Applied Ergonomics 80 (2019) 168–174

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Display clutter and its effects on visual attention distribution and financial risk judgment

Svetlana Ognjanovic[a,*], Manfred Thüring[b], Ryan O. Murphy[c], Christoph Hölscher[a]

[a] ETH Zurich, Chair of Cognitive Science, Clausiusstrasse 59, 8092, Zürich, Switzerland
[b] Technische Universität Berlin, Faculty V of Mechanical Engineering and Transport Systems, Department of Psychology and Ergonomics, Chair of Cognitive Psychology and Cognitive Ergonomics, Marchstraße 23, D-10587, Berlin, Germany
[c] Morningstar Investment Management, Head of Decision Sciences, 22 W. Washington St, Chicago, IL, 60602, USA

ARTICLE INFO

Keywords:
Display clutter
Expertise
Eye-tracking
Stocks

ABSTRACT

Display clutter is a widely studied phenomenon in ergonomics, where information density and other properties of task-relevant visualizations are related to effective user performance and visual attention. This paper examines the impact of clutter in the context of financial stock visualizations. Depending on their expertise, traders can use a variety of different cues to judge the current and future value of a stock and to assess its riskiness. In our study, two groups of participants (novices and experts) judge the riskiness of 28 pairs of stocks under two clutter conditions (low and high). Consistency of judgments and group concordance serve as measures for judgment performance, while mean fixation duration, fixation frequency, and transition matrix density are employed to capture visual attention. Our results reveal significant effects of display clutter and expertise on both the performance measures as well as the visual attention measures.

1. Introduction

To judge the riskiness of a stock, one can attend to many different information cues, such as stock volatility, traded volume, stock price, opening price, daily price range, 52-weeks high and low, Beta measure, price chart, earnings per share, competitors' price change, and many others (Hens and Rieger, 2010). Customizable user interfaces of trading applications (Ziegler et al., 2008; Ang and Quek, 2006) can visualize all this stock information, but how much should they show to be supportive? At what point does additional information lead to a cluttered display which fails to help decision makers and instead causes inefficient visual behavior and rather inaccurate decisions?

Research from various domains has shown that more available information to the decision maker does not necessarily lead to more accurate judgments. In a study with clinical psychologists, psychology graduate students, and undergraduate students, participants were asked to judge a case based on different amounts of patient information (Oskamp, 1965). Beyond a certain amount, the accuracy of the judges decreased while their confidence increased thus leading to overconfidence. In this line of research, however, the focus was on the amount of information in relation to accuracy and confidence of the judges, but not on the visualization of information.

The current literature on financial information visualizations does not offer much empirical research that addresses the questions introduced above. However, in accounting research, the effects of information presentation format on judgment and decision making have been studied. Kelton et al. (Palmer, 1994) compared tables against graphs and concluded that a combination of both support financial decision making best (Kelton et al., 2010). In a more recent study, accounting visualizations were compared against interactive information presentation, or the combination of both. While interactivity alone led to overconfidence, the combination of visualizations and interactivity led to the most accurate decisions (Tang et al., 2014). Although the just mentioned work, and other decision-making research suggest that the presentation of a task (i.e., presentation format) can influence decision outcomes and confidence, the aspect of the abundance of visualized information (i.e., clutter) was not addressed. In the field of engineering psychology or ergonomics, display clutter has been studied extensively because it can lead to major safety issues by deteriorating user performance and visual attention in complex work environments. However, clutter is not a display property per se, but depends on certain characteristics of the user, the task, and the situation. With this in mind, the challenge is to find the ideal middle ground between excessive data and insufficient information (Moacdieh and Sarter, 2015). The consideration of expertise can help to find this middle ground. Expertise is usually defined as the optimal adaptation to tasks in a specified domain,

* Corresponding author.
E-mail address: svetlana.ognjanovic@gess.ethz.ch (S. Ognjanovic).

https://doi.org/10.1016/j.apergo.2019.05.008
Received 1 December 2017; Received in revised form 6 December 2018; Accepted 20 May 2019
Available online 03 June 2019

which results in outstanding performance (Boot and Ericsson, 2013). To understand the cognitive processes of expert performance, continuous measures are needed. Eye-tracking provides such measure (Holmqvist et al., 2011).

In the present paper, we aim to understand the interplay of clutter and expertise: How can display clutter influence financial judgment performance and visual attention distribution? What role does expertise play in this context? And which potential advantage do experts have over novices when exposed to high clutter?

In the following, we address these questions by reporting a study in which we investigated display clutter and expertise in connection with the visualization of financial stocks. Based on typical visualizations, we created two clutter levels for the same group of stocks. The level of expertise was varied by inviting two groups of participants, i.e. novices and experts in trading. Judgment performance and visual attention were measured under four resulting conditions.

1.1. Display clutter

Display clutter is usually associated with the abundance of information available to a system user that can have a negative influence on the interaction and the overall performance (Tullis, 1983; Rosenholtz et al., 2005). Since display clutter has been identified in many different domains (e.g. radar research, electro-optical imaging, aviation research), there are a number of different approaches to objectively define it, such as display density, display layout, target-background, and the task-irrelevance approach (Moacdieh and Sarter, 2015; Alexander et al., 2008). The display density approach was also referred to as “numerosity clutter” by Wickens, Hollands, Banbury, and Parasuraman (2013), based on the set-size effect (Palmer, 1994). With the increase of number of items on the screen, search time is expected to increase. Numerosity clutter is expected to hinder selective attention and therefore influence the visual search process (Wickens et al.,2013 Parasuraman). In the performance and attentional costs approach, the actual user-display interaction is measured and therefore includes display properties as well as user characteristics (Moacdieh and Sarter, 2015). We employ this extended view of clutter to capture the interaction between display properties and decision maker, based on performance and attention related measures. Performance costs in ergonomics have been operationalized by search time (to find a target or solve a task) and accuracy of the task output (e.g., number of false-positive or false-negative responses). Search time tends to increase with higher levels of clutter, whereas accuracy tends to decrease (Moacdieh and Sarter, 2015). A potential disadvantage in assessing clutter by taking performance as the solely measure, is the strong focus on task outcome rather than task process. Hence eye movement measures can serve as a valuable complement to more traditional performance measures. A number of studies have reported a delay in visual search due to clutter (Neider and Zelinsky, 2011). Depending on the type of visual search (e.g., local or global search), various eye-tracking metrics can be employed. Our interest lies in global search patterns in terms of mean fixation duration, fixation frequency, and the directness of scanpaths. Increased mean fixation duration has been associated with higher levels of clutter (Beck et al., 2010; Henderson et al., 2009). Furthermore, high number of fixations have been linked to decreased search efficiency in cluttered displays (Goldberg and Kotval, 1999; Grahame et al., 2004). We test the applicability of these clutter effects on financial visualizations. A measure that addresses the directness of global search, is the transition matrix density. It was introduced by Goldberg and Kotval (1999) and describes the visual scan path allocation over time. In their influential study, Goldberg and Kotval (1999) showed, among others, that the visual scan path was more random with an illogical arrangement of display features than with a logical one. This indicates that the measure is able to reflect the confusion of participants. We expect to induce this confusion with the high clutter condition and therefore predict a more random scanpath under this condition for this group of participants as well.

Most of the previous studies listed above focused on performance and attentional costs in relation to display clutter (Pankok and Kaber, 2018). Our work extends this approach by including different expertise levels of the decision maker into the investigation.

1.2. Eye movements and expertise

Experts require little time to perceive and encode information in their domain (Reingold and Sheridan, 2011). For example, perceptual advantages due to skilled performance were found in the seminal work on expertise in chess by de Groot, Gobet, and Jongman (de Groot et al., 1996) and Chase and Simon, 1973a, 1973b.

In other domains such as medicine or aviation, the central connection between skilled performance and eye movements have been greatly acknowledged as well. For example, in an early study with radiologists and medical students, experienced radiologists were able to detect a larger number of abnormalities, showed a more efficient scan path and required fewer fixations to find the abnormalities than the students (Kundel and La Follette, 1972). Jarodzka et al. (2010) showed that biologists (experts) not only performed more accurately and with a more efficient visual search, but their findings also indicated more diverse gaze patterns from experts than from novices (students). This suggests that experts employ different strategies to solve the same complex task (Jarodzka et al., 2010).

In the field of aviation research eye movement studies that compare different levels of expertise are not as widespread as in medical domains; however, the work that has been done does include display clutter. Beck et al. (2012), for example, reported that the accuracy in solving a search task was significantly better for pilots compared to students under a high-local-clutter condition. Pilots showed longer fixation durations than students (expertise main effect). Also, the authors found an interaction effect for the fixation duration measure, indicating higher fixation durations from pilots in all trial types except for low global clutter. In a study on website clutter, Grahame et al. (2004) reported that older participants, who were expected to have less experience with websites, performed slower in the high clutter condition than students. These findings highlight the interrelation between expertise and clutter and we expect to find the same type of interaction effects between clutter and expertise in the context of stock visualizations.

1.3. Financial risk judgment

To analyze the interaction between expertise and clutter of stock visualizations, we have to operationalize performance in the context of financial trading. We decided to concentrate on one aspect of stock evaluations that is especially important, i.e., the risk judgment of a stock. Before acting, the trader has to judge the stock in terms of its riskiness, so he can match the outcome with his risk preference and some other factors (e.g., existing wealth, previous experiences), to make a decision. However, many different cues are related to the riskiness of a stock. The trader weights all potentially relevant cues in a certain order to judge a respective stock (Hens and Rieger, 2010; Tversky et al., 1988). Volatility is the most widely used measure of objective risk and is expected to be weighted the highest. It represents the standard deviation of the annualized returns over a given period of time. We therefore take this measure as a starting point for the creation of our stock visualization stimuli, to ensure that the variance between our stock volatilities lies within a controlled range. Other risk attributes, such as Beta measure or the stock trading volume, are considered important as well, but their contribution to subjective risk perception cannot be precisely defined due to the complexity of the concept risk (Slovic, 2016). For this reason, we chose to use a method that allows for relative risk values, measured through pair-wise comparisons. Based on the output of pair-wise comparisons, the judgment consistency for each participant can be computed in terms of the number of circular triads or transitivity, which in our case will represent the individual risk judgment performance (Kendall and Smith, 1940; Bortz et al., 2008). Since it has been shown that experts differ from novices in their ability to provide coherent judgments and to make discriminations (Weiss et al., 2006; Einhorn, 1972, 1974; Spence and Brucks, 1997), we expect our expert participants to show more consistent risk judgment performance.

Group disagreement can be a sign of multidimensionality of the task at hand; for example, each participant considers a different attribute to judge the stimulus on the same (risk) criterion (Kendall and Smith, 1940; Bortz et al., 2008). We employ the concordance measure as a second variable of judgment performance and expect the experts to agree more with each other than the novices.

In the present study we examine how well the findings from other domains, on the effects of clutter and expertise, can be applied to financial stock visualizations. Heretofore, we are not aware of any empirical research on financial visualizations associated with display clutter and expertise effects. We relate two judgment measures to the display clutter factor, consistency and concordance, to test the relative performance of users with different expertise levels. We furthermore employ the eye-tracking methodology to capture the visual attention processes during risk judgment as well as potential attentional costs induced by the high clutter condition.

1.4. Hypotheses

Our statistical hypotheses for performance, confidence and eye movement measures were as follows:

H1.1 Consistency, as a measure of individual risk judgment performance, will be negatively influenced by the high *clutter* condition.
H1.2 The consistency of *experts* will be higher than the consistency coefficients of *novices*.
H1.3 An interaction effect of *clutter* x *expertise* will appear with a negative influence of high clutter on the judgment consistency of novices.
H2.1 Concordance, as a measure of rank proximity between individuals and their group, will be negatively influenced by the high *clutter* condition.
H2.2 The concordance of *expert group* will be higher than the concordance measures of the *novice group*.
H2.3 An interaction effect of *clutter* x *expertise* will appear with a negative influence of high clutter on the concordance of the novice group.
H3.1 The self-reported confidence levels will be positively influenced by the high *clutter* condition.
H3.2 The confidence levels of *experts* will be higher than the reported confidence levels of *novices*.
H3.3 An interaction effect of *clutter* x *expertise* will occur for the confidence levels.
H4.1 Mean fixation duration will be increased in the high *clutter* condition.
H4.2 The mean fixation duration of *experts* will be higher than the mean fixation duration of *novices*.
H4.3 An interaction effect of *clutter* x *expertise* will appear with higher fixation durations for experts under the high clutter condition.
H5.1 Fixation frequency will be higher in the high *clutter* condition.
H5.2 The fixation frequency of *experts* will be lower than the fixation frequency of the *novices*.
H5.3 An interaction effect of *clutter* x *expertise* will appear with higher fixation frequencies for novices in the high clutter condition.
H6.1 Transition matrix density, as a measure of directness of visual search behavior, will be higher in the high *clutter* condition.
H6.2 The transition matrix density of *experts* will be lower than the one of *novices*.
H6.3 An interaction effect of *clutter* x *expertise* will lead to higher transition matrix density for novices in the high clutter condition.

2. Experiment

2.1. Methods

The experiment was based on a 2×2 factorial, between-participants design with two display clutter levels and two expertise groups. The dependent variables included behavioral and eye movement measures. The behavioral measures were judgment consistency (as a relative performance measure), group concordance, and judgment confidence. The eye movement measures were mean fixation duration (in milliseconds), fixation frequency (count per second), and the transition matrix density (in %) based on the definition by Goldberg and Kotval (1999).

2.1.1. Participants

A total of 98 participants volunteered to take part in our experiment by signing up through the University of Zurich and ETH Zurich online recruiting tool. An online pre-survey was used to screen the background of potential participants. We recruited students with a background in Economics without any trading experience, as well as students with investment banking internship (i.e., min. 3 months full-time internship) or stock trading experience (i.e., regular trading in free time, or industry work experience). While the first group has no trading experience, the second one is familiar with the basics of trading and the key performance indicators of stocks. To distinguish both kinds of participants, we call them novices and experts. Although the second group is definitely more knowledgeable than the first one, it must be noted that its members are no professional traders. The participants from our expert group ($n = 50$) were on average 25 years old (standard deviation [SD] = 2.3), including 37 males and 13 females; the participants belonging to the novice group ($n = 48$) were on average 24.2 years old ($SD = 2.9$), with 27 males and 21 females. Participants from both groups were randomly assigned to the low or high clutter condition.

2.1.2. Stimuli

For each clutter condition, the same stocks were used. The stimuli represented real company stocks, obtained from Yahoo finance, and were carefully selected beforehand. Eleven different stimuli, eight stocks for the main experiment and three for the training phase, were created in collaboration with a financial domain expert for each display clutter condition. The layout of our stock visualizations is inspired by standard stock information tools, such as Thomson Reuters, eTrade, Yahoo Finance, and Google Finance. This resulted in a generic financial interface with static stock information.

Fig. 1 shows the versions of high and low clutter for one stock. The header area consisted of price, opening price, day range 52-week range, average volume, and Beta value; the price range and the Beta values were consistent across the eight stocks. In the middle of each stimulus, a graph with a 12 months price line was shown, which was centered across the stocks to make it comparable. The price line was controlled for skewness, kurtosis, run length, and final trend (i.e., going up or down). The eight stocks were ranked in terms of objective riskiness based on their volatility (i.e., standard deviation of return). In the high clutter condition, additional information was shown below the graph. The lower left table contained P/E ratio trailing 12 months, EPS trailing 12 months, shares outstanding, dividend yield, and quarterly dividend. The table on the lower right included the name, price, and price change from seven competitor companies. All company names and competitors' names were invented to avoid any biasing due to previous knowledge about specific companies.

2.1.3. Apparatus

A high-resolution monitor with a screen size of 24 inches and an image resolution of 3840×2160 was used to present the stimuli. Eye

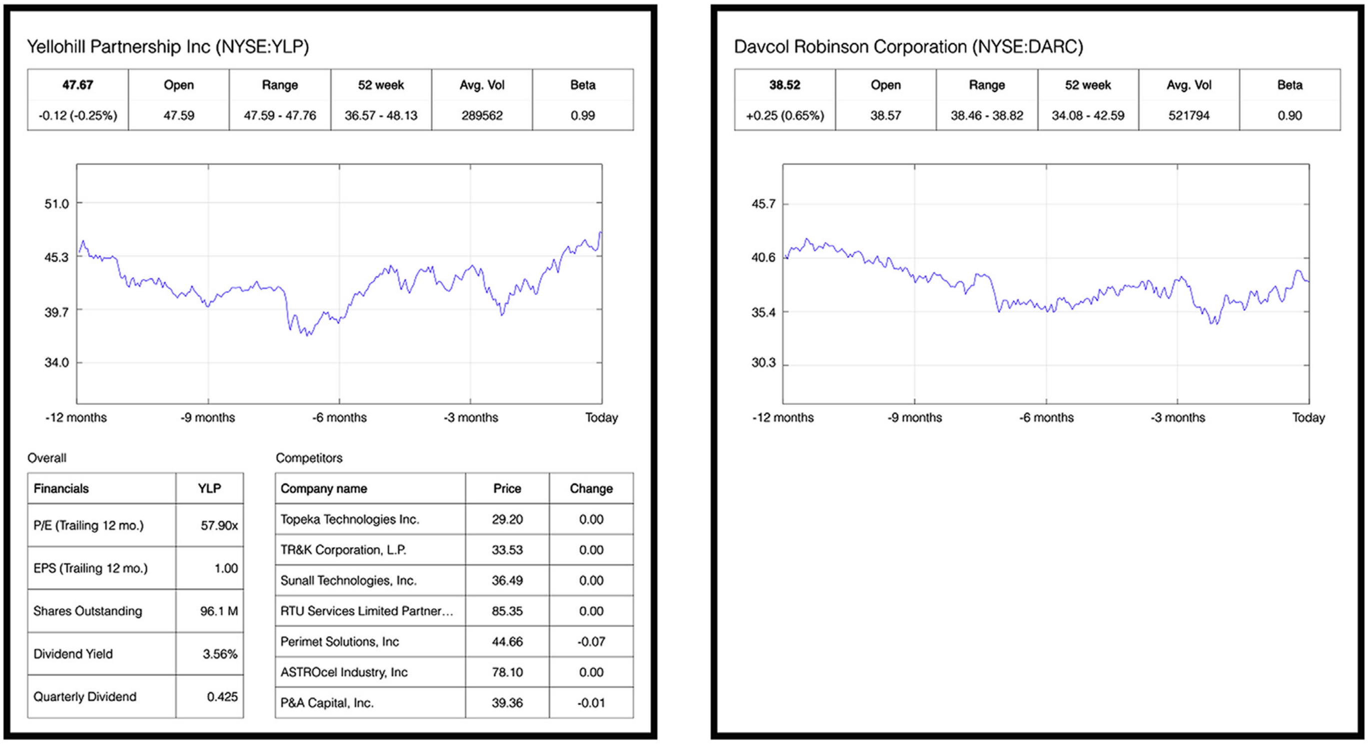

Fig. 1. High and low clutter stock visualizations. For each of the eight stocks, a high and a low clutter version was designed.

movements were recorded with the system SMI RED250mobile, which provides a sampling rate of 250 Hz. It is a remote eye-tracking system, allowing a monitor-mounted and contact-free setup. After calibration, the average error of visual angle in this system is 0.4°. PsychoPy experimental presentation software was used for stimulus presentation and data collection (Peirce et al., 2011). The luminance of the stimuli presented on the monitor was controlled with a lux meter.

2.1.4. Procedure

Each participant was welcomed by the researcher and asked to read and sign an informed consent form. Then the participants completed a demographic questionnaire. Detailed written instructions were given onscreen, and three training trials were run to introduce the participants to the experimental task. Before the eye tracking started, the participants' chair position was fixed to keep their height and their distance to the display (60–70 cm) constant. Next, participants underwent a standard 5-point calibration procedure followed by the test phase. This phase consisted of 28 trials, in which a complete pairwise comparison of the eight stock stimuli, i.e., $(8 \times 7)/2$, was performed in random order. In each trial, a pair of stocks was presented on the screen and the participants had to click on the stock that they considered as riskier. Before each trial, a black cross was presented in the center of a white screen, to control the gaze position. After each trial, participants rated the confidence in their judgment. After 14 trials the calibration procedure was repeated to correct for any degradation of eye-tracking precision. At the end of the experimental phase, participants filled out a questionnaire to capture their previous experience in trading. Finally, they were paid 30 Swiss francs for their participation.

2.1.5. Measures and analyses

For data visualization and statistical analysis, the commercial Software Package for Social Sciences (SPSS) was used (IBM). The distribution of data was checked by analyzing the standardized residuals and running the Shapiro Wilk test of normality. When the data was normally distributed a 2×2 factorial analysis of variance (ANOVA) for independent samples was computed, otherwise the hypothesis was tested with the non-parametric Mann-Whitney test.

Two dependent measures served to investigate the quality of risk judgments: consistency in terms of transitivity and concordance in terms of rank proximity between individual participants and their group. Additionally, the subjective confidence in their forced-choices was rated by each participant.

Consistency. The consistency coefficient K is a measure, where the number of circular triads d observed by individual participants are divided by the maximum possible number of triads d_{max} depending on the total number of stimuli, which was eight stocks in our experiment, and subtracted from 1 (Kendall and Smith, 1940; Bortz et al., 2008). A circular triad appears in case of intransitive judgment combinations; if stock *a* was perceived as riskier than stock *b*, and stock *b* was riskier than *c*, then stock *a* should be chosen over stock *c*; by choosing stock *c* instead of *a*, the participant produces a circular triad and therefore this combination of judgments would be considered inconsistent (Fig. 2).

To calculate the number of triads d we produced a dominance matrix for each participant (Bortz et al., 2008) and added up the respective preference scores across the eight stock stimuli N. Based on the preference scores S_i we were able to identify the circular triads d. This data processing was performed off-line using the commercial software package Matlab (8.6 and The MathWorks, 2015).

Concordance. This measure was used to analyze the similarity of judgments under each of the four experimental conditions. Based on the preference scores S_i, a rank order of the eight stocks was established for each participant. The dominance matrices from all members of a group were aggregated and a group stock rank order was created. These were the same dominance matrices that had been produced earlier to compute the consistency coefficient. Kendall's tau correlations were computed between individual participant's ranking and the ranking of the according group.

Confidence. The confidence measure was based on a continuous scale ranging from 50% (not certain at all) to 100% (completely certain). Participants saw a triangle shaped slider on top of the scale as soon as they started moving the mouse along the scale. When clicking on the scale, the value from that position was logged and the next

$$K = 1 - d/d_{max}$$

$$d = \frac{N(N-1)(2N-1)}{12} - \frac{1}{2}\sum_{i=1}^{N} S_i^2$$

$$d_{max} = N(N^2-4)/24$$

N=6	A	B	C	D	E	F	S_i
A		+	+	+	+	+	5
B	–		–	+	–	+	2
C	–	+		–	–	–	1
D	–	–	+		+	+	3
E	–	+	+	–		+	3
F	–	–	+	–	–		1

Fig. 2. Formulas for the consistency coefficient K and circular triads d, as well as a dominance matrix that represents preference scores, which are needed to calculate circular triads.

screen initiated.

Eye-tracking measures. The eye-tracking measures adopted in this experiment were mean fixation duration (ms), fixation frequency (count/s) across all 28 trials per participant, and the transition matrix density (%) for the scanpath analysis (Goldberg and Kotval, 1999). The experimental output file containing the eye movement data was imported into the SMI BeGaze software for further processing. The eye-tracking measures were computed based on the velocity-based event detection algorithm as part of this software. The following saccade detection parameters were set in SMI BeGaze: Peak velocity threshold of 40°/s, minimal fixation duration of 100 ms, peak velocity start at 20% of saccade length and end at 80% of saccade length.

Transition matrix density. This measure was used to analyze participants' visual scan path during the pairwise comparisons. For all pairs of stimuli, gridded areas of interest (AOIs) were created to divide the visible screen space into content-independent areas. The size of the grids was chosen based on the resolution accuracy of the eye-tracking device, which resulted in a grid of 17 × 30 cells and therefore 510 AOIs. Based on the AOIs, a transition matrix was computed for each participant and trial by using the SMI BeGaze software. To determine the density of the matrix, the R statistics package was used. The sum of non-zero cells was divided by the total number of matrix cells. Also, the matrix diagonal was removed, since scan path transitions within the same AOI were counted as transitions (Holmqvist et al., 2011).

3. Results

For all six dependent variables, inferential statistics were performed to test the hypotheses listed above. All effects are reported significant at $p < .05$.

Consistency. Since the consistency coefficients were not normally distributed, the non-parametric Mann-Whitney test was used for analysis. The judgment consistencies did not differ between the low clutter ($Mdn = 0.90$) and the high clutter ($Mdn = 0.90$) groups, $U = 1101.0$, $z = -0.63$, *ns*, $r = -0.06$. The expertise, however, had a significant influence on judgment consistency. The experts ($Mdn = 0.93$) were more often consistent than the novices ($Mdn = 0.85$), $U = 911.5$, $z = -2.09$, $p < .05$, $r = 0.21$. Additionally, in the high clutter condition the consistency of the experts was higher ($Mdn = 0.95$) than the consistency of the novices ($Mdn = 0.75$), $U = 159.0$, $z = -1.97$, $p < .05$, $r = -0.30$ (Fig. 3), which was in line with our hypothesis.

Concordance. Kendall's Tau correlation coefficients were used to evaluate the rank proximity between an individual participant and the according group. These coefficients were further processed to test the concordance hypotheses. Since they were not normally distributed, the Mann Whitney test was employed again. A clutter effect was found, indicating a higher group concordance in the low clutter condition ($Mdn = 0.72$) than in the high clutter condition ($Mdn = 0.56$), $U = 690.5$, $z = -3.55$, $p < .05$, $r = -0.36$. The expertise levels did not have any effect on the concordance measure, i.e., the experts did not agree more with each other ($Mdn = 0.66$) than the novices ($Mdn = 0.64$), $U = 1158.0$, $z = -0.30$, *ns*, $r = -0.03$. Also, in the high clutter condition, the two expertise levels did not differ from each other (Fig. 3).

Confidence. The confidence ratings were normally distributed. An ANOVA revealed that confidence measures between the low clutter condition (M = 0.79, SD = 0.07) and the high clutter condition (M = 0.80, SD = 0.08) did not differ, F (1, 94) = 0.45, $p = .50$, $d = -0.13$. Nor was the confidence of the experts (M = 0.81, SD = 0.08) higher than the confidence of novices (M = 0.79, SD = 0.07), F (1, 94) = 1.24, $p = .13$, $d = 0.27$. There was no significant interaction effect.

Fixation duration. The fixation duration measures were normally distributed. Therefore, an ANOVA was computed. The mean fixation duration did not differ between the low (M = 270 ms, SD = 28) and the high clutter condition (M = 268 ms, SD = 29) levels, F (1, 94) = 0.18, $p = .34$, $d = 0.10$. The main effect of expertise was not significant neither, but we found a trend towards our stated hypothesis, F (1, 94) = 2.48, $p = .06$, $d = 0.29$. Also, there was a significant interaction effect of clutter and expertise, F (1, 94) = 3.37, $p < .05$. Contrast tests showed that experts' fixation duration (M = 277 ms, SD = 26) was significantly longer than those of novices (M = 258 ms, SD = 26) in the high clutter condition, F (1, 94) = 5.28, $p < .05$, $d = 0.70$ (Fig. 4).

Fixation frequency. The fixation frequency measures were normally distributed and an ANOVA was computed. The fixation frequency (count/s) did not differ between the low and the high clutter condition. Also, the difference between the two expertise levels did not reach the required significance level of 5%, but there was a trend, F (1, 94) = 2.03, $p = .079$, $d = 0.23$. Furthermore, an interaction effect between expertise and clutter was found, F (1, 94) = 3.46, $p < .05$ (Fig. 4). Contrast tests revealed that in the high clutter condition, novices showed a higher fixation frequency (M = 3.75, SD = 0.55) than the experts (M = 3.44, SD = 0.41), F (1, 94) = 4.90, $p < .05$, $d = 0.64$ (Fig. 4).

Transition matrix density. This measure was not normally distributed. Therefore, the Mann-Whitney test was used to test the hypotheses. It showed a significant difference between the two clutter conditions for the transition matrix density measures, U = 812.5, z = −2.68, p < .05, r = −0.27. The two expertise groups did not significantly differ in the way they searched for information, U = 1137.0, z = −0.45, p = .33, r = −0.05. Also, in the high clutter condition, no differences appeared between the two expertise groups.

4. Discussion

The purpose of this study was to understand how display clutter can affect visual attention distribution and risk judgment performance in the context of financial stock visualizations. Our results support the assumption that not only display clutter, but also the expertise of the decision maker has significant implications for the interaction with financial information presented as a combination of text and graphics.

We found that the individual judgment performance in terms of consistency differed between the expertise groups, but not between the clutter conditions. This confirms our hypothesis on the expected ability

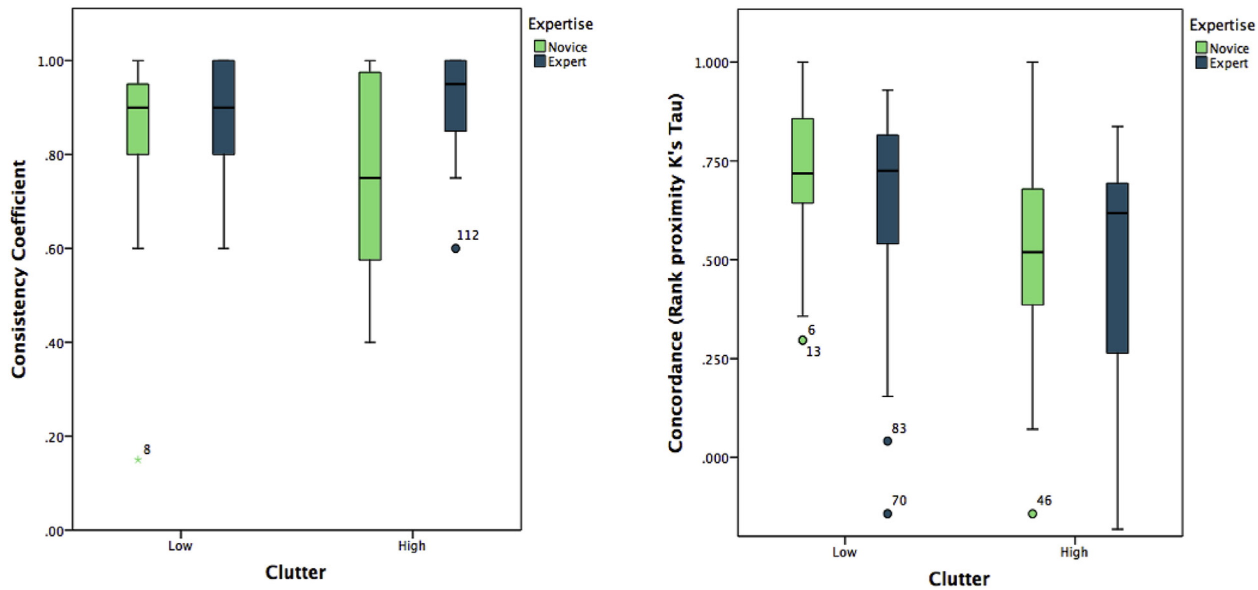

Fig. 3. Medians of consistency and concordance coefficients for clutter and expertise levels.

of experts to provide more coherent judgments than novices (Einhorn, 1972, 1974; Spence and Brucks, 1997). The effect size of this result was in the small/medium range. The reason for the rather small effect size may result from the fact that the experts, despite their experience in trading, were not as familiar with the type of stock information we provided. Moreover, the effect may be greater when professional traders instead of knowledgeable students serve as participants. The expected interaction effect revealed the negative influence of high clutter on novice participants' performance with a medium effect size. This finding is in line with other research in the field of ergonomics (Beck et al., 2012).

Our results on group concordance indicate a clutter effect of a medium size. In the high clutter condition, participants agreed significantly less with each other, than in the low clutter condition. This indicates a multidimensionality aspect of our task criterion (Bortz et al., 2008):

Participants judged the stimuli more or less consistently on an intra-individual level, when more attributes were available to them however (i.e., in high clutter), they chose other attribute combinations to judge the riskiness of the stock. The group concordance did not differ between expertise levels. Judgment confidence ratings were collected to test whether more information cues, i.e., high clutter, and more experience would lead to more confidence. However, the results do not support our hypotheses in that respect. The difference between the expert and novice participants in terms of number of years of experience may not have been pronounced enough to produce a confidence effect.

Display clutter and expertise had a significant effect on the distribution of visual attention. The results indicate that an interaction of these variables (clutter x expertise) impact both mean fixation duration as well as fixation frequency. Fixation durations were affected with a medium to large effect size, so that the fixation duration of experts was significantly higher than the one of novices under the high clutter condition. Additionally, fixation frequencies were influenced by the interaction with a medium to large effect size, suggesting that the experts did focus their attention on fewer and potentially more relevant items on the screen under the high clutter condition. This confirms our hypotheses and is in line with reported ergonomics research (Beck et al., 2012).

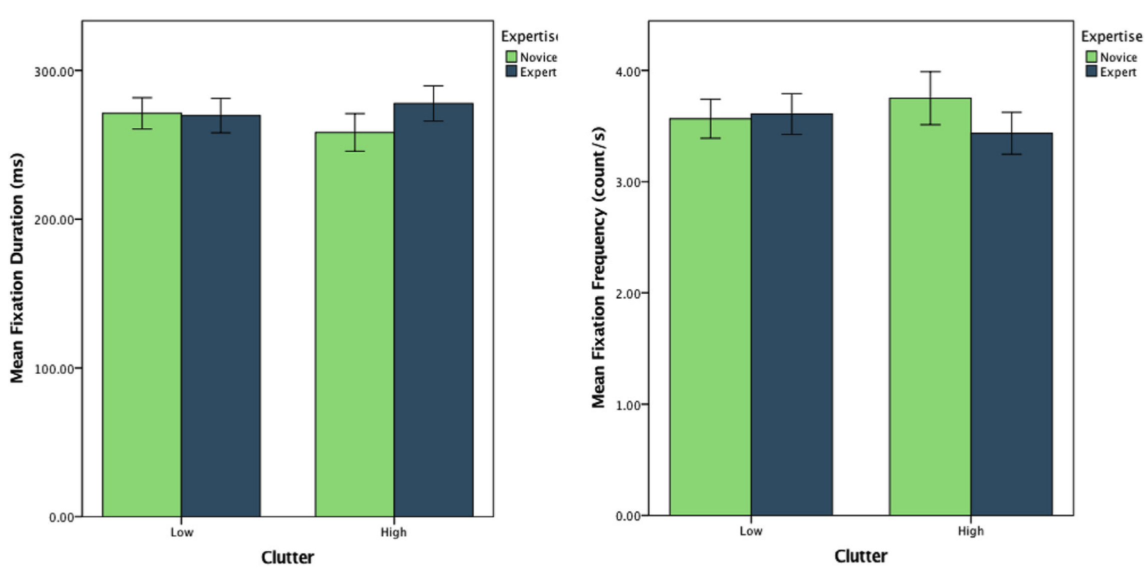

Fig. 4. Mean fixation duration and fixation frequency for low and high clutter condition.

For the mean fixation duration measure, the hypothesis on the expertise differences could not be confirmed because it did not reach the 5% significance level. However, there was a trend in favor of our hypothesis. As for the pure clutter impact on fixation durations and frequencies, our data did not support the hypotheses. The high clutter condition may not have been "cluttered enough" to have a sufficient impact on the experts' visual attention and judgment performance.

The directness of visual scanpath, which was measured by transition matrix density, was significantly influenced by clutter, but not by expertise. As predicted, the low clutter condition led to more directed visual scanpath and the high clutter condition to more transitions and thus to more random scanpaths (Goldberg and Kotval, 1999). The effect size was in the medium range, maybe because the stimuli did not differ enough to cause a large effect, especially for the expert group.

5. Conclusion

Our findings suggest that clutter in financial visualizations has a negative impact on the judgment performance and visual information processing of novices or lay people compared to experts. Furthermore, our results indicate that high clutter leads to some disagreement within groups and to rather random visual scanning behavior. Additionally, we found evidence that previous experience with stock trading leads to better risk judgment performance.

Our study is a first step to investigate effects of clutter and expertise in the context of financial visualizations. More research is needed to provide comprehensive insights into the consequences for visual attention and judgmental performance. For instance, a diversity of clutter variations could help to evaluate more fine-grained clutter impact. Also, more complex stock visualizations should be employed and tested with professional experts to test the ecological validity of our results.

Acknowledgements

This work was supported by the Swiss National Science Foundation SNSF Doc. CH. The authors also wish to thank Iva Barisic for her help with programming the experiment in Python and PsychoPy as well as Daniel Grosshans for their advice and help with creating experimental stock stimuli.

References

MATLAB 8.6. The MathWorks Inc., Natick, MA.
Alexander, A.L., Stelzer, E.M., Kim, S.H., Kaber, D.B., 2008. Bottom-up and top-down contributors to pilot perceptions of display clutter in advanced flight deck technologies. In: In Proceedings of the Human Factors and Ergonomics Society Annual Meeting. vol. 52. Sage Publications, September, pp. 1180–1184 No. 18.
Ang, K.K., Quek, C., 2006. Stock trading using RSPOP: a novel rough set-based neuro-fuzzy approach. IEEE Trans. Neural Netw. 17 (5), 1301–1315.
Beck, M.R., Lohrenz, M.C., Trafton, J.G., 2010. Measuring search efficiency in complex visual search tasks: global and local clutter. J. Exp. Psychol. Appl. 16 (3), 238.
Beck, M.R., Trenchard, M., van Lamsweerde, A., Goldstein, R.R., Lohrenz, M., 2012. Searching in clutter: visual attention strategies of expert pilots. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Sage Publications.
Boot, W.R., Ericsson, K.A., 2013. Expertise.
Bortz, J., Lienert, G.A., Boehnke, K., 2008. Verteilungsfreie Methoden in der Biostatistik. Springer-Verlag.
Chase, W.G., Simon, H.A., 1973a. Perception in chess. Cogn. Psychol. 4 (1), 55–81.
Chase, W.G., Simon, H.A., 1973b. The Mind's Eye in Chess.
de Groot, A.D., Gobet, F., Jongman, R.W., 1996. Perception and Memory in Chess: Studies in the Heuristics of the Professional Eye. *Van Gorcum & Co*.
Einhorn, H.J., 1972. Expert measurement and mechanical combination. Organ. Behav. Hum. Perform. 7 (1), 86–106.
Einhorn, H.J., 1974. Expert judgment: some necessary conditions and an example. J. Appl. Psychol. 59 (5), 562.
Goldberg, J.H., Kotval, X.P., 1999. Computer interface evaluation using eye movements: methods and constructs. Int. J. Ind. Ergon. 24 (6), 631–645.
Grahame, M., Laberge, J., Scialfa, C.T., 2004. Age differences in search of Web pages: the effects of link size, link number, and clutter. Hum. Factors 46 (3), 385–398.
Henderson, J.M., Chanceaux, M., Smith, T.J., 2009. The influence of clutter on real-world scene search: evidence from search efficiency and eye movements. J. Vis. 9 (1) 32-32.
Hens, T., Rieger, M.O., 2010. Financial Economics. Springer, " Berlin.
Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2011. Eye Tracking: A Comprehensive Guide to Methods and Measures. *OUP Oxford*.
Jarodzka, H., Scheiter, K., Gerjets, P., van Gog, T., 2010. In the eyes of the beholder: how experts and novices interpret dynamic stimuli. Learn. Instr. 20 (2), 146–154.
Kelton, A., Pennington, R., Tuttle, B., 2010. The effects of information presentation format on judgment and decision making: a review of the information systems research. J. Inf. Syst. 24 (2), 79–105.
Kendall, M.G., Smith, B.B., 1940. On the method of paired comparisons. Biometrika 31 (3/4), 324–345.
Kundel, H.L., La Follette Jr., P.S., 1972. Visual search patterns and experience with radiological images 1. Radiology 103 (3), 523–528.
Moacdieh, N., Sarter, N., 2015. "Display clutter a review of definitions and measurement techniques," *Human Factors*. The Journal of the Human Factors and Ergonomics Society 57 (1), 61–100.
Neider, M.B., Zelinsky, G.J., 2011. Cutting through the clutter: searching for targets in evolving complex scenes. J. Vis. 11 (14) 7-7.
Oskamp, S., 1965. Overconfidence in case-study judgments. J. Consult. Psychol. 29 (3), 261.
Palmer, J., 1994. Set-size effects in visual search: the effect of attention is independent of the stimulus for simple tasks. Vis. Res. 34, 1703–1721.
Pankok, C., Kaber, D., 2018. The effect of navigation display clutter on performance and attention allocation in presentation-and simulator-based driving experiments. Appl. Ergon. 69, 136–145.
Peirce, J., Gray, J., Halchenko, Y., Britton, D., Rokem, A., Strangman, G., 2011. PsychoPy–A Psychology Software in Python.
Reingold, E.M., Sheridan, H., 2011. Eye movements and visual expertise in chess and medicine. Oxford Handbook on Eye Movements 528–550.
Rosenholtz, R., Li, Y., Mansfield, J., Jin, Z., 2005. Feature congestion: a measure of display clutter. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, April, pp. 761–770.
Slovic, P., 2016. The Perception of Risk. " Routledge.
Spence, M.T., Brucks, M., 1997. The moderating effects of problem characteristics on experts' and novices' judgments. J. Mark. Res. 233–247.
Tang, F., Hess, T., Valacich, J., Sweeney, J., 2014. The effects of visualization and interactivity on calibration in financial decision-making. Behav. Res. Account. 26 (1), 25–58.
IBM SPSS Statistics for Macintosh, Version 22.0.
Tullis, T.S., 1983. The formatting of alphanumeric displays: a review and analysis. Hum. Factors 25 (6), 657–682.
Tversky, A., Sattath, S., Slovic, P., 1988. Contingent weighting in judgment and choice. Psychol. Rev. 95 (3), 371.
Weiss, D.J., Shanteau, J., Harries, P., 2006. People who judge people. J. Behav. Decis. Mak. 19 (5), 441–454.
Wickens, C.D., Hollands, J.G., Banbury, S., Parasuraman, R., 2013. Engineering Psychology & Human Performance. Psychology Press.
Ziegler, H., Nietzschmann, T., Keim, D., 2008. Visual analytics on the financial market: pixel-based analysis and comparison of long-term investments. In: International Conference on Information Visualisation, pp. 287–295 July.