Discussion Papers in Economic Behaviour

DPEB
05/13

Statistical formats to optimize evidence-based decision making: a behavioral approach

> Iván Arribas
University of Valencia, ERI-CES, Ivie, Spain
> Irene Comeig
University of Valencia, LINEEX, Spain
> Amparo Urbano
University of Valencia, ERI-CES, Spain
> José Vila
University of Valencia, ERI-CES, Spain

October, 2013

Statistical formats to optimize evidence-based decision making: a behavioral approach

Iván Arribas, University of Valencia, ERI-CES, Ivie, Spain

Irene Comeig, University of Valencia, LINEEX, Spain

Amparo Urbano, University of Valencia, ERI-CES, Spain

José Vila, University of Valencia, ERI-CES, Spain

October 2013

Please send correspondence to: Iván Arribas, Department of Economic Analysis, University of Valencia, Avenida Tarongers, s/n, Valencia, 46022, Spain, Ph: +34 961 625 351, Fax: + 34 963 828 249 (email: ivan.arribas@uv.es).

Acknowledgments: The authors thank Enrique Fatás (University of Norwich, UK) and Ismael Rodríguez (University of Valencia, Spain) for their insights, and valuable and constructive criticisms. Thanks also goes to anonymous referees for their insightful suggestions as well as for bringing the authors' attention to the work of Robin Hogarth. Iván Arribas, José Vila and Amparo Urbano wish to thank the Ministry of Science and Technology, the European Feder Funds under project ECO-2010-20584, and the Generalitat Valenciana under the Excellence Programs Prometeo 2009/068 and ISIC 2012/021 for their financial support. Irene Comeig also wishes to thank the Generalitat Valenciana GV/2012/047 and the University of Valencia UV-INV-AE112-66116 for their financial backing.

Abstract

Statistical information is crucial for managerial decision making. The decision-making literature in psychology and mathematical cognition documents how different statistical formats can facilitate certain types of decisions. The present analysis is the first of its kind to assess the impact of statistical formats in the presentation of data from market research on both the optimality of market decisions and the time required to perform the decision-making process. An economic experiment provides the data for this study. The experiment presents statistical information in *simple frequencies* and *relative frequencies* using *numerical* and *pictorial* representations in the context of different informational environments. The key findings are that statistical information presented in terms of relative frequency formats gives rise to more accurate decision making than data presented in terms of simple frequencies, independently of the informational environments. When time is the relevant variable, numerical formats lead to a faster interpretation than pictorial ones. Since the number of factors defining the four statistical formats and the different informational environments is quite large, an orthogonal design offers a suitable experimental design. This design keeps the experiment manageable without substantially reducing its analytical power.

Keywords: Economic experiments, Statistical formats, Probability judgment, Orthogonal design, Judgment under uncertainty

1. Introduction

Statistical information is crucial for decision making. Several studies discuss how people have difficulty interpreting probabilistic information (e.g., Gigerenzer and Edwards 2003; Osman and Shanks 2005). While difficulties with probabilities sometimes seem to result from a lack of experience or education, social scientists argue that the problem lies in the presentation of information (Gigerenzer and Edwards 2003). Specifically, previous research in psychology shows that different statistical formats can facilitate some types of decisions. Regarding Bayesian reasoning, Gigerenzer and Hoffrage (1995, 1999), and Hoffrage, Lindsey, Hertwig, and Gigerenzer (2000) reveal that presenting the statistical information as simple frequencies (e.g.,1 out of 10 visits to an e-commerce website leads to an actual purchase), rather than as single-event probabilities (e.g., the probability that a visit to an e-commerce website leads to an actual purchase is 0.1), can improve people's decisions. These authors argue that the facilitation of natural frequencies comes from: i) reducing the computations needed to solve the problem; and ii) fitting better with the ways in which humans have experienced statistical information over history (see also Brase 2007). Ghosh and Ghosh (2005) review several applied studies that indicate how medical students and practitioners have difficulty explaining, interpreting or calculating probability statistics specific to medicine. Changing the format of the probability information has a significant impact on the subsequent use of this information.

Research in mathematical cognition about how people process numerical quantities also supports the facilitation of presentations in terms of frequencies, and therefore of whole numbers, rather than in terms of rates (i.e., x/y, as fractions or proportions) (see Butterworth 2007; Ni and Zhou 2005).

Another related question is the role of statistical-processing biases and their role in decision making. For instance, Soyer and Hogarth (2010) conduct a survey to 257 academics designed to test the ability of economists to make probabilistic predictions from regression outcomes presented in a manner similar to those published in leading economic journals. Many respondents underestimate uncertainty by failing to take into account the standard deviation on the estimated residuals. The addition of graphs fails to substantially improve inferences.

The important questions about how good or accurate people are at making judgments and what these judgments depend on provoke considerable controversy in the psychological literature. The psychological research on judgments under uncertainty typically asks subjects to solve a statistical problem and compares decisions between two choices with different wordings. However, as Girotto and Gonzalez (2007) point out, standard word problems are not the best instruments to test general hypotheses about the nature of human judgment (Macchi 2000; Macchi and Bagassi 2007; Macchi and Mosconi 1998).

Brase (2002) simply and directly evaluates whether—apart from simplifying statistical inference tasks—simple frequencies defined on small reference groups are clearer than other formats. The author reveals that, after eliminating possible confusion for subjects in the calculations, subjects perceive more clearly simple frequencies (based on small reference groups, e.g., 1 out of 5) and to some extent relative frequencies (percentages, e.g., 40%) than absolute frequencies (e.g., 90 million Americans) and single-event probabilities (e.g., 0.33).

Motivated by the work of Brase (2002), this study goes a step further and asks: which statistical data performs better in decision making—10 out of 100 (simple frequencies), or 10% (relative frequencies)—and under which representation—numerical or pictorial? To answer this question an economic experiment provides data to compare simple and relative frequency formats, each of them under numerical and pictorial representation. No existing experiment attempts to investigate the communication of statistical market information in business reporting. Therefore, this is the first paper to present an economic experiment on the way that statistical representation of frequencies facilitates decision making.

Three features of the experiment are worth of highlighting. First, the experiment is an economic experiment, which means that experimental subjects receive a payment, depending on their performance, to establish suitable behavioral incentives. Second, the present analysis assesses the impact of different statistical formats in the presentation of data from market research (framed in several informational environments) on the accuracy of market decisions. Thus, the analysis deals with a problem of managerial decision making, where the experimental subjects are students enrolled in several academic degrees related to business administration. The methodology is that of typical decision-making tasks, namely choices between alternatives. Both simple frequencies and relative frequencies appear as numerical and pictorial representations. These statistical formats appear in the context of several informational environments, namely sample size, number/percentage of buyers in the sample and supply costs in the different versions of the managerial problem. Further assumptions are that two sample sizes exist (two different numbers/percentages of buyers) and that the market has asymmetric costs (either over-supplying costs or under-supplying costs). Moreover, the choices present three levels of difficulty. The impact of each statistical format in decision making undergoes quantification in terms of both the time required to make decisions and the accuracy of these decisions. Third, in this experiment, the number of factors to consider, namely the statistical formats and the different informational environments, is quite large. Specifically, five factors receive attention (statistical formats, sample size, number/percentage of buyers, costs and responses). Three of these factors have two levels, one factor has three levels and the remaining factor has four levels. The possible combinations of these factors define 96 different choices for the subjects. Such a high number of decisions might overwhelm subjects or require several sessions with different subjects. An orthogonal design whereby the number of alternatives reduces to 25 (a quarter of the total number) helps resolve this problem, thus allowing the experiment to run in just one session and with the same pool of subjects. The methodology of orthogonal design allows for a reduction in the number of decisions without substantially lowering the capacity for analysis (mainly related to the interaction order between factors). Thus, this approach also offers a contribution to the experimental design methodology.

The key findings of the paper are the following. (1) Statistical information presented in terms of relative frequency formats (i.e., percentage of buyers) gives rise to better decision making than information in terms of simple frequencies (i.e., the total number of buyers) independently of the informational environments. (2) Numerical formats are better than pictorial frequency formats. Therefore numerical relative frequency formats perform best for decision making. (3) The time required for decision making is lower under numerical formats than under pictorial ones. (4) Under pictorial statistical formats, relative frequencies perform faster than simple frequencies. (5) The sample size matters, and the lower the sample size, the lower the time needed to make decisions. Nevertheless, because the sample sizes are 100 and 500, the normalization of 100 may also drive the result as well as the smaller quantity. (6) None of the interactions between an informational environmental factor and the statistical format factor significantly affects either the accuracy or the time to make decisions. The outperforming of the numerical relative frequency format may lie in its ability to provide an appropriate set structure of the problem by reducing computational demands. Finally, the paper develops the managerial implications of these findings and establishes good practices to present statistical market information in business reporting.

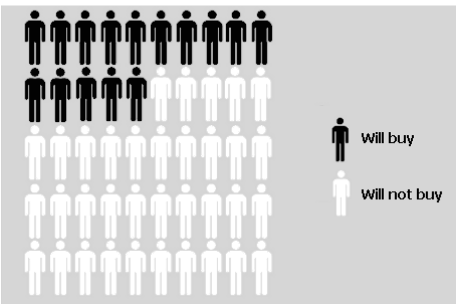
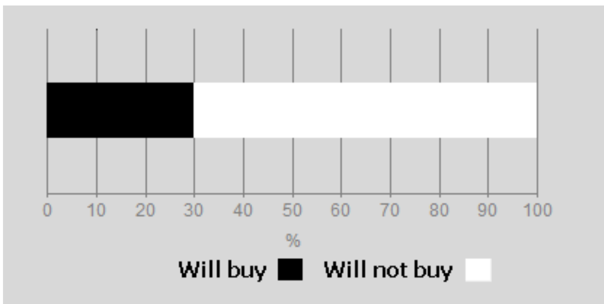
2. Experimental Design

An economic experiment provides an appropriate tool to examine which type of frequency (simple or relative) and which type of representation (numerical or pictorial) give more accurate and fast responses to managerial decision making. The experiment presents the following managerial decision-making environment. Subjects receive the results of a simple market research exercise and must choose one out of two alternative provision levels (number of units) of a good to be allocated in a market. The market research provides information on the percentage of the customers in the sample that will buy a unit of the good (only one unit per customer) under the assumption that the survey is completely reliable (a customer will buy a unit of this product if and only if she or he demonstrates her or his intention to buy during the survey). This information is displayed using four statistical formats:

Providing information with these four types of presentations avoids possible word confusions among the four statistical formats. Moreover, the cost of an unsold product unit (over-supply cost) and the cost of leaving a subject without the product (under-supply cost) are different. Subjects choose between two possible answers in each round. These two possible answers can be of three types: i) Correct and Focal; ii) Correct and Opposite; and iii)

Correct and Extreme. *Correct* means the right answer to the decision problem, namely that the number of units of the product to allocate to the market to maximize expected profit given the sample information, the market size and the asymmetry of supply costs. *Focal* refers to the option of allocating to the market a number of units proportional to the number of clients that will buy the product in the sample, with no other consideration on the asymmetry of over-supply and under-supply costs (i.e., given a market size of 1000 potential buyers, if 30% of clients in the sample buy the product, the number of units to allocate to the total market will be 300, no matter if the over-supply cost is larger or smaller than the under-supply cost). The answer *Opposite* refers to an answer that is smaller than the focal one, in those cases where the right answer is larger than the focal answer and vice versa. The final option is *Extreme*, whereby the number of units allocated to the market is larger than the right number in those cases where the right answer is also larger than the focal answer. The converse is also true, namely that a choice is extreme when the choice is lower than the right answer and the right answer is also smaller than the focal option.

Figure 1. Example of the four possible types of statistical formats.

<table>
<tr><td>Numerical Simple:
Of the 500 people interviewed, 150 state that they will buy the product.</td><td>Numerical Relative:
Of the 500 people interviewed, 30% state that they will buy the product.</td></tr>
<tr><td>Pictorial Simple:

</td><td>Pictorial Relative:

</td></tr>
</table>

The experiment considers two cost treatments (higher over-provisioning cost and higher under-provisioning cost). Each treatment includes two sample sizes (500 and 100) as well as two different proportions of elicited buyers in the sample (30% and 70%).

As previously mentioned, subjects had to deal with 96 possible decision-making problems (cards). The use of an orthogonal design reduced the number of cards to 25 (a quarter of the total). From the medical literature the experiment borrows the methodology of orthogonal designs, which is novel in the design of economic experiments. Such a design keeps the experiment manageable without substantially reducing its analytical power. Specifically, such a design enables the inclusion of multiple factors in an experiment because the sample size remains small, thus making the estimation of both the main effects and all the two-way interactions possible. However, this sample size reduction comes at a statistical power cost, rendering the estimation of some three-way interactions or higher-order interactions impossible. Therefore, the advantages of the orthogonal design are: i) obtaining information of the effects of all the factors in a single experiment of moderate size; ii) enabling the analysis of the effect of one factor under varied conditions of others; and iii) facilitating the analysis of the two-way interactions, which may be the most important reason to include more than one factor in an experiment (see Arribas, Comeig, Vila, and Urbano 2013).

3. Experimental procedures

The experiment took place in the Laboratory of Research in Experimental Economics (LINEEX) of the University of Valencia (Spain) in a session with 60 undergraduate students (39 males and 21 females) from several degrees related to business administration. The average ± standard deviation age was 21 ± 2.4. The session consisted of 25 rounds of decisions preceded by a detailed training session that included three unpaid practice rounds. The experiment randomly presented the twenty-five cards to each one of the subjects. The language was Spanish. Subjects received payment for twelve rounds drawn at random out of the twenty-five decisions they made. Subjects received 10 ECUs (the Experimental Laboratories standard ad hoc currency, with an arbitrary exchange rate with the euro) per correct answer. The session lasted approximately 120 minutes, and subjects earned 15€ on average. At the end of the experiment, subjects completed the Holt and Laury (2002) test to estimate their degree of risk aversion and received payment for one round drawn at random in the Holt and Laury task. Subjects also completed a standard socio-demographic questionnaire to control for differences in technical skills and wealth. The responses to this questionnaire reveal no significant differences in technical skills and wealth among subjects. Also, under a constant relative risk aversion utility function, an estimate of the average subject's risk aversion is 0.28 (slightly risk averse), which is in accordance with the generally observed risk aversion of the population.

4. Results

4.1 Analysis of the decisions' accuracy

On average, the subjects give right answers in 78.4% of their decisions. This figure varies greatly depending on the specific characteristic of the statistical representation formats (see Table 1). The lowest percentage of right answers emerges in informational environments with possible responses *Correct* and *Focal* (66.8%). On the contrary, high percentages of correct answers arise in environments with *Correct* and *Opposite* responses (90.8%) and statistical formats with Numerical relative frequency (83.3%). Other factors such as the sample size, the percentage of buyers, and costs do not affect the percentage of right answers (see Table 1).

Table 1. Right answers (%) and time to finish the 25 rounds (minutes). (Average ± S.D.)

Factor		Right answers	Time to finish
Total		78.4±18.1	40.1 ±30.6
Statistical format			
Numerical simple freq.		77.0 ± 42.1	35.2 ± 28.0
Numerical relative freq.		83.3 ± 37.3	38.4 ± 29.4
Pictorial Simple freq.		75.3 ± 43.2	48.9 ± 33.6
Pictorial Relative freq.		79.3 ± 40.6	42.9 ± 31.3
Informational environment			
Sample size			
	100	78.9 ± 40.8	35.1 ± 27.1
	500	77.7 ± 41.7	47.7 ± 33.8
Percentage of buyers			
	30%	78.9 ± 40.8	40.0 ± 31.2
	70%	77.7 ± 41.7	40.2 ± 29.6
Costs			
	Higher over-supply cost	79.6 ± 40.3	40.3 ± 31.4
	Higher under-supply cost	76.7 ± 42.3	39.9 ± 29.4
Responses			
	Correct and Focal	66.8 ± 47.1	42.6 ± 33.4
	Correct and Opposite	90.8 ± 28.9	35.5 ± 27.2
	Correct and Extreme	76.7 ± 42.4	44.5 ± 30.0

A Generalized Linear Random Intercept Model for Binary Responses tests whether the statistical formats and the informational environmental factors significantly affect the percentage of right answers. The following are the main findings. i) Statistical formats are relevant: in formats with numerical relative frequencies the percentage of right answers is significantly higher than in formats with simple frequencies (either numerical or pictorial, p-values equal 0.0162 and 0.0083, respectively). However, no significant difference emerges between numerical relative frequencies and pictorial relative frequencies (p-value = 0.1705). ii) Sample sizes, percentages of buyers, and costs do not affect the percentage of right answers (p-values > 0.1). iii) The percentage of right answers is significantly higher in informational environments whose responses include the *Correct* and the *Opposite* choice than the others (p-value < 0.001).

Expanding on finding i) leads to the conclusion that the statistical format is relevant. No significant difference exists in the percentage of right answers between numerical formats (simple frequencies and relative frequencies) and pictorial formats (p-value= 0.3670). However, the difference is significant in the percentage of right answers between simple formats (either numerical or pictorial) and relative formats (either numerical or pictorial) (percentages equal to 76.4% and 81.3%, respectively, p-value = 0.0134) (see Table 2). Thus, a key point is the use of relative frequencies (explicit inclusion of the percentage of buyers in the statistical format) rather than the use of simple frequencies.

Table 2. Right answers (%) and time to finish (minutes) as a function of statistical formats.

Factor	**Right answers**	**Time to finish**
Simple frequencies	76.4 ± 42.4	39.7 ± 30.7
Relative frequencies	81.3 ± 39.0	40.7 ± 30.4
Numerical formats	79.1 ± 40.7	36.3 ± 28.5
Pictorial formats	77.3 ± 41.9	45.9 ± 32.6

However, the effect of visualizing the percentage of buyers on the proportion of right answers depends on the representation used: numerical versus pictorial. Under pictorial statistical formats, no difference emerges in the percentage of right answers between simple frequencies and relative frequencies (percentages of 75.3% and 79.3%, respectively, p-value = 0.2080), meanwhile restricting the analysis to numerical formats provokes significant differences (percentages of 77.0% and 83.3%, respectively, p-value = 0.0166). Thus, the use of relative frequencies increases the percentage of right answers, especially under numerical presentations. Finally, with respect to the decision accuracy, results show an absence of significant two-way interactions between the informational environment factors and statistical formats.

4.2 Analysis of the time to make decisions

On average, subjects need 40.1 minutes to answer the 25 rounds. However, this time depends on both the informational environment factors and the statistical formats. The most relevant findings are the following. i) The time depends on the statistical format: the time needed to answer the rounds is significantly lower with the numerical relative frequency format than with pictorial formats—either simple or relative (p-value < 0.01)—whereas the time spent with pictorial simple frequency formats is significantly higher (p-value < 0.01). ii) The time spent does not depend on the percentage of buyers or costs (p-values > 0.1). iii) The time depends on the sample size: the large the sample, the longer the time (p-value < 0.0001). iv) When the responses include the *Correct* and the *Opposite* options, the time reduces significantly (p-value < 0.0001) (see Table 1).

The effect of the statistical formats on time is the opposite of that for the percentage of right answers. The time spent by the subjects differs significantly between numerical formats and pictorial formats (36.3 minutes and 45.9 minutes, respectively, p-value < 0.0001) (see Table 2). This difference also appears after restricting the analysis to simple frequencies (p-value < 0.0001) but vanishes in the case of relative ones (p-value = 0.0573). On the other hand, no difference exists in the time spent between simple frequencies and relative ones (39.7 minutes and 40.7 minutes, respectively, p-value = 0.5459). This situation also holds upon restricting the analysis to numerical frequencies (p-value = 0.0846), but a different result emerges in the case of pictorial formats. In the latter case, the time spent differs significantly between simple frequencies and relative frequencies (48.9 minutes and 42.9 minutes, respectively, p-value = 0.0153). Regarding the time to make decisions, results show no significant two-way interactions between the informational environment factors and statistical formats.

The relationship between the time to finish and the accuracy level is inverse and moderate but significant. The Pearson's correlacion coefficent is equal to -0.28, with a p-value = 0.0330 for testing the null hypothesis that the correlation is equal to zero. Thus, individuals with a higher number of hits need less time to finish.

5. Conclusions, implications and limitations

This paper is the first study of the effect of statistical formats on managerial decision making. The main conclusion of the analysis is that the statistical format used to present statistical information has a significant impact on both the time required to make a decision and its accuracy. A relevant finding is that statistical information for business reporting is better understood when presented in relative frequencies rather than in simple frequencies. As regards with the use of pictorial representations, the results of the experiment support that numerical formats are more convenient than graphical formats, since the former enable a more accurate and a quicker decision than the latter. Therefore, numerical relative frequency formats perform better for decision making.

The results of this paper have some relevant methodological and managerial implications. Regarding methodology, this paper presents the first application of an economic experiment investigating the impact of statistical formats on the accuracy and response time in business decision making. Moreover, the paper applies an innovative methodology in economic experiments: the use of an orthogonal design to define the combination of the factors (statistical formats and informational scenarios) under which the experimental subjects make choices. This methodology admits a larger number of factors without running different treatments with different groups of subjects.

The main managerial implications of these findings pertain to the improvement of statistical information presentations in different types of business reports. Simple frequencies are widespread as a statistical format in business reports, advertisement, newspapers, etc. The findings show that the use of simple frequencies can reduce the accuracy of evidence-based decisions, specifically when simple frequencies appear pictorially. Expressions such as *1 out of 4* clients… are less effective than *25% of clients…* and business reports should avoid pictorial simple frequency formats. Similarly, pictorial formats do not improve the accuracy of decision making and increase the time required to process the statistical information during the decision-making process. A good practice to write efficient and effective business reports is to use numerical relative statistical formats (percentages) for simple information and leave pictorial formats for more complex statistical information. In a general environment, where IT analysis tools facilitate and generalize the production of sophisticated reports, this study's conclusions support simplicity as good practice for an effective presentation of quantitative information in business reporting.

Finally, this paper has some limitations. For instance, the statistical information presented in this economic experiment is quite simple and easy to understand (sample market share of a product). The paper does not consider more complex statistical information where the visualization of pictorial statistical formats may increase the understanding of the information and then reduce the processing time and increase the accuracy of the decision. For instance, the graphical display of the distribution of the full history of returns of a risky asset may reduce the risk associated to investment products and help investors make more risky allocations than otherwise (Kauffmann, Weber, and Haisley, 2012). Future research will seek to investigate the importance of graphical displays for complex statistical information. Furthermore, as in many decisions the format of statistical information is not subject to the firm's or anybody's choice, managers should have the competence to evaluate correctly statistical information under any format. As Soyer and Hogarth (2010) point out, evaluating the predictive ability of statistical models should be an important component of the teaching of statistics and econometrics for future managers.

References

Arribas, I., Comeig, I., Urbano, A. and Vila, J. (2013). Orthogonal design in experimental economics, MIMEO.

Brase, G. L. (2002). Which Statistical Formats Facilitate What Decisions?, *Journal of Behavioral Decision Making,* 15, 381-401.

Brase, G. L. (2007). Omissions, conflations, and false dichotomies: Conceptual and empirical problems with the Barbey & Sloman account, *Behavioral and Brain Sciences,* 30, 258-259.

Butterworth, B. (2007). Why frequencies are natural, *Behavioral and Brain Sciences,* 30, 259-260.

Ghosh, A. K. and Ghosh, K. (2005). Translating evidence-based information into effective risk communication: Current challenges and opportunities, *Journal of Laboratory and Clinical Medicine*, 145(4), 171-180.

Gigerenzer, G. and Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insights, *British Medical Journal,* 327, 741-744.

Gigerenzer, G. and Hoffrage, U. (1995). How to improve Bayesian reasoning without instructions: frequency formats, *Psychological Review,* 102, 684-740.

Gigerenzer, G. and Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: a reply to Lewis and Keren (1999) and Mellers and McGraw (1999), *Psychological Review,* 106, 425-430.

Girotto, V. and Gonzalez, M. (2007). How to elicit sound probabilistic reasoning: Beyond word problems, *Behavioral and Brain Sciences,* 30, 268.

Hoffrage, U., Lindsey, S., Hertwig, R. and Gigerenzer, G. (2000). Communicating statistical information, *Science*, 290, 2261-2262.

Holt, C.A. and Laury, S.K. (2002). Risk aversion and Incentive Effects, *The American Economic Review*, 92(5), 1644-1655.

Kauffmann, C., Weber, M. and Haisley, E. (2012). The Role of Experience Sampling and Graphical Displays on One's Investment Risk Appetite, *Management Science,* article in press, http://dx.doi.org/10.1287/mnsc.1120.1607.

Macchi, L. (2000). Partitive formulation of information in probabilistic problems: beyond heuristics and frequency format explanations, *Organizational Behavior and Human Decision Processes,* 82, 217-236.

Macchi, L. and Bagassi, L. (2007). The underinformative formulation of conditional probability, *Behavioral and Brain Sciences,* 30, 274-275.

Macchi, L. and Mosconi, G. (1998). Computational features versus frequentist phrasing in the base-rate fallacy, *Swiss Journal of Psychology.* 57, 79-85.

Ni Y. and Zhou, Y. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias, *Educational Psychologist,* 40, 27-52.

Osman, M. and Shanks, D.R. (2005). Individual differences in causal learning and decision making, *Acta Psychological*, 120, 93-112.

Soyer, E. and Hogarth, R.M. (2010). Econometrics and decision making: Effects of presentation mode, *WP 1204, Universitat Pompeu Fabra*.