Effective Visualization of Time-Critical Notifications in Virtual Reality

Uwe Gruenefeld[1], Marie-Christin Harre[2], Tim Claudius Stratmann[1], Andreas Lüdtke[2], Wilko Heuten[3]

Media Informatics and Multimedia Systems Group, University of Oldenburg[1]
Human Centered Design Group, OFFIS - Institute for Information Technology[2]
Interactive Systems Group, OFFIS - Institute for Information Technology[3]

uwe.gruenefeld@uol.de, harre@offis.de, tim.claudius.stratmann@uol.de, andreas.luedtke@offis.de, wilko.heuten@offis.de

Abstract

Virtual Reality (VR) devices empower users to be fully immersed into a virtual environment. In an enclosed environment such as this, notifications indicating risks (e.g., risk of injury when bumping into walls) must be perceived as quickly and correctly as possible. Different to findings from previous work using displays to investigate fast response times, it seems that immersion into a virtual environment, width of field of view, and use of near-eye displays observed through lenses may all be factors having considerable impact on the perception of time-critical notifications. Therefore, we studied the effectiveness of five different visualization types (color, shape, size, text, number) in two different setups (room-scale, fixed-position) with 20 participants in VR. Our study consisted of a part in which we tested one notification and a part with multiple notifications showing up at the same time. We measured reaction time, correctness and subjective user evaluation. Our results show that visualization types can be organized by a consistent effectiveness ranking for different numbers of notification elements presented. Further, we offer promising recommendations regarding which visualization type is best for showing time-critical notifications in future VR applications.

1 Introduction

In recent years, Virtual Reality (VR) technology has made considerable improvements (e.g., very high refresh rates, higher resolution). These advances have led to the production of VR headsets (e.g., Oculus Rift) that offer users a more immersive experience (Fittkau et al., 2015). Especially in enclosed environments, there exist time-critical notifications that must be perceived as quickly and correctly as possible by the user (e.g. notifications to warn about upcoming danger, such as walls in the user's play area, while using the room-scale experience). Time-critical notifications may also be utilized in less hazardous situations, such as if the door bell or telephone were to ring, but the user could no hear it because of their earplugs. Since we aim to visualize such notifications, our first step was to build upon existing knowledge for designing information visualizations.

In the desktop domain, guidelines for designing information visualizations exist (e.g., (Ware, 2004, 2010)), and some of these methods may be applicable to VR systems (Burns and Hajdukiewicz, 2013; Ostendorp, Feuerstack, et al., 2016). However, the main problem is that the available guidelines focus on multiple information elements being presented at the same time, while in many scenarios only one notification is being shown to the user. Furthermore, the available guidelines focus on desktop systems rather than VR systems. This is problematic because immersion into a virtual environment, wider field of view, and use of near-eye displays observed through lenses may have considerable impacts on the perception of time-critical notifications. In addition, prior works have shown that identification accuracy decreases rapidly across visual angle (Gutwin et al., 2017). To our knowledge, there is no prior work that has tested whether these guidelines are transferable to the VR environment for effective visualization of time-critical notifications.

In this paper, we investigate to what extent available guidelines can be applied to show time-critical notifications in Virtual Reality, and to what degree those guidelines must be specialized or adapted to fully capture the potential of VR systems. To examine this aspect, we conducted a user experiment with the VR headset Oculus Rift, in which users had to react to presented notifications as quickly and correctly as possible. Users were asked to state whether the presented notifications were critical or not critical. For this, we tested different visualization types based on prior guidelines (see Figure 1) in two different setups, with two different numbers of notifications presented at the same time (one, nine).

We propose the following contribution:

- A ranking of five visualization types (color, shape, size, text and number) with regard to reaction time, correctness and subjective feedback for encoding time-critical notifications in VR systems.

2 Related work

On an abstract level our study addresses the derivation of effective visualization (with regard to both time taken to perceive a certain piece of information and the correctness level of the perception). In this regard, we built upon knowledge from well-established work in the information visualization domain.

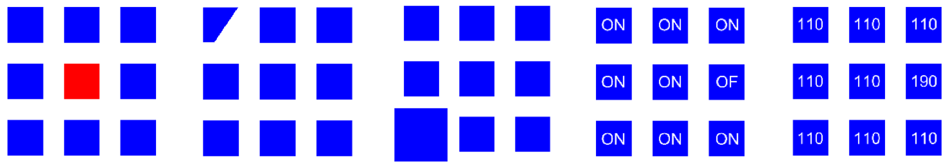
Cleveland and McGill (Cleveland and McGill, 1984) presented an efficiency ranking of visual attributes. In their work, the authors revealed which kinds of information visualizations (e.g. color hue, shape) are more efficiently perceived (with regard to time and accuracy). This ranking was later concertized by Mackinlay et al. (Mackinlay, 1986), who established different rankings for nominal value types, quantitative value types, and ordinal value types. Nevertheless, it has never been tested whether these rankings are still valuable in VR environments, in which the positions and backgrounds of the information visualizations are more variable and less controllable. Both work on the same abstraction level for defining different visualizations, e.g. comparing color hue vs. shape rather than comparing exact instances of those (e.g. red vs. blue). This abstraction level is based on the works of Carpendale (Carpendale, 2003), in which the author presented the consideration of visual variables as a basis for visualization and Bertin's Semiology of Graphics (Bertin, 2011). The visual variables used on this abstraction level are variables such as position, coordinates in space, size, orientation, and color hue, and are not concerned with individual properties in either dimension (e.g. red/blue). As Bertin's work is fundamental to the visualization area, we follow this abstraction level for defining visualizations in this paper.

Olshannikova et al. (Olshannikova et al., 2015) presented relevant challenges with regard to Big Data processing and analysis with a focus on visualization of Big Data and its integration with AR and VR. In this regard, the authors discuss visualization problems and claim that understanding issues related to human perception is critical for optimizing visualizations. For this, the authors present a future research agenda and data visualization challenges. In their opinion, simplicity in visualization has to be achieved, and psycho-physical studies can provide answers to questions regarding perception of visualizations in VR. With our paper, we present the results of such a psycho physical study to contribute to that research challenge. In the Augmented Reality domain, Huang et al. (Huang et al., 2013) investigated the effect of different visualizations for speed monitoring on the time to perceive the information. They experimented with different colors and shapes as well as different levels of background transparency for the head-up display. Huang et al. specifically focused on different visualizations for a speedometer. In contrast to this, we are working at a much higher level of abstraction, investigating which kind of information visualization (e.g. color hue, shape, volume, text) can be perceived most effectively.

For this purpose, we conducted an experiment to both test whether the rankings are still valuable and examine the effect of different VR setup (room-scale, stand-only) on perception of information. For this, we applied the adapted ranking of Ostendorp et al. (Ostendorp, Friedrichs, et al., 2016), in which the Mackinlay ranking was enhanced with knowledge about preattentive perception (Treisman, 1985; Healey et al., 1996; Gutwin et al., 2017). Preattentive perception states that some elements can be perceived in milliseconds by the human visual system, as these visual features produce a kind of "pop-out" effect when presented alongside less salient elements.

3 Experiment in Virtual Reality

For our experiment in VR, we derived three different visualization types (color, size and shape) from previous work (Cleveland and McGill, 1984; Mackinlay, 1986; Ostendorp, Friedrichs, et al., 2016) and combined them with two additional visualization types (word and number) that are frequently used for displaying time-critical notifications (see Figure 1). We included text



(a) Color (b) Shape (c) Size (d) Word (e) Number

Figure 1: Multiple notifications with different visualization types for one time-critical notification. Best seen in color.

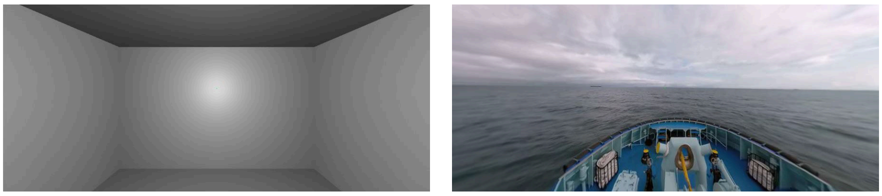
(word and number) to specifically analyze whether this would negatively impact performance compared to visual presentation forms (color, size and shape). For our experiment, we asked participants to rate displayed notifications as critical or non-critical, as we presented one or nine notifications each, depending on the part of the experiment. Further, we tested the notifications in two different setups (see Figure 2). We counterbalanced the two parts (one vs. nine notifications) over all participants and counterbalanced the two setups (fixed-position vs. room-scale) within each part. We used a blue square to represent non-critical notifications. To evaluate the different visualization types we randomly (equal amount) changed one or no parameter of a randomly selected notification to make it critical in each trial (e.g., color was changed to red to make the displayed notification critical).

We decided to use red to encode a critical notification due to its association with danger, while we used blue for a non-critical state due to its association with calmness (Maehr et al., 2008). For the shape, we decided to use a simple symmetric form to indicate non-criticality and a more complex, asymmetric form for criticality. We based this decision on the fact that humans prefer symmetry and associate it with positive aspects (Tyler, 2003). For the size, we enlarged the square to indicate a critical state. In the case of text and number, it was important for us to choose words or numbers with similar lengths. Otherwise, the participants would have been able to estimate the state of the system by perceiving length, which is another visualization type according to the previously mentioned prior works. For non-critical information, we chose the number 110 and the word "on", while we chose a higher value (190) and the word "of" for criticality. These values can, for instance, indicate the state of an alarm system (on/off) or critical temperature levels (110/190). In this case, it is critical if the alarm system is turned off or the computer's graphics card reaches a temperature around 190° Celsius.

3.1 Study design

Our experiment was conducted as a within-subjects controlled laboratory study. Our study had three independent variables: number of notifications presented at the same time (one vs. nine), VR setup (fixed-position vs. room-scale) (see Figure 2), and visualization with five levels (color vs. shape vs. size vs. text vs. number) (see Figure 1). All notifications were shown in foveal or para-foveal vision (Kalat, 2015). Given the foregoing, we had 2 (numbers of notifications presented at the same time) x 2 (VR setups) x 5 (visualization types) x 6 (iterations)[1], resulting

[1] We derived this value from pilot testing in which it worked best with regard to study duration.



(a) Room-scale (3D Environment). *(b) Fixed-position (360° video).*

Figure 2: Different VR setups. Best seen in color.

in 120 runs per participant. For conducting the experiment, we used the Oculus Rift - a head-mounted VR glass which is fully immersive.

Our hypotheses were:

H_1 If one notification is presented, we hypothesize that the visualization type 'size' performs worse with respect to correctness compared to 'color', 'shape', 'number' and 'text'.

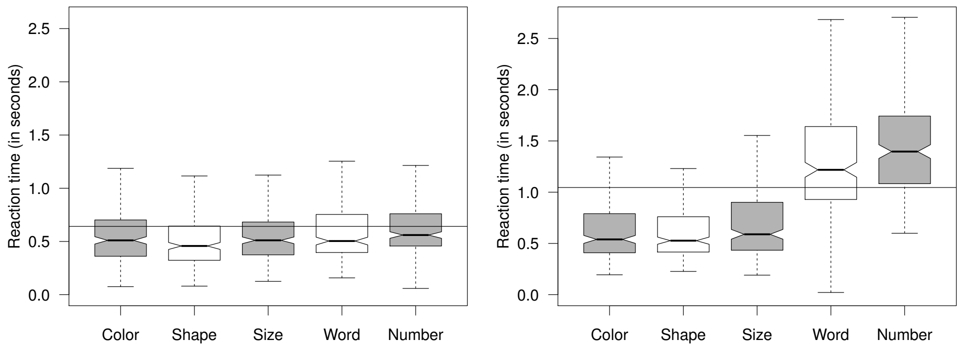
H_2 Based on prior work, we hypothesize that color performs best with regard to measured reaction time independent of the number of presented notification and the VR setup.

H_3 The VR setup 'room-scale' has a negative impact on the measured reaction times.

3.2 Procedure

The experiment took place in an empty office room with participants sitting on a chair (for fixed-position) or moving around while standing (for room-scale). For the fixed-position scenario, participants were sitting on a chair both because this is the typical use-case of this VR setup and because it prevented them from moving around. For the room-scale experience, we used a virtual copy of the office room the participants were placed in, in order to have the physical borders within the VR setup. We divided the study into two parts, investigating one notification in one part and nine notifications presented at the same time in the other part. For each part, participants had a short training phase with a black background to become familiar with the different visualizations. After the training phase, the participants were asked to estimate the system state as quickly and correctly as possible in each run. The different conditions of each part were presented in a counter-balanced design.

The task of the participants in each run was to state whether one of the presented notifications was critical or not as quickly and correctly as possible. For this, each participant worked with the five different visualization types for showing the criticality of a notification. In each run, the participant had to signal readiness by pressing a trigger. After this, one or nine notifications appeared and the participant had to press the trigger again to indicate he/she had fully perceived the situation. After that the visualization disappeared. Thereby, we were able to measure the time in ms needed to perceive the situation with the given visualization type. After the visualization disappeared, the participant had to state if one of the notifications shown before was critical or not critical. With the answers given, we were able to measure the correctness of the percept.



(a) Single notification. *(b) Multiple notifications.*

Figure 3: Boxplot of mean reaction time (line indicates overall mean reaction time).

After each part, the participants were asked to fill in a questionnaire, rating the different visualizations as Likert-scale items. Overall, each participant took approximately 40 minutes to finish the experiment.

3.3 Participants

In the experiment, we had 20 participants (7 female) aged between 16 and 61 years (M=29.65, SD=11.6). None of them suffered from color vision impairment. All had normal or corrected to normal vision.

3.4 Results (single notification)

Reaction time We consider the effects of the two factors (Visualization, Setup) on mean reaction time. The mean reaction times for the visualizations are: Color=0.602s, Shape=0.637s, Size=0.704s, Word=0.633s, Number=0.634s. The reaction times are compared in Figure 3a.

A Shapiro-Wilk-Test showed that our data is not normally distributed ($p < 0.001$), and thereafter we ran a Friedman test that revealed a significant effect of visualization on reaction time ($\chi^2(4)$=68.12, $p < 0.001$, N=20). A post-hoc test using Wilcoxon Signed-rank with Bonferroni correction showed significant differences between the visualizations (see Table 1a).

To investigate if there is a significant effect of VR setup on reaction time, we directly performed a Wilcoxon Signed-rank with Bonferroni correction. The mean reaction times for the two setups are: Room-scale=0.622s and Fixed-position=662s. Here, we found no significant differences (W = 90486, Z = 0.079, $p = 0.937$, ϕ = 0.00).

Correctness We consider the effects of the two factors (Visualization, Setup) on correctness (where correctness means how accuratly user were able to specify the position of critical notifications). The total numbers of correctly perceived critical notifications are: Color (231/240 = 96.3% accuracy), Shape (230/240 = 95.8% accuracy), Size (217/240 = 90.4% accuracy), Word

Comparison	P-value	ϕ-value
Color vs. Shape	0.002**	0.20
Color vs. Size	0.121	0.10
Color vs. Word	0.029*	0.14
Color vs. Number	<0.001***	0.24
Shape vs. Size	<0.001***	0.30
Shape vs. Word	<0.001***	0.30
Shape vs. Number	<0.001***	0.38
Size vs. Word	0.540	0.04
Size vs. Number	0.021*	0.15
Word vs. Number	0.108	0.10

(a) Single notification.

Comparison	P-value	ϕ-value
Color vs. Shape	0.475	0.05
Color vs. Size	0.066	0.12
Color vs. Word	<0.001***	0.78
Color vs. Number	<0.001***	0.85
Shape vs. Size	0.015*	0.16
Shape vs. Word	<0.001***	0.81
Shape vs. Number	<0.001***	0.85
Size vs. Word	<0.001***	0.78
Size vs. Number	<0.001***	0.85
Word vs. Number	<0.001***	0.36

(b) Multiple notifications.

Table 1: Pairwise comparisons of visualization techniques.

(217/240 = 90.4% accuracy) and Number (223/240 = 92.1% accuracy).
The order is: **Color $>$ Shape $>$ Number $>$ Word, Size.**

Questionnaire At the end of the task, we asked participants to rate all visualizations with 6-point Likert items. Participants stated they very much liked color (Md=6, IQR=0), shape (Md=5.5, IQR=1.25), and size (Md=4.5, IQR=1), while they disliked word (Md=2.5, IQR=1) and number (Md=2, IQR=2). Overall, 16 participants preferred color, three preferred shape, and one preferred number.

3.5 Results (multiple notifications)

Reaction time Again, we consider the effects of the two factors (Visualization, Setup) on mean reaction time. The mean reaction times for the visualizations are: Color=0.723s, Shape=0.729s, Size=0.744s, Word=1.407s, Number=1.625s. The reaction times are compared in Figure 3b.

A Shapiro-Wilk-Test showed that our data is not normally distributed ($p < 0.001$), and thereafter we ran a Friedman test, which revealed a significant effect of visualization on reaction time ($\chi^2(4)$=574.45, $p < 0.001$, N=20). A post-hoc test using Wilcoxon Signed-rank with Bonferroni correction showed significant differences between the visualizations (see Table 1b).

To investigate if there is a significant effect of VR setup on reaction time, we directly performed a Wilcoxon Signed-rank with Bonferroni correction. The mean reaction times for the two setups are: Room-scale=1.043s and Fixed-position=1.049s. Here, we found no significant differences (W = 83157, Z = -1.582, $p = 0.114$, ϕ = 0.05).

Correctness We consider the effects of the two factors (Visualization, Setup) on correctness (where correctness means how accuratly users were able to specify the position of critical notifications). The total numbers of correctly perceived critical notifications are: Color (221/240 = 92.1% accuracy), Shape (216/240 = 90.0% accuracy), Size (216/240 = 90.0% accuracy), Word (210/240 = 87.5% accuracy) and Number (210/240 = 87.5% accuracy).
The order is: **Color $>$ Shape, Size $>$ Word, Number.**

Questionnaire At the end of the task, we asked participants to rate all visualizations with 6-point Likert items. Participants stated they very much liked color (Md=6, IQR=0), shape (Md=5, IQR=0), and size (Md=5, IQR=1), while they disliked word (Md=3, IQR=1.25) and number (Md=2, IQR=1.25). Overall, 16 participants preferred color while two preferred shape, and two preferred size.

3.6 Discussion

Single vs. multiple notifications presented at the same time In our study, we compared the presentation of time-critical notifications with one or nine presented at the same time. We did this to test the performance of different visualization types independent of neighboring notifications. In our hypothesis H_1 we assumed that size would perform worse compared to other visualization techniques when only one notification is presented. With regard to mean reaction time, we found this to be the case. However, the median did not significantly differ. Therefore, we cannot accept H_1, but the high mean reaction time reveals that participants were insecure about this visualization type. Further, the correctness for one notification is lowest for the visualization type size. Therefore, we recommend to not use size for encoding time-critical notifications, especially when only one notification is presented at a time.

Comparison to existing guidelines The results reveal the transferability of guidelines and recommendations from the desktop domain (Cleveland and McGill, 1984; Mackinlay, 1986; Ostendorp, Friedrichs, et al., 2016) to some extent. According to works in the desktop domain, the effectiveness ranking for estimating whether a value is critical is: color > shape > size > text (Ostendorp, Friedrichs, et al., 2016). Based on this prior knowledge, we built our hypothesis H_2 that color performs best with regard to measured reaction time. Within the experiment we were able to show that this is only true with regard to correctness and subjective ranking. However, with regard to median reaction time, the visualization type shape was slightly faster than color in all conditions. For multiple notifications presented to the user, we suggest the following effectiveness ranking in VR:

Multiple notifications
color > shape, size > word, number

For time-critical notifications in VR that are presented without other notifications, we suggest a slightly different effectiveness ranking:

Single notification
color > shape > number, word, size

Since the number of notifications is not controllable for most systems, we recommend using the ranking for single notifications.

Influence of VR setup We wanted to investigate to what extent the VR setup - which is less controllable for most VR applications (the user decides) - affects the effectiveness ranking of the different visualization of notifications. In hypothesis H_3 we expected the VR setup 'Room-scale' to have a negative impact on the measured reaction time. We stated this because we assumed a moving player would have to deal with more distraction, which would lead to worse reaction times. However, our results showed no significant differences in reaction time and therefore, we cannot accept our hypothesis H_3.

Transferability to head-mounted displays Although we placed the experiment in the VR domain, we assume that an experiment in the AR domain would lead to similar results: "Augmented reality can be considered a type of virtual reality. Rather than experiencing physical reality, one is placed in another reality that includes the physical along with the virtual." (Sherman and Craig, 2003). Nevertheless, a future experiment situated in the AR domain (e.g. using a Hololens or Meta2) may strengthen the certainty of such guidelines for AR. In the future, it may also be interesting to investigate wether the results change if different colors, shapes, sizes, words, and numbers are used.

Limitations In our study we limited ourselves to one representation per visualization type. While this was sufficient as a first step towards an efficiency ranking in VR, further work is needed to verify our results. The same argument applies to the tested VR setups. One per setup may not be sufficient.

4 Conclusion

In this paper, we presented the results of a perception study situated in the VR domain. As VR experiences become increasingly immersive, VR developers need guidelines to derive effective visualizations of time-critical notifications. Effective in this sense means that the information visualized can be perceived as quickly and as correctly as possible and is strongly favored by the users. In our experiment, we tested different visualization types (color, shape, size, text, number) for displaying time-critical notifications in two VR setups (fixed-position, room-scale) with two different numbers of notifications presented at the same time (one vs. nine). We were able to set up an effectiveness ranking for the presentation of single or multiple notifications. This ranking can be used by VR developers to make decisions about how to most effectively present visualizations of time-critical notifications in future applications.

References

Bertin, J. (2011). *Semiology of graphics: Diagrams, networks, maps*. ESRI Press.

Burns, C., & Hajdukiewicz, J. (2013). *Ecological interface design*. CRC Press.

Carpendale, M. S. T. (2003). *Considering visual variables as a basis for information visualisation*. University of Calgary. Calgary, AB.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), pp. 531-554.

Fittkau, F., Krause, A., & Hasselbring, W. (2015). Exploring software cities in virtual reality. In *2015 ieee 3rd working conference on software visualization (vissoft)* (pp. 130–134). doi:10.1109/VISSOFT.2015.7332423

Gutwin, C., Cockburn, A., & Coveney, A. (2017). Peripheral popout: The influence of visual angle and stimulus intensity on popout effects. In *Proceedings of the 2017 CHI conference on human factors in computing systems, denver, co, usa, may 06-11, 2017.* (pp. 208–219). doi:10.1145/3025453.3025984

Healey, C. G., Booth, K. S., & Enns, J. T. (1996). High-speed visual estimation using preattentive processing. *ACM Trans. Comput.-Hum. Interact. 3*(2), 107–135.

Huang, C., Chao, C.-W., Tsai, T., & M.-H., H. (2013). The effects of interface design for head-up display on driver behaviour. *Life Science Journal*, *2*(10).

Kalat, J. (2015). *Biological psychology*. Nelson Education.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics TOG*, *5*(2), 110–141.

Maehr, M., Karabenick, S., & Urdan, T. (2008). *Social psychological perspectives*. Advances in motivation and achievement. Emerald [Group.

Olshannikova, E., Ometov, A., Koucheryavy, Y., & Olsson, T. (2015). Visualizing big data with augmented and virtual reality: Challenges and research agenda. *Journal of Big Data*, *2*(1), 22.

Ostendorp, M.-C., Feuerstack, S., Friedrichs, T., & Lüdtke, A. (2016). Engineering automotive hmis that are optimized for correct and fast perception. In *Proceedings of the 8th acm sigchi symposium on engineering interactive computing systems* (pp. 293–298). EICS '16. Brussels, Belgium: ACM.

Ostendorp, M.-C., Friedrichs, T., & Lüdtke, A. (2016). Supporting supervisory control of safety-critical systems with psychologically well-founded information visualizations. In *Proceedings of the 9th nordic conference on human-computer interaction*. NordiCHI '16. Gothenburg, Sweden: ACM.

Sherman, W. R., & Craig, A. B. (2003). *Understanding virtual reality: Interface, application, and design*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics and Image Processing*, *31*(2), 156–177.

Tyler, C. (2003). *Human symmetry perception and its computational analysis*. Taylor & Francis.

Ware, C. (2004). *Information visualization: Perception for design*. Interactive Technologies. Elsevier Science.

Ware, C. (2010). *Visual thinking: For design*. Morgan Kaufmann series in interactive technologies. Elsevier Science.