ARTICLE IN PRESS
JBR-08616; No of Pages 6
Journal of Business Research xxx (2015) xxx–xxx
Contents lists available at ScienceDirect
Journal of Business Research

Extracting business information from graphs: An eye tracking experiment☆

Jose Vila [a,*], Yolanda Gomez [b]

[a] Erices, LINEEX and University of Valencia,46022 Valencia, Spain
[b] LINEEX and University of Valencia,46022 Valencia, Spain

ARTICLE INFO

Article history:
Received 1 February 2015
Received in revised form 1 August 2015
Accepted 1 September 2015
Available online xxxx

Keywords:
Data visualization
Statistical graph
Eye tracking
Experimental economics

ABSTRACT

Information graphics are visualizations that convey information about data trends and distributions. Data visualization and the application of graphs is increasingly important in business decision making, for instance, in big data analysis. However, relatively little information exists about how people extract information from graphs and how the framing of the graphic design defines may 'nudge' and bias decision making. As a contribution to fill this gap, this study applies the methodology of experimental economics to the analysis of graph reading and processing to extract underlying information. Specifically, the study presents the results of an experiment whose baseline treatment includes graphical and numerical information. The authors analyze how the information extraction changes in other treatments after removing the numerical information. The experiment applies eye-tracking technology to uncover subtle cognitive processing stages that are otherwise difficult to observe in visualization evaluation studies. The conclusions of the study establish patterns in the process of graph analysis to optimize data visualization for business and policy decision making.

1. Introduction

Data visualization commonly helps to inform data users in a wide range of thematic areas, from health behavior to traffic accidents. Specifically, data visualization and the application of graphs are increasingly important in business decision making, policy making, and scientific research. Despite of the importance and almost universal application of data graphs and statistical charts, relatively little information exists on how people actually extract the information from graphical representations and how the framing of the graphic design defines may bias decision making and persuade users to a specific decision (Feeney, Hola, Liversedge, Findlay, & Metcalf, 2000; Lewandowsky & Spence, 1989). As a contribution to fill this gap, this study presents the results of an eye-tracking economic experiment to analyze how human beings actually scan and process statistical graphs to extract their underlying information. The application of the economic experimental methodology, including economic incentives for a successful completion of tasks, increases the validity of the results in comparison to the results from surveys or psychological experiments (Hernandez & Vila, 2014; Holt & Laury, 2002).

The study helps to answer a twofold research question. From a descriptive viewpoint, the study provides experiment-based insights on how real people look at a statistical chart and which are the pieces of information that they actually process and integrate in their decision-making process. The research also studies how different visualization patterns lead to more or less convenient decisions. From a *normative* viewpoint, this research provides guidelines to facilitate a more effective analysis of basic statistical charts. Answering these descriptive and normative questions has not only a scientific interest—the enhancement of the understanding of the cognitive processes founding active reading of statistical charts—but also a clear managerial implication, providing with an experimental sound contribution on how to improve statistical charts to lead to better decision making in business and policy.

2. Theoretical framework

The descriptive and normative research questions in the introduction are not new. Starting from the seminal work of Buswell (1935), this literature is exponentially growing during the last years with the quick development of big data methods and visualization software (Shixia, Weiwei, Yingcai, & Mengchen, 2014). However, most of these studies follow a computer science approach (Diehl, Beck, & Burch, 2010; Wang, Chou, Su, & Tsai, 2007) and do not focus on providing and analyzing empirical evidence on the underlying cognitive procedures or the effectiveness of alternative visualization frames for a proper understanding of the information in a graph.

The literature presents two experimental approaches to deal with these research questions. The first one is the 'black box' approach.

☆ The authors are grateful to contributions from Irene Comeig, Univertsity of Valencia, and Iván Arribas, University of Valencia, for their careful reading and suggestions on revising this essay. The authors thank the Ministry of Science and Technology, European FEDER Funds (ECO-2010-20,584), and Generalitat Valenciana (Excellence Programs Prometeo 2013/019).

* Corresponding author at: LINEEX-Erices and University of Valencia, Research Institutes Building, Serpis, 29, 46022 Valencia, Spain.
E-mail address: jose.e.vila@uv.es (J. Vila).

http://dx.doi.org/10.1016/j.jbusres.2015.10.048
0148-2963/© 2015 Elsevier Inc. All rights reserved.

Under this approach, experimental subjects receive the same information in different visualization frames. Then, these studies measure and compare a series of indicators of the quality of the decision making (accuracy of the answer to a question, time required to make the decision, etc.) among different frames. Although most studies apply psychological experiments, where subjects do not receive economic incentives, Arribas, Comeig, Urbano, and Vila (2014) provide an interesting example of the application of an incentivized economic experiment. They present statistical information using numerical and graphical representations and assess the impact of such statistical formats on both the optimality of market decisions and the time required to complete decision making. Although the 'black box' approach allows for analyzing the effectiveness of alternative visualization frames, this approach does not provide information on subjects' scanning patterns: order, number of times, and during how long subjects actually look at each specific element of the statistical chart, etc. A good understanding of the underlying cognitive processes and the establishment of evidence-based normative guidelines to improve data visualization requires this type of information and the consideration of alternative approaches.

A proposal for this alternative is data collection using eye-tracking tools, which are able to provide real-time information on where a subject is locating her or his eyes and allows for the identification of actual visualization patterns. Eye movements may seem very stochastic but are indicators of higher-level cognitive processes (Conati & Maclaren, 2008). An eye-tracking methodology helps uncover subtle cognitive processing stages that are otherwise difficult to observe in visualization evaluation studies. Although completion time and accuracy on specific tasks may indicate that differences or problems exist, a deeper understanding of visual scanning strategies on information graphics may help to determine specific guidelines for designing graphs and selecting graph types for particular datasets and tasks. An eye-tracking methodology can help to observe these visual scanning strategies, providing richer information beyond that available from response time and accuracy-based methodologies (Goldberg & Helfman, 2011). However, the price for the information that sophisticated non-intrusive eye-trackers provide is the impossibility of running experiments with large samples able to support reliable statistical inference. For instance, sample size is 32 subjects in Goldberg and Helfman (2011) or 38 subjects in Burch, Heinrich, Konevtsova, Hoferlin, and Weiskopf (2011). This tradeoff between information accuracy and sample size is not a specific issue of eye trackin, but is a common feature of all neuro-economic experiments, where the complexity and time consumption in data collection usually limits data analysis to descriptive analysis, precluding the application of more sophisticated inference methods or statistical hypothesis testing.

The literature presents several examples of the application of eye-tracking methods to analyze graph-scanning strategies. For instance, Goldberg and Helfman (2011) present an illustrative eye-tracking study to compare how radial and linear graphs support value lookup tasks for both one and two data dimensions. This experiment presents linear and radial versions of bar, line, area, and scatter graphs to the participants, who complete each a counterbalanced series of tasks. Eye tracking also helps to classify error strategies and to support the establishment of improvement guidelines in the design of radial and linear graphs. As another example, Burch et al. (2011) apply eye tracking to identify visual exploration behaviors of participants solving a typical hierarchy exploration task by inspecting a static tree diagram. To uncover exploration strategies, they examine fixation points, duration, and saccades of participants' gaze trajectories. Huang (2013) highlights that graph esthetics usually come from common senses and personal intuitions—thus, their relevance to effectiveness is not clear—and conducts two eye-tracking studies in an attempt to understand the underlying mechanism of edge crossings. These studies show that eye tracking is an effective method for gaining insights into how people read graphs and how the esthetics conditions human graph-reading behavior.

However, the literature includes no previous examples in of the design and implementation of incentivized economic experiments with eye tracking: recording subjects' scanning strategy while they read basic statistical charts with information for a successful development of tasks with an economic incentive.

The present study is a first attempt to fill this gap and get together the potentiality of eye-tracking methods with the reliability of incentivized economic experiments. To this end, experimental subjects observe simple horizontal bar charts with the information on the percentage of purchasers of ten movies. Then, participants answer a simple multiple-choice question (the percentage of sales of the fourth most purchased movie) using this information. The task has an economic incentive and those subjects who answer the question properly receive 2 euros. Graphs present the information in two different frames, defining the two treatments of the experiment. Treatment 1 does not provide the numeric value of sales percentage and this information only appears on the ax of the chart. Treatment 2 shows this numerical value beside the bars, making the information in the horizontal ax redundant. During graph reading, a non-intrusive eye tracker records subjects' scanning strategy. This type of eye tracker provided more accurate information (precise recording of eye movement almost at a pixel level) and has lower impact on the subjects.

The structure of this study is as follows. This section summarizes the theoretical framework and Section 3 presents the economic experiment and the features of the eye tracker. Section 3 also describes the methodology used to analyze eye-tracking data, based on the definition of areas of interest and descriptive comparison of heat maps. Section 4 summarizes the results of the experiment and Section 5 presents a brief discussion of these results.

3. Method and experimental design

An eye-tracking economic experiment provides an appropriate tool for this study to examine the visualization patterns when reading simple statistical graphs. In the experiment, subjects receive information on the sales levels of ten movies through a bar graph that represents the percentage of participant in a previous experiment that actually purchased each movie. In a first step, subjects look at the graph during 15 s for a general exploration. After that, during 30 s and under the graph, the screen shows the question 'Which is the sale percentage of the fourth most sold movie?' with four multiple-choice answer options: 'From 16% to 25%', 'From 26% to 35%', 'From 36% to 45%', and 'From 46% to 55%'. Subjects receive and economic incentive of 2 euros if they answer the question properly. The experiment considers two treatments. In the first one, the bar chart does not include the numerical values of the sales percentage of each movie. Treatment 3 actually shows these values beside the top of each bar (Fig. 1). To avoid learning effects, the experimental design is between subjects, with half of the sample participating in each of the treatments.

The eye-tracker system Tobii T120 records eye fixation data. A processor located in a dedicated computer embedding the eye tracker calculates the gaze data. This eye-tracking system integrates a 17-in. TFT monitor, allowing for non-invasive data collection, because with this experiment, subjects need not wear special eye-tracking glasses that could make them feel uncomfortable and affect their behavior during the experiment. The eye-tracking device has a sampling rate of 120 registers per second and its precision values measured on human eyes are based on stimulus points on the native TFT screen (1280×1024 pixels). The interaction and analysis software Tobii Studio version 3.2.1 streams the data for further processing.

The experimental software-development environment includes the web technologies php and MySQL database. Tobii Studio allows to display webpages to participants simultaneously to the recording process. This software uses the website URL or a local address to open the site in Internet Explorer and automatically records all mouse clicks, key strokes, and webpages visited during the test.

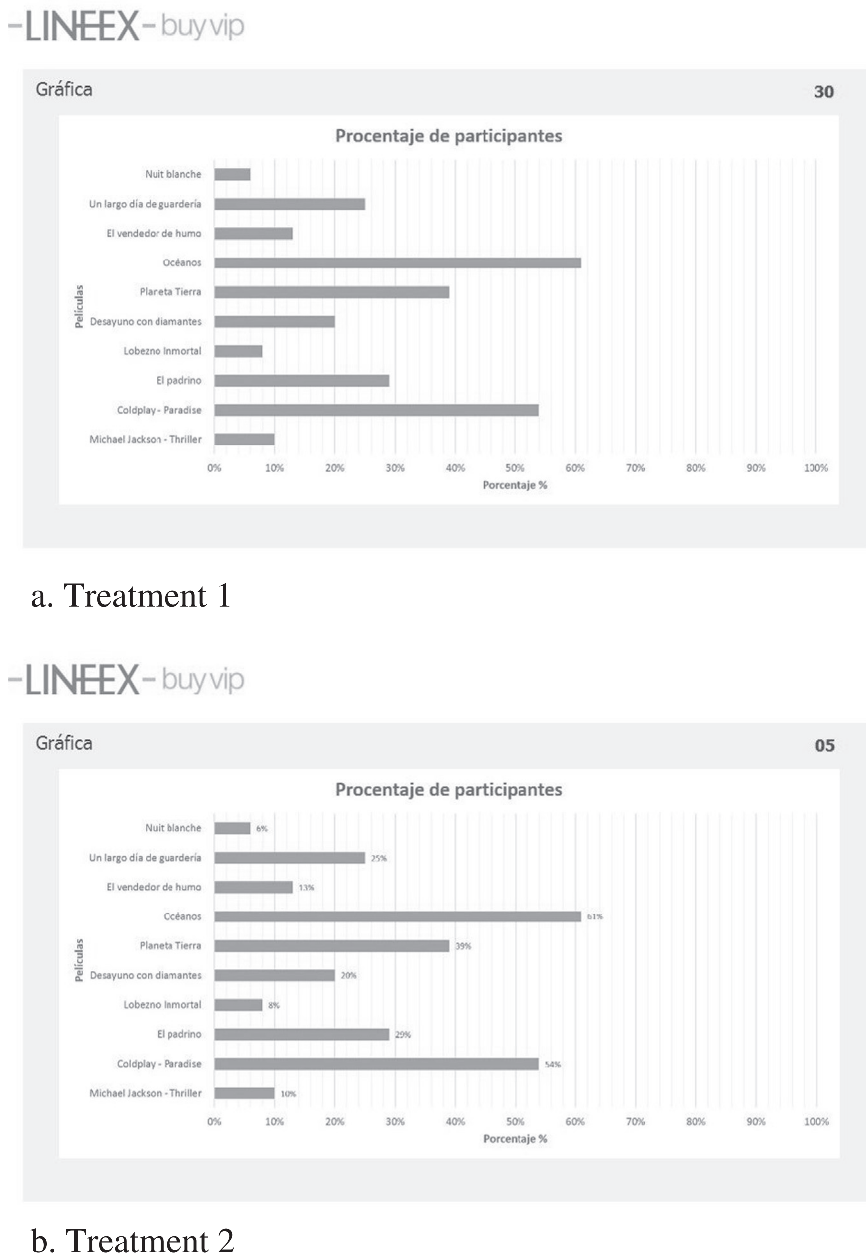

Fig. 1. Bar charts presented in both treatments.

To use the eye-tracking system, each subject participates in the experiment individually. Adjustment of the eye-tracker height and distance to each participant is necessary, as well as a 5-point calibration before each session. These characteristics of data collection make the procedure quite time consuming and preclude the use of large samples. This feature is common in most economic neuro-experiments. The final sampling size is 30 subjects, which allows to support a descriptive analysis. LINEEX, Laboratory for Research in Experimental and Behavioral Economics of the University of Valencia, was in charge of the production of the experimental software and the running of the experiment in April 2014.

One of the main challenges in the application of eye-tracking technologies in research is the definition of a procedure able to transform the large amount of information from the eye-tracker (coordinates of fixation of the eyes 120 times per second) into significant and interpretable results. Two complementary approaches to reaching this goal are the following:

- Establishment of areas of interest (AoI). Under this approach, the researchers select a series of areas in the computer screen and the eye tracker provides different metrics on how and when the subject has fixed her or his eyes within each of these areas. Some instances of these metrics are the time to first fixation of the eyes in the AoI or the total time that the subject fixes his or her eyes in the AoI, also called total visit duration (TVD). This study uses this last metric for the analysis of the eye-tracking data.

Fig. 2 shows the 45 areas of interest in the graphs, which correspond to the titles of the ten movies (10 AoIs); top of the bars in the histogram, including the numeric value of the percentage of sales in treatment 2

Please cite this article as: Vila, J., & Gomez, Y., Extracting business information from graphs: An eye tracking experiment, *Journal of Business Research* (2015), http://dx.doi.org/10.1016/j.jbusres.2015.10.048

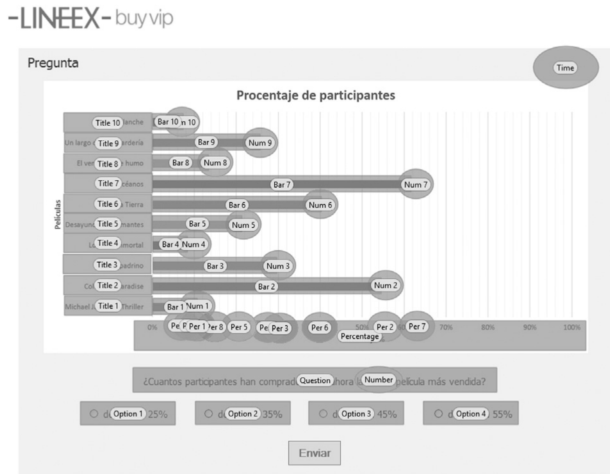

Fig. 2. Definition of the areas of interest, AoIs.

(10 AoIs); colored area corresponding to each bar (10 AoIs); areas in the horizontal ax (percentage ax) under the top of each bar (10 AoIs); text of the question (1 AoI) and potential answers to the question (4 AoIs). According to these definitions, a vector of values corresponding to the metrics applied to the 45 AoIs can summarize each visual exploration.

- Production of heat maps. These maps represent, using a greyscale, the total time that a subject o group of subjects fixed their eyes in each pixel of the image. The darker the color of the pixel, the larger the time spent in looking at this pixel. For instance, the TVD in the AoIs with the title of the movies of the subjects in Fig. 3.1.b is larger than that of the subjects represented in Fig. 3.1.a, because the heat map is darker in the first figure than in the latter in the areas of interest of the titles.

4. Results

This section presents the main results of the descriptive analysis of the visualization patterns of the subjects. The analysis distinguishes between two complementary groups of patterns: visualization patterns of the 43.3% of subjects who answer the question properly (for short, effective visualization patterns) and of 56.7% who are not able to provide the right answer (for short, ineffective visualization patterns).

4.1. Visualization patterns for a general exploration of the chart

Before the presentation of the question, 'Which is the sales percentage of the movie that is in fourth position is sales percentage?' subjects do not know yet which pieces of information are relevant for a successful performance the incentivized task.

Fig. 3 shows that effective visualization patterns have two main differential features. Firstly, subjects who answer the question properly spend less time looking at the chart and they concentrate their attention—the concentration of their TVD indicates—in a small number of AoIs. Specifically, effective visualization patterns focus only on AoIs covering the titles of the movies and the top of the bars (Treatment 1) or the numerical information (sales percentages in Treatment 2). Effective visualization patterns spend very little time to visualize non-informative AoIs such as the colored area of each bar or the horizontal ax in Treatment 2. On the other hand, ineffective visualization patterns

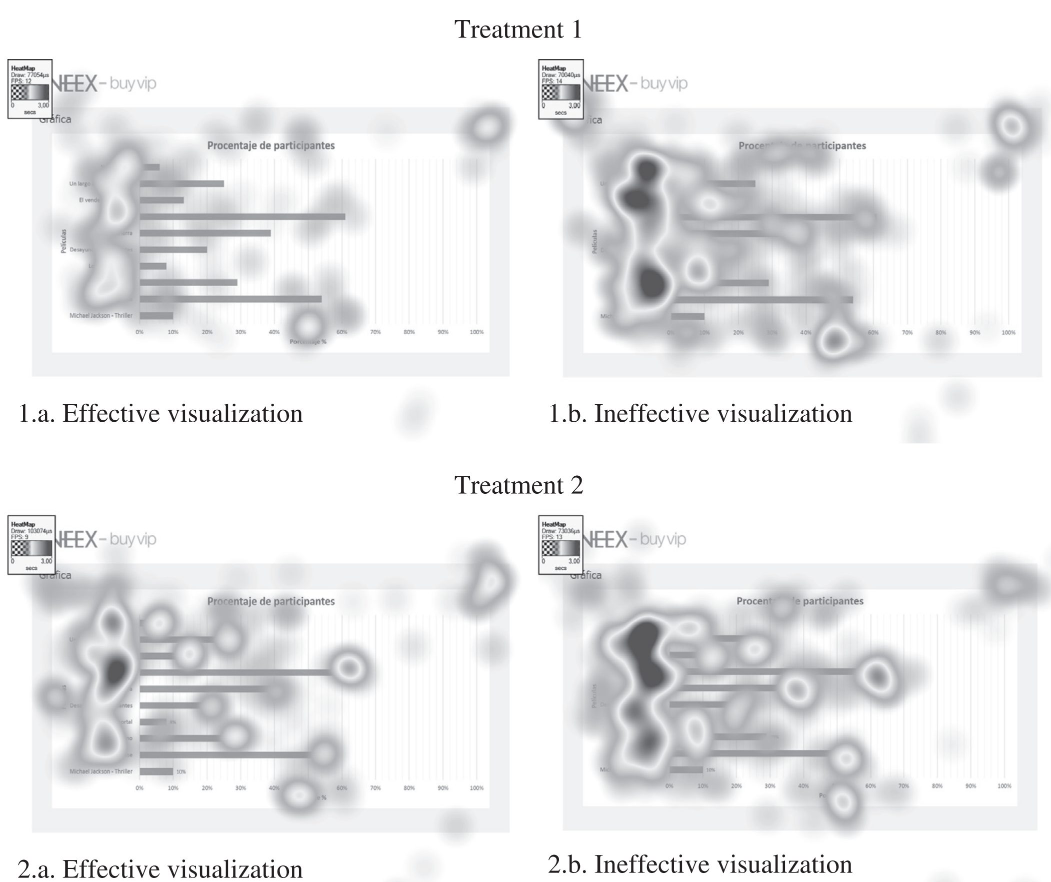

Fig. 3. Heat map of the effective and ineffective patterns before showing the question to be answered.

devote more time to look at the charts, visiting and spending TVD in most of the AoIs. Ineffective visualization includes eye fixation in the solid areas of the bars and in the horizontal ax, even in Treatment 2, where the information in the ax is completely redundant.

The inclusion of the sales percentages in Treatment 2 has no impact in the configuration of ineffective visualization patterns. However, a treatment effect arises for effective patterns: The inclusion of numerical information does not only increase the time required to read this information (time allocated to the 10 AoIs covering the top of the bars), but also increases the lecture detail of the titles of the movies.

4.2. Visualization patterns for answering the question

The presentation of the specific question defining the incentivized tasks, sales percentages of the fourth most purchased movie, discriminates the informative value of the different AoIs. For instance, the information on the titles of the movies is now completely irrelevant, as well as the sales percentages of all the movies but the fourth one.

Fig. 4 presents the heat maps for both treatments and effective and ineffective visualization patterns once subjects are aware of the specific question. One of the main conclusions of the descriptive analysis of these heat maps is that ineffective visualization patterns do not change with the disclosure of the informative value of each AoI. As in the first step, the total time to read the chart is longer and experimental subjects look at most of the AoIs, including titles. However, effective visualization patterns clearly discriminate between relevant and irrelevant areas, concentrating subjects' TVD in a small number of informative areas, namely, (1) the tops of the bars—Treatment 1—or the numerical percentages—Treatment 2—of the fourth most purchased movie and the other two movies with similar sales, and (2) the value of the horizontal ax corresponding to the sales level of the fourth most purchased movie, when this numerical information is not available beside the top of the bar in Treatment 1.

Treatment effect is similar to that detected in subsection 3.1. The inclusion of numerical information has no impact in how those subjects that do not answer the question look at the graph. However, effective visualization in Treatment 1 allocates a larger TVD in the AoIs covering the percentages and the top of the bars, specifically for the fourth most purchased movie and those with similar sales levels. In other words, the areas getting the attention in both treatments are the same, but the presence of numerical values increases the time required for the identification of the fourth most purchased film.

Finally, effective visualization patterns focus in the right answer (from 26% to 35%) and the next one (from 36% to 45%), while ineffective visualization allocates the TVD among the four possible answers. Eye fixation in the AoIs corresponding to different answers shows that

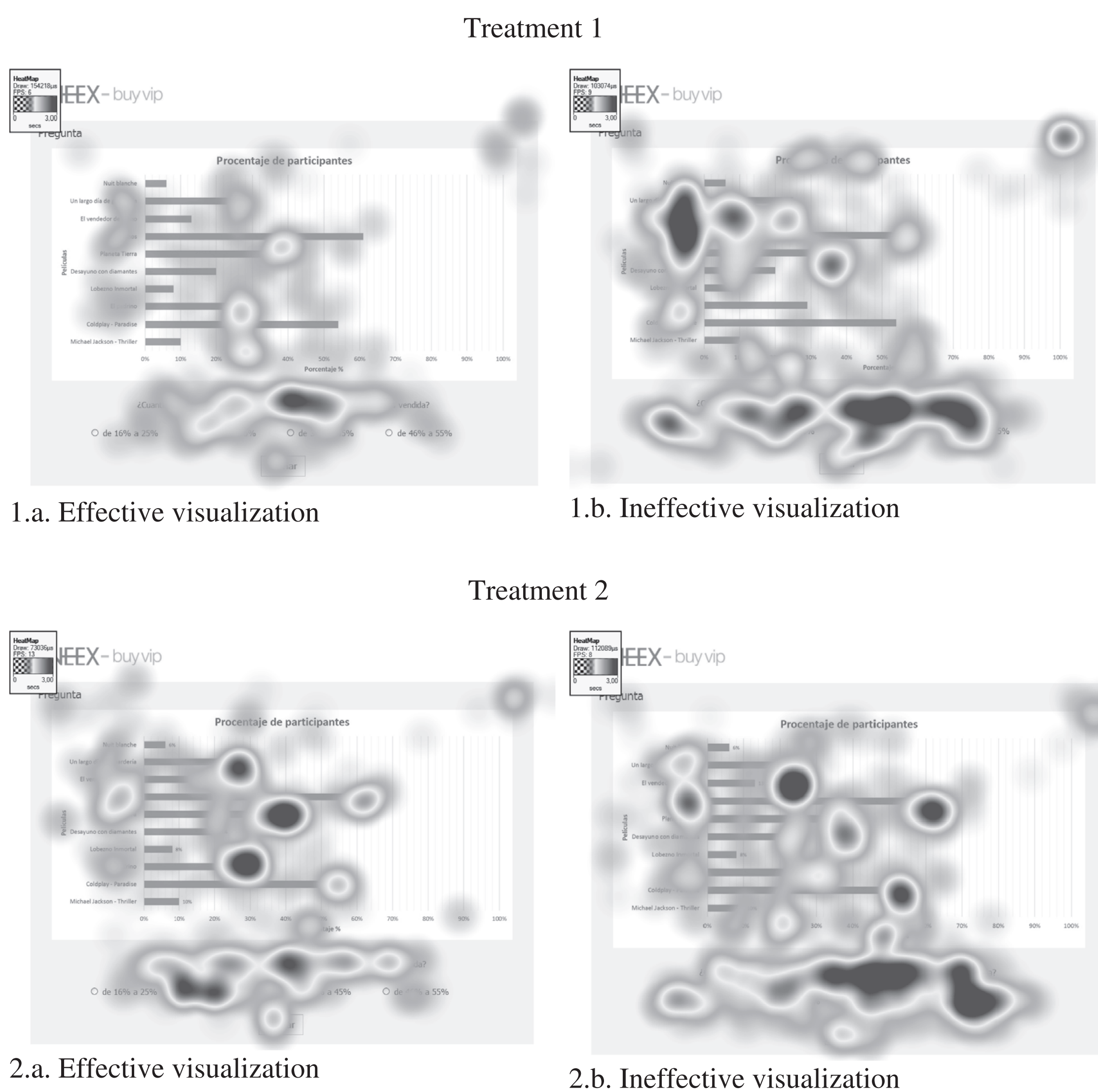

Fig. 4. Heat map of the effective and ineffective patterns after showing the question to be answered.

subjects failing to complete the incentivized task are actually doubting between different options before finally selecting a wrong option.

5. Conclusions, limitations, and implications

Although data visualization trough basic statistical graphs is a common practice to support evidence-driven business decision making, this study shows that the extraction of relevant information from such graphs may become a difficult task even in very simple situations. Specifically, 56.7% of subjects make an error when answering a simple question, even if the information required for a proper answer is available in the statistical chart shown in the screen during the decision process. The cause of such errors is not a lack of interest in performing the tasks, which had an economic incentive, as shown by (1) the long time that those subjects providing wrong answers spend in their scanning of the graph; (2) their visualization pattern, including a detailed reading of most of the areas of interest; and (3) the long time spent looking at the remaining time to complete the task. In other words, subjects with wrong answers seem to do their best to extract the required information but are not able to discriminate between relevant and irrelevant pieces of information.

The main result of the study is that visualization strategy is very different between the subjects who were able to answer the question properly and those providing a wrong answer. The key feature distinguishing both strategies is that subjects responding the question properly focused their attention in a small subset of relevant and informative AoIs, whereas the others spare their attention in a wider set of AoIs, some of them completely irrelevant to answer the proposed question.

As already highlighted in previous results in the literature (Arribas et al., 2014), the inclusion of numerical values in the graphs seems to have no impact in the accuracy of the response but increases the time required to extract the information. Specifically, the scanning time of those subjects providing a right answer is larger when looking at the graphs presenting the numerical values of the sales percentages beside the bars (Treatment 2).

The results of the experiment provide an empirical foundation for two key guidelines to improve chart presentation. Firstly, the experiment shows that those subjects who are not able to answer the question properly seem to have difficulties to discriminate relevant from irrelevant elements. The application of graphical framings that (1) stress the key pieces of information of the chart (for instance, the top of each bar) and (2) avoid highlighting other elements in the graph that provide no information and could distract the subjects during the scanning process could be useful to increase the effectiveness of data visualization. The role of the numerical values in the scanning process supports the establishment of a second guideline. Because these values increase the total time required in the scanning and deviate attention to other not very relevant elements such as the title of the movies, graphical representations that do not include numerical values could induce a better performance of the chart readers. A good alternative to the inclusion of numerical values could be the application of clear scales in the axes that facilitate the association of an approximate numerical value to relevant bars. In summary, and as highlighted in general in the visualization literature, a good data representation consists of very simple graphs (1) avoiding esthetical non-informative graphical elements and numerical information and (2) highlighting a small number of key and highly informative elements.

This study is also a good example of the role that eye tracking can play to understand complex cognitive processes. At the same time, this research is also a good example of the key limitations that the application of this methodology generates. The main limitation, as in most of neuro-economic and eye-tracking experiments, is that the sample size cannot be very large due to the high cost in terms of time and resources required for the collection of the data from each subject. This fact precludes the application of statistical inference methods and hypothesis testing. Then, the analysis needs to follow a descriptive approach and does not allow checking the statistical significance of the results. The small sample size of eye-tracking experiments does not allow for answering relevant questions on the profiles of the subjects that follow different scanning patterns. For instance, an important unanswered question in this study is a characterization in terms of socio-demographic and education levels of those subjects who apply an efficient visualization pattern and can complete the task successfully and those subjects who cannot. Finally, another limitation is the simplicity of the statistical information and the incentivized tasks in the experiment, as a consequence of the requirements of the eye-tracking technology and the large amount of data that the eye trackers generate for each observation.

The study opens some interesting questions for further experimental research. The above conclusions suggest the importance of the application of frames able to nudge the chart reader to focus her or his attention in the most relevant graphical elements and hiding these other elements that provide no information and could mislead the subject during information extraction. A proposal of such graphical frames and the analysis of their actual impact in the scanning processes is the goal of further eye-tracking economic experiments for different types of statistical charts and with different profiles of users.

References

Arribas, I., Comeig, I., Urbano, A., & Vila, J. (2014). Statistical formats to optimize evidence-based decision making: A behavioral approach. *Journal of Business Research*, *67*(5), 790–794.

Burch, M., Heinrich, J., Konevtsova, N., Hoferlin, M., & Weiskopf, D. (2011). Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12).

Buswell, G. T. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

Conati, C., & Maclaren, H. (2008). Exploring the role of individual differences in information visualization. *Proceedings of the Working Conference on Advanced Visual Interfaces* (pp. 199–206). New York: ACM Press.

Diehl, S., Beck, F., & Burch, M. (2010). Uncovering strengths and weaknesses of radial visualizations—an empirical approach. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 935–942.

Feeney, A., Hola, A.K.W., Liversedge, S.P., Findlay, J.M., & Metcalf, R. (2000). How people extract information from graphs. *Lecture Notes in Artificial Intelligence*.

Goldberg, J., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Journal of Information Visualization*, *10*(3), 182–195.

Hernandez, P., & Vila, J. (2014). Measuring value levers: Experimental and contingent approaches. *Journal of Business Research*, *67*(5), 775–778.

Holt, C., & Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Huang, W. (2013). Establishing aesthetics based on human graph reading behavior: Two eye tracking studies. *Personal and Ubiquitous Computing*, *17*(1), 93–105.

Lewandowsky, S., & Spence, I. (1989). The perception of statistical graphs. *Sociological Methods & Research*, *18*(2–3), 200–242.

Shixia, L., Weiwei, C., Yingcai, W., & Mengchen, L. (2014). A survey on information visualization: Recent advances and challenges. *The Visual Computer*, *30*(12), 1373–1393.

Wang, K.C., Chou, S.H., Su, C.J., & Tsai, H.Y. (2007). More information, stronger effectiveness? Different group package tour advertising components on web page. *Journal of Business Research*, *60*(4), 382–387.

Please cite this article as: Vila, J., & Gomez, Y., Extracting business information from graphs: An eye tracking experiment, *Journal of Business Research* (2015), http://dx.doi.org/10.1016/j.jbusres.2015.10.048