information using traders and non-traders

Santiago Alonso Díaz*

Resumen

Este estudio evaluó si las características superficiales de una prueba probabilística afectaban el desempeño. En la prueba se utilizaron cartas que tenían significado en relación con la historia usada como contexto; esto es, los participantes tenían que decidir si el precio de una acción subía o bajaba a partir de una combinación de las cartas con significado. La población de estudio estuvo compuesta por dos grupos. Unos eran traders, y su trabajo se relacionaba con la historia de contexto. Los otros eran sujetos sin ninguna experiencia en trading. Los datos fueron analizados usando la metodología propuesta por Meeter et al. (2006). Los resultados mostraron que la prueba fue más difícil tanto en comparación con otros estudios como con los datos previamente recogidos en 2009. De hecho, parece que el aprendizaje probabilístico es sensible a efectos de encuadramiento (framing-effects), en los que las características superficiales son importantes para aprender o decidir apropiadamente.

Palabras clave: traders, aprendizaje probabilístico, categorización probabilística, economía conductual, aprendizaje de múltiples entradas.

Código JEL: D03, D80 y D83.

* Financista de la Universidad Externado de Colombia. MSc en Neurociencia y Educación de la Universidad de Columbia. Catedrático de Neuroeconomía en la Universidad de los Andes, la Universidad Central y la Universidad Nacional.

Probabilistic categorization with meaningful information using traders and non-traders

Santiago Alonso Díaz

Abstract

This study evaluated if facial features of a probabilistic task affected performance. In particular, cards/cues had meanings, related to the cover story. Participants had to predict if the price of a share went up or down, given a combination of the meaningful cards. Two types of participants were tested. One group was related to the cover story: they were traders. The other was composed by subjects without any trading experience. The data was analyzed using strategy analysis by Meeter et al. (2006). The results showed that the task was harder, both in comparison to other studies and previously collected data in 2009. In fact, it seems that the paradigm is prone to framing-effects, where the superficial features are important in learning or deciding properly.

Keywords: traders, probabilistic learning, probabilistic categorization, behavioral finance, multiple-cue learning

JEL Code: D03, D80 y D83.

1 Introduction

Many decisions take place in real time and, to complicate things further, the connection between an action and an outcome is not evident; in fact it could be probabilistic. Such type of fast-paced-probabilistic-outcome decisions often occur in the financial world. Traders in particular make a living in deciding how and where to allocate resources. A big part of their work is to try to predict, with the best available external information, if a market is going up or down, if it is a good time to sell or buy a bond, if currency X is going to increase its price, if country Y's debt is a good buy, if investing in commodity Z is wise; and in occasions they have to do it fast.

An initial plausible supposition is that they must be integrating relevant information before taking a decision. That is, they see some information, they relate it to an outcome and then they take a course of action. This fairly simple model is actually the description of a known paradigm in cognitive psychology, used to evaluate probabilistic learning. In general it has taken three forms: Disease prediction (e.g. Gluck & Bower, 1988), Weather prediction (Knowlton, Squire, & Gluck, 1994) and Ice Cream Flavor prediction (Hopkins, Myers, Shohamy, Grossman, & Gluck, 2004). In any of these versions, subjects have to predict a binary outcome from external information. For example, in the weather prediction version, the probabilistic outcome can take two forms: rain or sun. The subject bases his prediction on external information, represented in a combination of one, two or three cards (taken from a set of four, for a total of 14 possible combinations) he sees in a screen. The combinations of cards carry information on the probability of the weather (e.g. if card 1 and 3 appear together they predict sun with probability X and rain with probability 1-X). The subject is able to assign a combination to an outcome (i.e. learn) because it always receives feedback on the correctness of its prediction after each trial.

It is important to notice that the task has two structural components that the experimenter can tinker with. The first one is the probability structure. That is, each combination has an assigned probability (Table 1) and each card has an assigned likelihood (Figure 1). The second structural component is the "facial" features. That is, the task uses cues and cover stories, which experimenters change to apply the paradigm for particular situations or just because they prefer to use one set of figures connected to a particular story.

One issue that arises is whether the task is sensitive to changes in the structural components. In general, changes in the probability structure do affect the difficulty of the task because the maximum rate of optimal answers diminishes or increases when the conditional probabilities in Figure 1 change. For example, Hopkins et al. (2004) decided to use conditional probabilities of 0.8, 0.6, 0.4 and 0.2 for their cues because they wanted to reduce frustration on the amnesic patients they evaluated. Therefore it is known that the task is sensitive to changes in the probability structure because it changes the amount of optimal answers that can be attained.

What is not clear is the sensitivity of the task to the other structural component. The facial features are stable across recent studies on probabilistic categorization and use either geometric shapes or facial features of Mr. Potato Head (some studies have used other shapes like sailboats, bulbs, butterflies, e.g. Shohamy, Myers, Hopkins, Sage, & Gluck, 2008, but since 1994 must have used geometric shapes or Mr. Potato Head features). More important, and independently of which shapes studies use, these have usually (if not always) been disconnected from the cover story. With the exception of the disease prediction version, in which subject had to use symptoms to predict diseases, cover stories are always disconnected from the cues used to predict the event (but even in the disease prediction version this is debatable; subjects were not medical doctors in these studies and the connection of the cues/symptoms to the cover story was not evident to the subjects; but that is another issue). The reason to conduct the paradigm in this manner is not clear and it leaves an important part of the task, the facial features, without a clear role/effect.

This research evaluated professional traders and non-traders on an adapted version of a probabilistic learning task. The adaptation consisted in using cues with meanings related to the cover story (Figure 1). The reason to evaluate traders is that their work is connected to the cover story used in the present study: predict the increase or decrease of the price of a stock. This was important because other professions (e.g. architects) might detach the meaning from the cues and the objective was to see how learning is affected when cards have a meaning. This process of detaching meaning was a general assumption and to check if the effects found were indeed related to trading experience, a second group was run. They were labeled as the comparison group. Importantly, they did not have any trading experience but they were working or studying in business, finance or economics.

There were two reasons to choose this type of subjects as comparison. First, the comparison group could not to be too foreign to the cover story and the meaning of the cues. The whole objective was to see how meaningful cues affect learning in probabilistic tasks. Choosing psychology students, or architecture students, or art students, would be a potential confounding variable. With participants that have studied or worked in business related areas there could still be some issues, but in general it is expected that they have taken a course or have heard about the meaning of, for example, a new investor in a company. Second, they could not be traders because if learning was indeed affected it would be premature to state that trading was a possible explanation. Other type of subjects was required.

The present research can also be seen as a study of behavioral finance in the sense that evaluates learning behaviors of traders. It dwells, superficially, into the following question: how do traders evaluate external information to learn probabilities? Some studies have found that professional trading influences behavioral biases, heuristics and rational behavior (Anderson & Sunder, 1995; Haigh & List, 2005a; List & Haigh, 2005b; Alevy, Haigh & List, 2006; List & Haigh, 2010). Even though this study does not address the same concepts as those studies, and in fact a whole different task is used, they do support the possibility that trading influences behavior. To do so, strategy analysis was used, in particular the one developed by Meeter, Myers, Shohamy, Hopkins & Gluck (2006). With it, it is possible to see what approach traders use, when they change it, and how many times.

Methods

2

2.1 Participants

26 professional traders were recruited from different trading companies (11 in total). The average age was 36.5. The youngest participant was 24 years old and the oldest 48 years old. The average experience was 9 years, ranging from 3 to 19 years. 14 of the 26 participants had post-graduate studies, but none beyond master or MBA level. 12 participants only had undergraduate studies.

22 subjects were also run and labeled as the comparison group. The criterion for selecting them was that they should not have any experience with professional trading. The average age was 26. The youngest participant was 19 years old and the oldest was 40 years old. Most of them (16) were undergraduate students from accounting or economics. 3 were university professors in the business faculty of a university in Colombia but their expertise was not related to finance. 2 were independent entrepreneurs and the final one was working as the general administrator of a family business.

30 additional subjects were analyzed. The data from these subjects was collected at a different time (i.e. 2009), in the Language and Cognition Lab, at Teachers College, Columbia University, for an eye-tracking experiment. All were graduate students in the faculty of education. The reason to analyze these additional subjects can be found in the results section. In general, they were used as a comparison point. They did a similar probabilistic task as traders and the comparison group, but with cards meaningless to the event.

2.2 Stimulus

An adapted version of the weather prediction task in Knowlton et al. (1994) was used as stimulus. The probability structure was maintained but with different cards and cover story. At the beginning of the task, the participant saw on-screen instructions telling him to predict the increase or decrease of the price of a company's share. He based his decision on a combination of cards that appeared on-screen. The cards used are shown in Figure 1. In any given trial, one, two or three cards could appear on the screen (for a total of 14 possible combinations). The combinations that appeared signaled the probability of the event (Table 1). The subject reported his prediction (i.e. the price goes up or down) by pressing 1 or 2 on a computer keyboard. During the first trials the subject would be guessing blindly, but as trials advance he learns to relate particular combinations with the occurrence of either of the probabilistic events. This is so because he receives feedback, at the end of every trail, on the correctness of his prediction. Every participant endured 200 trials divided in 4 blocks of 50.

The reason to use the more strict version of Knowlton et al. (1994) (i.e. in other studies the probabilities of each card are a bit more discernible: 0.8, 0.6, 0.4, 0.2; compare them with the ones in Figure 1) was to compare the results with unpublished data collected in 2009, which used that version.

The meaning of the cards was shown to the participants before beginning the task (Figure 1). In previous papers, the cues/cards were neutral (e.g. geometric shapes or facial features of Mr. Potato Head). The rationale to assign meanings to the cards was to simulate a more "realistic" situation so that it closely relates to the cover story.

Figure 1

Cards used in the experiment

Subjects saw these cards on a computer screen, in different combinations (Table 1) to predict the change in price (up or down) of a company's share. Card 1: New investor; Card 2: New regulation/law; Card 3: New competition; Card 4: New CEO. The numbers in the base are the conditional probabilities of the price going UP if the card appears on screen. Card 1 is a strong card for the price going up, while card 4 is a strong card for the price going down. Images were taken from Microsoft Office Clip Art.

Strategy analysis: One question of interest in probabilistic learning literature is: how do subjects solve the task? One of the first efforts to give an answer was Gluck et al. (2002). They identified 3 basic strategies that subjects could use. The first one was called *One-Cue*, and its name is almost self-explanatory. Subjects identify one cue/card of the four and every time they see it, with other cards or by itself, they always answer with one, and only one, of the binary events (e.g. price-up or price-down). They do not care about the other cards. The second one was called *Singleton*. In this strategy, the participant learns optimally the outcome associated with all four combinations that only have

Table 1					
Probabilistic structure of the task					
Cards					
Combination	1 (Investor)	2 (Regulation)	3 (Competition)	4 (CEO)	P (up | combination)
1	0	0	0	1	0.15
2	0	0	1	0	0.38
3	0	0	1	1	0.1
4	0	1	0	0	0.62
5	0	1	0	1	0.18
6	0	1	1	0	0.5
7	0	1	1	1	0.21
8	1	0	0	0	0.85
9	1	0	0	1	0.5
10	1	0	1	0	0.82
11	1	0	1	1	0.43
12	1	1	0	0	0.9
13	1	1	0	1	0.57
14	1	1	1	0	0.79

one card (e.g. combinations 1, 2, 4 and 8 in Table 1). All the other combinations are answered randomly. The third strategy was called *Multi-cue*, and it is the optimal strategy. The subject manages to learn the optimal outcome associated with all combinations and answers accordingly.

A couple of years later, in an effort to expand strategy analysis, Meeter et al. (2006) developed a more refined way to identify strategies. In general, the authors gave more detail on the original strategies proposed by Gluck et al. (2002) and developed a method to recognize which ones participants used; they even managed to pinpoint aproximate trials were a strategy shift took place.

Because Meeter et al. (2006) proposal is the latest version of strategy analysis, it was the one used in this paper. A brief summary (as brief as it can be; in fact, the reader is invited to check the original paper for more detail) of the method will follow.

The analysis is based on a pre-established set of strategies. To be more precise, Meeter et al. (2006) established 11 different strategies that were classified in 5 groups (Table 2). As it can be seen in Table 3, each strategy has a vector of probabilities, with 14 elements, one for each combination. The probabilities are actually *likelihoods of answering that the price will go up* given a particular combination. For example, let us assume that a participant is following the strategy "Singleton" in Table 2. The table assigns a likelihood of 0.95 that the subject will answer "price-up" to combination 4 (C4). This assignment has to do with the fact that in "Singleton" strategies the subject manages to learn the outcome associated with each of the singletons (i.e. C1, C2, C4, C8); and because C4 is a singleton, and is more closely associated with the outcome "price-up" (see Table 1), the likelihood of answering that the price goes up is high. The likelihood assigned to C4 is not 100% because of the probabilistic nature of the task, and sometimes the subject could answer otherwise (e.g. just to check). Also notice that in the "Singleton" strategy all the other combinations are assigned a likelihood of 50% because he answers randomly on all non-singleton combinations (for more detail on the description of strategies see Appendix 1)

With the values of Table 2 and the answers given by each of the subjects, it is possible to estimate the likelihood that the subject is following a particular strategy given his set of answers. This is done by computing the likelihood, with the binomial distribution, of answering X times "price-up" and Y times "price-down" to a combination, given a particular strategy. Once the likelihoods for all the combinations are computed, the next step is to multiply them with each other. The result is the likelihood of following that strategy. This is repeated for all strategies and the strategy with the highest likelihood is the one assumed to be used by the subject. This strategy is called *the best fit strategy*.

Table 2

Strategies														
Description of combination	D	d	dD	u	uD	ud	udD	U	UD	Ud	UdD	Uu	UuD	Uud
Combination number	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
Singleton														
Strong Singleton	0.05	0.5	0.5	0.5	0.5	0.5	0.5	0.95	0.5	0.5	0.5	0.5	0.5	0.5
Singleton	0.05	0.05	0.5	0.95	0.5	0.5	0.5	0.95	0.5	0.5	0.5	0.5	0.5	0.5
Intermediate														
Singleton + Prototypes	0.05	0.05	0.05	0.95	0.5	0.5	0.5	0.95	0.5	0.5	0.5	0.95	0.5	0.5
2 versus 1	0.05	0.05	0.05	0.95	0.5	0.5	0.05	0.95	0.5	0.5	0.05	0.95	0.95	0.95
Optimal														
All but two strong cards	0.05	0.05	0.05	0.95	0.05	0.5	0.05	0.95	0.5	0.95	0.5	0.95	0.5	0.95
Perfect strategy	0.05	0.05	0.05	0.95	0.05	0.5	0.05	0.95	0.5	0.95	0.5	0.95	0.95	0.95
Single Cue														
Strong Up	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Weak Up	0.05	0.05	0.05	0.95	0.95	0.95	0.95	0.05	0.05	0.05	0.05	0.95	0.95	0.95
Weak Down	0.95	0.05	0.05	0.95	0.95	0.05	0.05	0.95	0.95	0.05	0.05	0.95	0.95	0.05
Strong Down	0.05	0.95	0.05	0.95	0.05	0.95	0.05	0.95	0.05	0.95	0.05	0.95	0.05	0.95
Random														
Random	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

The numbers in the cells are the likelihoods of answering that the price will go up given the combination.

The numbers are obtained using table 1 from Meeter et al. (2006) and setting the parameter π as 0.95.

D: Strong card for price down (card 4 in figure 1); d: Weak card for price down (card 3 in figure 1).

U: Strong card for price up (card 1 in figure 1); u: Weak card for price up (card 2 in figure 1).

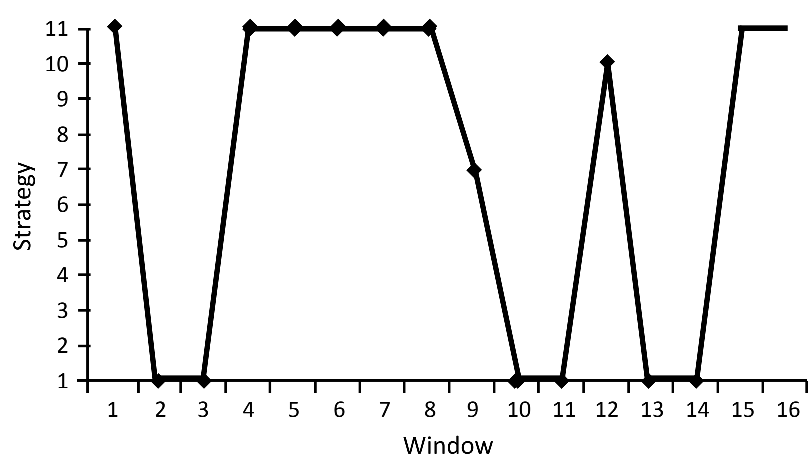
C1 - C14: Combination number. The same as in Table 1.

The previous process is done repeatedly in windows of 24 trials (i.e. using the answers of 24 trials). Each window has an overlap of 50% (i.e. it shares 12 trials with the previous window). Because of the importance of the last block, a final window with the last 40 trials is also used. The point of using windows of 24 trials is that subjects change strategies (i.e. because most subjects have higher than chance performance, this implies that they use, at least, a random strategy at the beginning and shift to an optimal strategy later) and the size of the window is a way to identify how many times a subject shifted to different ones.

This requires further detail. Each window is best fitted by different strategies because a different set of answers is used to compute the likelihoods. But often adjacent windows have the same "best fit strategy" because they share 50% of answers. Whenever two or more adjacent windows have the same best fit strategy, it is assumed that the subject is consistently using that strategy. This is called a consistent block. When a consistent block is broken, the method assumes that a strategy shift took place (Figure 2).

Figure 2

Visualization of the number of strategy shifts for one of the traders



The numbers on the strategy axis follow the same order as in Table 3. 1 = Strong Singleton; 2 = Singleton; 3 = Singleton+Prototypes; 4 = 2 vs 1; 5 = All but two strong cards; 6 = Perfect strategy; 7 = Strong Up; 8 = Weak Up; 9 = Weak Down; 10 = Strong Down; 11 = Random. This trader shifted 6 times his strategy. The number of shifts is the number of intervals where the graph is flat, plus an additional one for the change from random strategy, which is the one assumed being used at the beginning. Each window was composed of 24 trials with an overlap of 50%.

Once windows where a strategy shift took place are identified, it is also possible to estimate the trial of occurrence. The first thing to do is to compute, for the 12 trials before an analyzed trial, the likelihood of the previous best fit strategy (in the same manner it was described before). Then, the procedure is repeated but this time the likelihood is computed for the new best fit strategy and using the answers from the 12 trials after the analyzed trial. Then a subtraction is made between both likelihoods (i.e. the one for the 12 trials after minus the one for the 12 trials before). This is done for all the overlapping trials of the windows where a strategy shift took place. The trial where the subtraction is maximum is the shifting trial (for further details and special cases see Meeter et al., 2006). When two or more trials end up having the same subtraction and a maximum, the shifting trial is set at the earlier one.

In sum, the technique identifies strategies, the number of strategies used in the 200 trials and the approximate trial where a shift occurred, including the trial where the first shift, away from random strategy, happened. Meeter et al. (2006) validated the method with Monte Carlo simulations. Answers were simulated using one of the strategies in Table 2. The method was then applied to see if it would identify the strategy used to generate the answers. The results were above average, and fluctuated from 70% as high as 92% of correct identifications. Meeter et al. (2006) applied the method to previous data collected by Gluck et al. (2002) and Hopkins et al. (2004). Their results will be commented and compared in the following chapter.

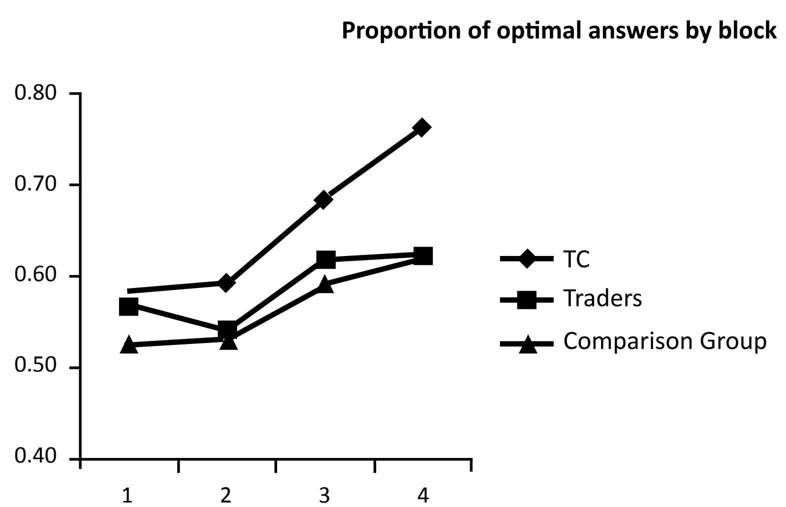
3

Experiment 1 (Traders)

3.1 Performance

The average performance of traders, in the 200 trials, was 58.79% with a standard deviation of 10.33. As the task advanced, a linear trend [$F(1,25) = 7.148$, $p = 0.013$] showed that performance increased (Figure 3). In the final block, they had, on average, 62% of optimal answers (Note: when performance is mentioned, it is always about optimal answers: those which reflect the dominant probability of the combination. Optimal answers are different from a correct prediction. In fact, an optimal answer is feasible even if an incorrect prediction was made in any given trial, i.e. because the outcomes are probabilistic).

Figure 3



There were 4 blocks in total, each one of 50 trials. Traders' and comparison's data was collected for this paper, using the cards from figure 1. The TC line is unpublished data collected from graduate students in the Language and Cognition Lab, at Teachers College, Columbia University. The stimulus used on TC participants was exactly the same as the one used in Knowlton et al. (1994), i.e. the weather prediction version.

Even though average performance on the 200 trials was above chance, traders seem to have worst performance than in previous papers using the original weather prediction task in Knowlton et al. (1994). An informal inspection of Table 3 shows that the performance of traders seems slightly below controls/students in these studies. Unfortunately, only one of these (i.e. Gluck, Shohamy, & Myers, 2002) reported, explicitly, the average and standard deviation (with these information is possible to calculate t-values and check for any statistical difference), but no significant differences were found with traders [$t(54) = 1.49$, $p = 0.14$]. Nonetheless, subjects from the study that reported standard deviations (i.e. subjects from Gluck et al., 2002) seemed to have the lowest performance in comparison with the other papers in Table 3. So, it is feasible that traders would underperform in comparison with subjects tested in the other studies.

To further test the possibility that traders underperformed in regards to previous studies, unpublished data, collected by the author for an eye tracking experiment in 2009, in the Language and Cognition Lab at Teachers College, Columbia University, was used (Figure 3). These data (from now on referred as TC data or subjects) is relevant because participants ($n = 32$) did the classical version of the task, with the same probability structure and meaningless geometrical shapes used in Knowlton et al. (1994). Average performance for these subjects, on the 200 trials, was 66% ($SD = 5.97$). Interestingly, on the last block, were it is assumed that learning already took place, traders clearly underperformed [$t(56) = 3.79$, $p<0.001$] in regards to TC participants (Figure 3). This is an indication that they had troubles and that the task seemed harder. Remember that both sets of data

Table 3

Reported performance on previous studies that used the version on Knowlton et al. (1994)

Paper	Control/ Students	Amnesic	Parkinson	Huntington
Knowlton et al. (1994)* 350 trials	Task1(10): Avg~67% Task2(15): Avg~67%	Task1(8): Avg~60% Task2(8): Avg~59%		
Knowlton et al. (1996a) 150 trials	Exp(15): Avg 65.43% SD: N.A.	Exp(12): Avg 59.63% SD: N.A.	Exp (20): Avg~59.53% SD: N.A.	
Knowlton et al. (1996b) 150 trials	Exp(12): Avg ~67% SD: N.A.			Exp(13): Avg~55% SD: N.A.
Gluck et al. (2002) 200 trials	Exp1(30): Avg 62.41% SD: 7.36			

The numbers in the parenthesis, after Exp, are the number of subjects.

* Only Task 1 and 2 of experiment 1 are reported. Task 3 only had 50 trials in Exp 1, and 90 trials in Exp 2.

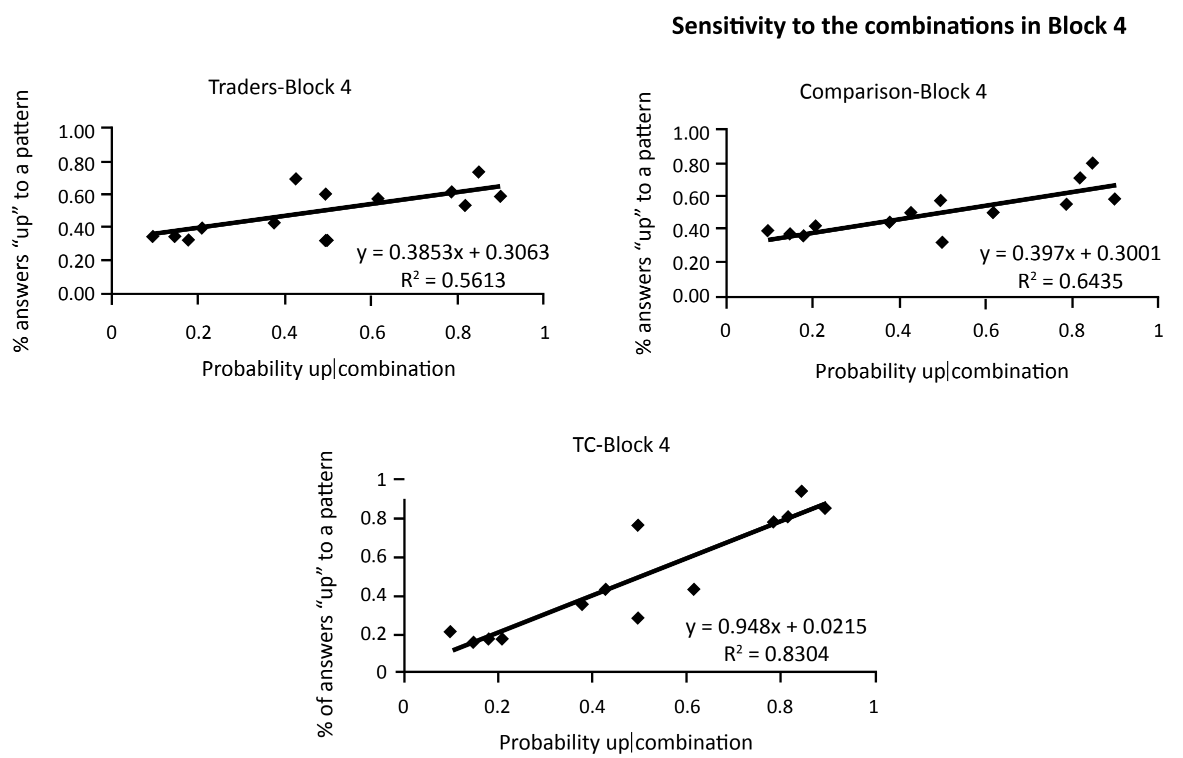
~ Personal aproximation of the average. The paper reported the performance in a special format, mostly graphs.

was collected using the same probability structure but with different facial features, so their difficulty may be explainable by the latter.

Another aspect from the performance of traders was their sensitivity to the strength of the combination (Figure 4). With strength it is meant the actual order of the combinations if they were to be put from the one least related to "price-up" to the one most related. The following would be the order (C stands for combination): C3, C1, C5, C7, C2, C11, C6, C9, C13, C4, C14, C10, C8, C12. The first term is C3 as it has a conditional probability of "price-up" of ~0.1, the lowest; while C12 is the last term because its conditional probability is ~0.9, the highest (Table 1). With sensitivity it is meant, the increase in percentage of answers "price-up" as the strength "up" increases (i.e. the slope of Figure 4).

Block 4 is particularly revealing of the final sensitivity, for both traders and TC subjects. First, the slope for TC subjects was almost one with intercept close to 0 (Figure 4). This means almost perfect sensitivity to the combination by the end of the task. As for traders, they had a less inclined slope and an intercept much more distant from zero. In other words, combinations with less strength were answered "up" more often than it should be, and combinations that should be answered in that way were not.

Figure 4



Each point is a combination paired with the average proportion of answers "price-up" in the 200 trials; they are ordered from left to right by its conditional probability of up given the combination (i.e. the first point is C3 the last C12). All groups showed linear sensitivity but TC subjects had better fit, and almost identity trend (i.e. slope close to 1 and intercept 0).

3.2 Results from strategy analysis

Strategy analysis revealed that traders shifted strategies roughly the same number of times as TC subjects. The former used, on average, 3.5 strategies (SD = 1.55), while the latter used 3.91 (SD = 1.28). The difference was not significant [t(56) = 1.092, p = 0.280]. Additionally, the number of switches is almost the same as the one found by Meeter et al. (2006) in the students [M = 3.8, SEM = 0.29] and controls [M = 3.3, SEM = 0.44] of Gluck et al. (2002) and Hopkins et al. (2004) papers.

As for the trial where the first strategy switch occurred, two things were relevant. First, many traders shifted late (i.e. not in the first block) or did not shift at all (Figure 5). In fact, 42% shifted late or never shifted. This proportion is much lower in TC subjects: 22%. Interestingly, all TC subjects shifted at least once from random strategy, but there were 4 traders, which is 15% of the whole sample, who never shifted. It is important to remember that if the method finds that in a particular window the random strategy/behavior is the best fit one, it does not mean that the subject was not trying to use different approaches. It just means that the data fits best with random behavior. Nonetheless, the traders in Figure 5 that never switched from random strategy, also had close to random behavior (M = 48.88%, SD = 3.47), which indicates that strategy analysis is reasonably accurate. This late shift by many traders (42%) could explain the low performance because identifying a correct strategy early, different from a random one, improves performance.

Second, Meeter et al. (2006) found that, on average, controls/students from Gluck et al. (2002) and Hopkins et al. (2004) studies switched strategy for the first time around the 14th-15th trial. These numbers are significantly lower than the averages obtained from traders [M = 32.93, SEM = 8.7]. In particular, against Gluck et al. (2002) participants, traders [t(50) = 2.31, p = 0.02] shifted later. The same happened with Hopkins et al. (2004) controls: traders were slower [t(50) = 2.30, p = 0.02]. In fact, both groups are comparable (i.e. no significant differences were found) to the patients with amnesia in the Hopkins et al. (2004) study. Any conclusion would be premature because the data Meeter et al. (2006) analyzed was collected with a different probability structure. Instead of the 0.75, 0.55, 0.44 and 0.25 probabilities of Figure 1, their cards/cues used a less strict structure: 0.8, 0.6, 0.4 and 0.2. What is interesting is that changing the likelihoods does not alter the probabilistic and incremental nature of the task (Hopkins et al., 2004) but it seems to alter the time when the first strategy switch occurs. More on this issue would be addressed in the discussion section.

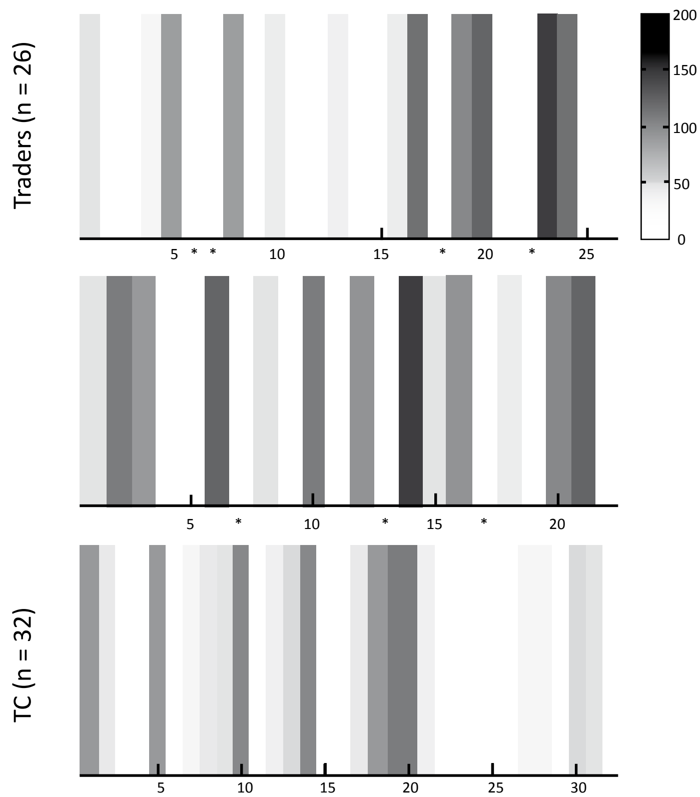
One final aspect that strategy analysis revealed was the predominant strategy used in the last 40 trials. These last trials are important because it is assumed that learning already took place. In this sense, a big proportion of traders seemed lost as they were still best fitted by random behavior (Figure 6). To be precise, in the last 40 trials more than 50% behaved as they were guessing. This is a clear contrast with Meeter et al. (2006) data, in which most of Gluck et al. (2002) subjects were using simple, intermediate or optimal strategies. Just a small proportion (less than 15%) was still using a random approach. Moreover, it also contrasts with TC data (Figure 6). These participants were using, predominantly, simple strategies (i.e. singleton or single-cue). This seems to confirm that many traders had trouble with this version of the task. Again, this will be commented later in the discussion section.

3.3 Traders by experience level

To further explore the performance of traders, two groups were formed based on experience. The criterion used was the average

Figure 5

Trial of first switch from random strategy, for traders, comparison and TC participants.



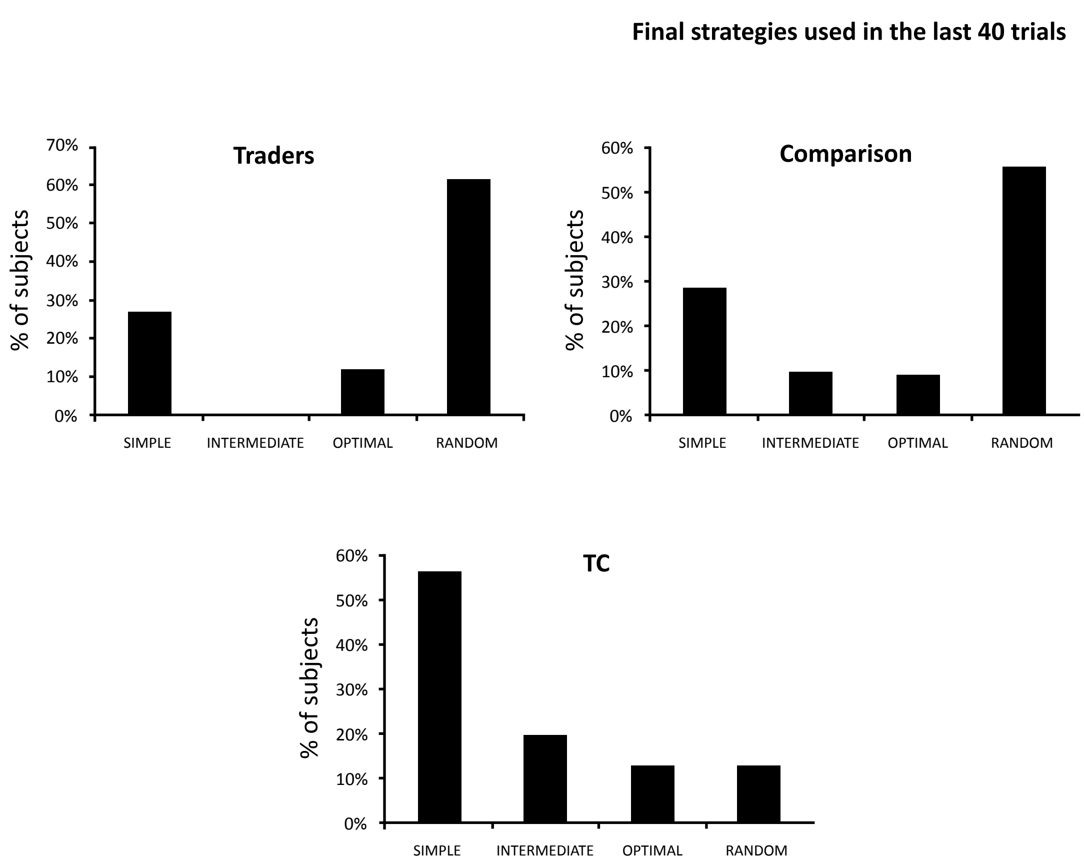
* Subjects that never change from random strategy

Brighter shades in the map represent the first block of trials. Subjects that shifted late, i.e. after the first 50 trials, are coded in darker shades. The symbol * represents subjects that never shifted strategy (in the map are coded as 0); according to strategy analysis they maintained random behavior. This did not happen with TC subjects.

experience of the whole group (M = 9.04, SD = 4.75). In other words, the first group had more than 9 years of experience and the second group had less than 9 years of experience. The former was named "Experienced Group" (n = 13, M = 12.92, SD = 3.27) and the latter was named "Amateur Group" (n = 13, M = 5.16, SD = 1.9). The name of the amateur group is symbolic, as some of the subjects (3 in total) in this group had 8 years of experience.

Their data was similar on all regards. No significant differences on performance, on the number of switches, on the trial number where the first switch occurred, were found (Table 4). Even the final strategies used were similar to those depicted in figure 6: in both experience groups, most subjects still behaved randomly towards the end of the task. The lack of difference in all the relevant measures of performance and strategy analysis shows, once again, that the task was harder in general, relative to previous studies and the TC experiment.

Figure 6



The percentage of traders and comparison subjects using a random strategy/behavior at the end of the task was higher than TC subjects. This last group used in the last 40 trials, predominantly, simple strategies (i.e. singleton or single cue).

4

Experiment 2 (Comparison Group)

4.1 Performance

Figure 3 shows that the comparison group improved its performance as the task advanced [$F(1,21) = 18.28$, $p<0.001$]; had almost the same performance as traders [$t(46) = 0.808$, $p = 0.423$] and similar learning disadvantages in the last block (i.e. in regards to TC participants [$t(52) = 5.27$, $p<0.0001$]). In fact, the sensitivity (i.e. slope) in Figure 4 is roughly the same for traders (0.38) and comparison participants (0.39).

4.2 Results from strategy analysis

Three things are relevant from strategy analysis. First, the comparison group had the highest proportion of subjects that switched late from random strategy: 54% of the sample (Figure 5). Second, the first strategy switch took place late ($M = 54.43$, $SD = 47.42$), and showed significant differences only with TC subjects [$t(49) = 2.44$, $p = 0.018$]. That is, the comparison group was indeed slower in switching from random strategies than TC subjects, but it was almost as fast as traders (no significant differences were found with traders $t(39) = 1.06$, $p = 0.29$ but a caveat has to be made. The proportion standard deviation over mean, for the trial where the first strategy shift took place, was 1.17 for traders, 0.98 for TC subjects and 0.87 for the comparison group. The high variability of the data meant that significant differences between traders and the comparison group were not found, but the mean was for the former 38.90 and the latter 54.42. The lack of differences is just a reflection of the high variability). And third, just as traders, the comparison group used,

Table 4

Averages and p-values for experienced and amateur subjects

	Performance on the 200 trials	Number of switches	Trial of first switch
Experienced	58%	3.38	29.54
Amateur	60%	3.61	36.31
p-value	0.29	0.71	0.70

predominantly, random strategies in the last 40 trials (Figure 6).

It is interesting that in most measures of performance and strategy analysis, traders and the comparison group underperformed in regards to TC subjects. Experiment 2 also shows that the difficulty of the task is not related to the fact of being a trader or not, because both groups had similar results. The reasons behind the difficulty for both groups are not clear, and they will be discussed in the following section, in particular: why would the facial features, which was the structural component changed, affect so much learning?

5

Discussion

The external features of the task, i.e. probability structure and cards with meaning, have an important influence. It has been previously established that relative small changes in the conditional probabilities of the cards, by making them more discernable, make it easier for subjects (Poldrack et al., 2001, Gluck et al., 2002, Hopkins et al., 2004, Lagnado, Newell, Kahan & Shanks, 2006, Vadhan, Meyers, Rubin, Shohamy, Foltin, & Gluck, 2008). Optimal performances can get as high as 83%. The original probabilities of Knowlton et al. (1994) only allow optimal performances as high as 75%. That is why most studies since 2002 have been using the more discernable probabilities, to make it less frustrating. The reason why this study decided to use the original probabilities was a practical one: to compare the results with previous collected data. The surprise was to find that, in addition to more stringent probabilities, the task became harder when the subjects were explicitly told the meaning of the cards (Figure 1).

It is possible that meanings put higher demands on cognitive process, not present when the cards/cues are meaningless, relative to the cover story of the task. There is evidence that two brain systems are recruited and interact when solving the task. One involved in declarative memory, the medial temporal lobe (MTL) and the other involved in procedural memory, the basal ganglia (Poldrack et al., 2001, Foerde, Knowlton, & Poldrack, 2006). Furthermore, for the assimilation of feedback, it has been shown that dopamine systems, important in error detection (Schultz, 2002), are also involved in the task (Aaron, Shohamy, Clark, Myers, Gluck, & Poldrack, 2004). Therefore, the task requires both an interaction of memory systems and an error detection process for integrating feedback. It is difficult to say which process is interrupted by the meanings of the cards, in fact any one of them could be affected. Declarative memory could be loaded with more demands because traders and comparison participants may be using their own previous knowledge on the impact of, for example, a CEO (card 4) on the destiny of companies. The error detection system could be comparing predictions both on the task itself and previous expectations (from previous knowledge) on the impact of, for example, competition (card 3) on the price of shares. All this is hypothetical and would require further studies but Foerde et al. (2006) showed that secondary tasks, done at the same time with a probabilistic learning task, do affect performance. That is, distraction is a factor and it is feasible that the meanings in Figure 1 were some kind of internal cognitive disturbance.

The sensitivity of the task to external features could be compared to other concept in economic literature: the framing-effect. Basically, framing-effects refer to the impact, on decision or appraisal of a situation, of the way the context is told, i.e. framed (Tversky & Kahneman, 1981). In other words, the external features of the problem matter. Interestingly, from the results of this paper, it seems that the concept of framing-effects does not only apply to the classical statical problems proposed by Tversky and Kahneman (where subjects are faced with situations in which they can not learn; they just decide from two static, instantaneous, probabilistic options). The data from traders and the comparison group can be seen as evidence that framing is also present in dynamic probabilistic categorization. In fact, the same categorization situation, framed differently, can have different and less efficient, learning trajectories.

Why probabilistic categorization also is sensitive to framing-effects? To try to answer, a different question has to be posed. Why would a subject, with normal health, could not be able to learn and have a close to chance performance, in this paradigm? The first obvious answer would be because he did not care, he was just answering randomly. And it is not so unlikely, taking into account the characteristcs of the subjects used in this study. Traders have busy schedules, and all were kind enough to take some time off to do the task, but they have many other different preoccupations and stresses. But at the same time is not a satisfactory explanation. In fact, it would be strange that 26 traders and 22 comparison subjects were stressed and uncaring in the two months that the data was gathered. Something else was happening and it could be related to pattern seeking behaviors. One of the classical examples of pattern seeking behaviors is the hot hand fallacy (and its cousin the gambler's fallacy): in random events, such as coin tossing, if someone is on a streak, a special ability is assigned to that person, even though the event is random. This is relevant because randomness could repeat strikes of events, in particular if the generated string is a binary one (e.g. see Spencer-Brown, 1957). Subjects doing a probabilistic categorization task might fell prey to pattern finding in random strings (the order of correct answers in the stimulus was random and unrelated from trial to trial), and in turn affect their learning. For example, a subject could be connecting two succesive trials by their content. That is, if he saw a combination X and then a combination Y, and the feedback was positive for his answer on pattern Y, he could conclude, wrongly, that every time combination Y is preceeded by combination X he has to answer, for example, "price-up". Some sort of narrative is generated that connect, otherwise independent, trials. This is a suboptimal strategy because the task is structured so that the probabilistic information is given by the cards and combinations, not by the way the trials unfold.

So, it is possible that the meaning of the cards encouraged this type of strategies. The problem is that formal evidence for this possibility was not collected.

Only informal talks with traders, after the task was completed, in which they reported that they thought the trials were connected, inspired this possibility (and texts by Taleb, 2007 and Mlodinow, 2008). If indeed this is what happened with traders and the comparison group, then framing-effects occur in probabilistic categorization tasks because random strings of binary events are happening, and it is almost instinctive to form a false narrative (at least at the beginning, when the participant is still lost and confused), more so when cards/cues have meanings (for more on narrative fallacy see Taleb, 2007). In other words, the way the task is framed alters learning because it increases the odds of inventing false patterns to connect trials. With neutral cards/cues, false narratives are awkward and most subjects would not engage in this strategy (or at least not keep it for a long time).

Another possibility is that the facial structure of the task alters how feedback, the evidence useful to learn, is integrated. Shadlen and colleagues (Shadlen, Hanks, Churchland, Kiani, & Yang, 2007) propose that decisions are made when sufficient evidence is collected, so that a threshold is surpassed. In fact, in a sequential version of the weather prediction task, Yang & Shadlen (2007) found that parietal neurons in primates were indeed behaving as if they were accumulating evidence (probabilistic evidence) as the cues appeared on screen. Most versions, if not all, of probabilistic categorization tasks in humans are not sequential (i.e. cues appear at the same time for a given combination), but is plausible that some sort of evidence is being collected, incrementally, in order to relate particular combinations with outcome one or outcome two (for evidence of incremental learning in the weather prediction task, see Lagnado et al., 2006).

The main evidence (if not the only one) in probabilistic tasks comes in the form of feedback at the end of each trial. So, and in broad terms, every time a subject answers optimally and receives a positive feedback, or answers suboptimally and receives negative feedback, a piece of evidence is collected that informs the participant about the probability assigned, by the experimenter, to the combination. The idea, if one accepts Shadlen et al. model, is that as trials go by a subject needs to accumulate enough evidence to learn the optimal answer for a given combination. If there is not enough evidence (i.e. an evidencial threshold has not been surpassed) he will continue to answer randomly. When cards/cues have meanings, this collected evidence must compete with evidence already assigned to the card, by means of education or prior biases. For example, a trader that in his professional career has experienced that regulations (card 2) hit share prices negatively, would have evidence that competes with the evidence he is collecting, in the task, via feedback (i.e. in the task card 2 is related stronger with price-up than price-down; see figure 1). This competition would make learning difficult and hinder performance.

6 Conclusion

Probabilistic learning seems to be affected by the "facial" structure of the task. It is interesting that something similar to framing-effects appears in dynamical learning. This implies that real-time information is not evaluated aseptically. In particular, this paper showed evidence that traders have learning difficulties when external input, that has a meaning, is used to predict.

Additionally, it was proposed that behind the learning difficulties of traders lies a pattern seeking behavior, not appropriate because the task is random (i.e. trials are not connected). This is quite paradoxical: Efforts to understand damage understanding. People working in turbulent business, such as trading, might find useful to take distance from his hypothesis/patterns and check for randomness first (even though this is also hard to determine).

References

7

Aaron, A., Shohamy, D., Clark, J., Myers, C., Gluck, M., & Poldrack, R. (2004). Human midbrain sensitivityto cognitive feedback and uncertainty during classification learning. *Journal of Neurophysiology*, 1144-1152.

Alevy, J., Haigh, M., & List, J. (2006). Information cascades: Evidence from an experiment with financial market professionals. *NBER Working Paper*.

Anderson, M., & Sunder, S. (1995). Professional traders as intuitive bayesians. *Organizational Behavior and Human Decision Processes*, 185-202.

Foerde, K., Knowlton, B., & Poldrack, R. (2006). Modulation of competing memory systems by distraction. *PNAS*, 11778-11783.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology*, 227-247.

Gluck, M. A., Shohamy, D., & Myers, S. (2002). How do people solve the weather prediction task? Individual variability in strategies for probabilistic category learning. *Learning & Memory*, 408-418.

Haigh, M., & List, J. (2005). Do professional traders exhibit myopic loss aversion? An experimental analysis. *The Journal of Finance*, 523-534.

Hopkins, R. O., Myers, C. E., Shohamy, D., Grossman, S., & Gluck, M. A. (2004). Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia*, 524-535.

Knowlton, B., Mangels, J., & Squire, L. (1996). A neostriatal habit learning system in humans. *Science*, 1399-1402.

Knowlton, B., Squire, L., & Gluck, M. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, 106-120.

Lagnado, D., Newell, B., Kahan, S., & Shanks, D. (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General*, 162-183.

List, J., & Haigh, M. (2005). A simple test of expected utility theory using professional traders. *PNAS*, 945-948.

List, J., & Haigh, M. (2010). Investment under uncertainty: Testing the options model with professional traders. *NBER Working Paper*.

Meeter, M., Myers, C., Shohamy, D., Hopkins, R., & Gluck, M. (2006). Strategies in probabilistic categorization: Results from a new way of analyzing performance. *Learning & Memory*, 230-239.

Meeter, M., Radics, G., Myers, C., Gluck, M., & Hopkins, R. (2008). Probabilistic categorization: How do normal and amnesic patients do it? *Neuroscience and Biobehavioral Reviews*, 237-248.

Mlodinow, L. (2008). *The drunkard's walk: How randomness rules our lives.* New York: Vintage Books.

Poldrack, R., Clark, J., Pare-Blagoev, E., Shohamy, D., Moyano Creso, J., Meyers, C., et al. (2001). Interactive memory systems in the human brain. *Nature*, 546-550.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, 241-263.

Shadlen, M., Hanks, T., Churchland, A., Kiani, R., & Yang, T. (2007). The speed and accuracy of a simple perceptual decision: A mathematical premier. In K. Doya, S. Ishii, A. Pouget, & R. Rao, *Bayesian brain: Probabilistic approaches to neural coding* (pp. 209-237). Cambridge, Massachusetts: MIT Press.

Shohamy, D., Myers, C., Hopkins, R., Sage, J., & Gluck, M. (2008). Distinct hippocampal and basal ganglia contributions to probabilistic learning and reversal. *Journal of Cognitive Neuroscience*, 1820-1832.

Spencer-Brown, G. (1957). *Probability and scientific inference.* London.

Taleb, N. (2007). *El cisne negro.* Barcelona: Paidós.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 453-458.

Vadhan, N., Meyers, C., Rubin, E., Shohamy, D., Foltin, R., & Gluck, M. (2008). Stimulus–response learning in long-term cocaine users: Acquired equivalence and probabilistic category learning. *Drug and Alcohol Dependence*, 155-162.

Yang, T., & Shadlen, M. (2007). Probabilistic reasoning by neurons. *Nature*.

Appendix

8

Strategies considered in the analysis done by Meeter et al. (2006)	
Strategy	**Description**
Random	All combinations have the same probability of being answered price-up or price-down
Singleton	Combinations that have one card (i.e. singleton) are optimally learned, the others are answered as in the random strategy
Strong Singleton	The only singleton combinations that guide the answer are the singletons with the highest probability of price-up and price-down (i.e. in this study, combinations 1 and 8). The other combinations are answered as in the random strategy
Singleton-Prototype	Combinations that ONLY have cards associated to price-up are answered as price-up, and combinations that ONLY have cards associated to price-down are answered as price-down. If a combination has cards both associated to price-up and price-down, they are answered as in the random strategy
2 vs. 1	Singleton-Prototype strategy + Majority rule in combinations with 3 cards. That is, if a combination has 3 cards, the subject answers price-up if 2 of the 3 cues are associated to price-up, and it answers price-down if 2 of the 3 cues are associated to price-down
All but two strong cards	All combinations are optimally learned with the exception of the ones that have two strong cards (i.e. in this study, combinations with cards 1 and 4). Those combinations are answered as in the random strategy
Perfect Strategy	The subject learns perfectly which combinations have a higher probability of price-up and price-down
Single card - Strong Up	Combinations with the strong price-up card are answered as price-up
Single card - Weak Up	Combinations with the weak price-up card are answered as price-up
Single card - Weak Down	Combinations with the weak price-down card are answered as price-down
Single card - Strong Down	Combinations with the strong price-down card are answered as price-down

Notes:

The reason to use the more strict version of Knowlton et al. (1994) (i.e. in other studies the probabilities of each card are a bit more discernible: 0.8, 0.6, 0.4, 0.2; compare them with the ones in figure 1) was to compare the results with unpublished data collected in 2009, which used that version.

DOCUMENTOS
DE INVESTIGACIÓN

ADMINISTRACIÓN
DE EMPRESAS

UNIVERSIDAD
CENTRAL
FACULTAD DE CIENCIAS ADMINISTRATIVAS,
ECONÓMICAS Y CONTABLES
Departamento de Administración de Empresas

Carrera 5ª Nº 21-38
PBX: 323 9868, ext.: 3751
Bogotá, D. C., Colombia

www.ucentral.edu.co