Journal of Risk Research
Vol. 14, No. 1, January 2011, 47–61

How do people perceive graphical risk communication? The role of subjective numeracy

Rebecca Hess*, Vivianne H.M. Visschers, Michael Siegrist and Carmen Keller

ETH Zurich, Institute for Environmental Decisions (IED), Consumer Behavior, Universitaetstrasse 22, Zurich 8092, Switzerland

(*Received 9 July 2009; final version received 22 April 2010*)

This study aims to evaluate directly how a graphical risk ladder is perceived and how this perception is related to people's subjective numeracy. Gaze durations and frequencies were used to examine visual attention. Participants ($N = 47$) appeared to focus on the target risk information, whereas referential information was less attended. Subjective numeracy was negatively correlated with total watching time and the absolute number of gaze events. Results suggest that participants with low subjective numeracy have more difficulty in comprehending the graph, and that they process the graphical information less efficiently than the participants with high subjective numeracy. In addition, the position of referential risks on risk ladders could influence people's risk perception. Based on these findings, we provide some implications for the design of risk communication graphs and for the use of graphs in informing persons with low subjective numeracy about risks.

Keywords: graphical risk communication; visual attention; eye-tracking; subjective numeracy; Paling perspective scale

Introduction

In today's Western society, people are confronted with many different everyday risks. These can range from potentially dangerous substances in the household to medical risks. In the medical domain, people have to decide whether to undergo a screening test for a certain disease or which medical treatment they should choose. Enabling patients to make informed decisions involves making such decisions based on well-understood information and corresponding to their own values (Marteau, Dormandy, and Michie 2001). This implies that doctors and/or medical counsellors have to fully inform patients about the risks of procedures and the meanings of test results. How risks can be successfully communicated is still an open question, however (Siegrist, Cousin, and Keller 2008). In this paper, therefore, we examine people's visual attention to a risk communication graph that can be applied in medical decision-making.

Graphical risk communication formats

Previous research has shown that people have substantial difficulties in understanding probability information (see Gigerenzer and Edwards 2003; Lipkus 2007; Visschers

*Corresponding author. Email: hessr@ethz.ch

ISSN 1366-9877 print/ISSN 1466-4461 online

DOI: 10.1080/13669877.2010.488745
http://www.informaworld.com

et al. 2009 for reviews of this literature). Some authors, therefore, suggest using graphical representations of numerical information for purposes of medical risk communication (Edwards, Elwyn, and Mulley 2002; Stallings and Paling 2001; Visschers et al. 2009). There are studies within the medical field indicating that certain graphical risk communication formats are indeed better understood than numerical information (Waters et al. 2006; Woloshin et al. 2000). However, this effect has not always been observed and seems to depend on the type of graph and the task at hand (Feldman-Stewart et al. 2000; Siegrist, Orlow, and Keller 2008; Weinstein, Sandman, and Hallman 1994).

Basically, there are two ways to depict the probability of a risk graphically. Either the risk is presented on its own (e.g. with an array of pictograms, see Siegrist, Orlow, and Keller 2008) or is related to other risks to put it into a broader context. The latter type of risk depiction might be especially helpful to convey risk information, as people intuitively seem to draw on analogies and risk comparisons to attach meaning to different risks (Bostrom 2008; see also Visschers et al. 2007). One special type of graph used to depict a target risk in relation to others consists of the so-called risk ladders. In these graphs, several risks are presented in an ascending order according to the level of risk to show where the target risk is located. Such risk ladders are often investigated in the field of environmental risks (Connelly and Knuth 1998; Sandman, Weinstein, and Miller 1994). For medical contexts, a special type of risk ladder called the Paling perspective scale (PPS) was designed to help doctors communicate risks to patients (Paling 2003; Stallings and Paling 2001). This graph consists of a panel that is subdivided into sections according to a logarithmic scale. The target risk and several reference risks can be depicted in this panel without explicitly referring to numeric values. Thus, the PPS offers two advantages: first, a target risk can be related to reference risks, and second, due to the logarithmic scale, risks of very different probabilities can be depicted and compared.

To our knowledge, no previous studies have explored whether these two characteristics of the PPS really do play a role in how people look at a risk that is presented in this graph. Therefore, the first aim of our study is to investigate whether people pay attention to the reference risks while looking at the PPS and how they perceive the logarithmic scale.

To measure the perception of the graph as directly as possible, we used an eye-tracker. This instrument follows a person's gaze and records eye movement data. Based on the presumption that the point of visual attention is also the focus of current cognitive processing (Just and Carpenter 1976), eye movement data have been used to investigate the processing of visual displays such as print advertisements (Rayner, Miller, and Rotello 2008), information graphics in newspapers (Holsanova, Holmberg, and Holmqvist 2009) and various types of graphs (Carpenter and Shah 1998; Korner 2004; Ratwani, Trafton, and Boehm-Davis 2008).

Numeracy

People's numeracy skills, which describe the ability to handle numbers, seem to affect the understanding of risk information (Apter et al. 2008; Nelson et al. 2008; Peters et al. 2007; Reyna and Brainerd 2007). Low numeracy skills appear to be associated with greater difficulties in understanding risk information and being more prone to influences by framing effects (Peters and Levin 2008; Peters et al. 2006; Schwartz et al. 1997). Graphical risk presentations have been recommended, therefore, especially for persons with low numeracy skills (Apter et al. 2008; Nelson et al. 2008). Only a few studies, however, have examined whether persons with low numeracy skills are better able to understand risk depicted graphically compared to risk depicted numerically. A recent study found that persons scoring low on a numeracy measure were unable to differentiate between high- and low-probability risks when presented in the PPS, whereas persons scoring high on this numeracy measure were able to do so (Keller and Siegrist 2009). As visual attention is a prerequisite for understanding a graph (see, e.g., the framework presented by Grunert and Wills 2007), the second aim of our study is to use eye movement data to investigate whether there is a relationship between numeracy skills and the visual perception of the PPS.

There are two mechanisms that can explain the relationship between understanding information depicted in a risk communication graph and numeracy. First, it is possible that graph reading, overall, is more difficult and complex for persons with low numeracy skills, as graph reading is part of numeracy itself. Studies about the comprehension of graphical displays in non-medical fields have shown that more complex graphs provoke more and longer fixations than simpler graphs due to more extensive cognitive processing involved in their interpretation (Carpenter and Shah 1998; Ratwani, Trafton, and Boehm-Davis 2008; Renshaw et al. 2004). Therefore, we expect a negative relationship between the participants' numeracy and the total time needed to look at the risk communication graph as well as the total number of gazes (*Hypothesis* 1).

Secondly, there is some evidence indicating that persons with high numeracy skills integrate more information in the decision-making process than persons with low numeracy skills (Peters and Levin 2008). Applied to the visual perception of the PPS, this could imply that persons with high numeracy skills extract more information from the graph than persons with low numeracy skills and, thus understand it better. This mechanism, in turn, can be described as having higher efficiency with which they process the PPS. Goldberg and Kotval (1999) suggest that, in search tasks, higher ratios of fixations on target regions of a graph per total number of fixations are indicative of search efficiency. We therefore expect a positive relationship between the participants' numeracy and the number of elements of the PPS they consider relative to the total number of gazes (*Hypothesis* 2).

Rationale of the study

In sum, this study has two general purposes. First, the general aim is to find out how people look at the PPS, in order to find out whether they actually take into account the special characteristics of the graph that are designed to foster comprehension of the depicted risk. Second, we want to explore the association between numeracy and visual attention to the PPS.

Methods

Participants

Participants (N = 52) were invited from a pool of persons who had taken part in a recent study and who had agreed to participate in the present study. They were chosen based on their numeracy value, which had been measured in the earlier study, so that a wide range of numeracy values could be represented in our sample. An independent person selected the participants based on their numeracy scores. Another person, who was blind to the potential participants' numeracy value, contacted them by telephone. Five datasets could not be analysed due to bad eye-tracking data quality (e.g. because of more than 20% missing data or strong offsets of eye movement measurements) and were excluded from further analysis. Thus, the sample used in the present study consisted of 47 participants: 15 females (32%) and 32 males (68%). The mean age was 51.5 years (SD = 13.4), and the education level ranged from lower-secondary school (8%, *n* = 4), upper-secondary vocational school (32%, *n* = 15) and upper-secondary school (11%, *n* = 5) to university/technical university (49%, *n* = 23). The participants received CHF 100 (approximately EUR 70) for their participation.

Materials

Informed consent

All the participants were asked to sign an informed consent form on which they indicated that the function of the eye-tracker had been explained to them, that they allowed us to analyse their data, which would be done anonymously, and that they could stop their participation in this study at any time.

Numeracy

Numeracy was measured with the subjective numeracy scale (SNS, Fagerlin et al. 2007; Zikmund-Fisher et al. 2007). This instrument consisted of an eight-item questionnaire about one's perceived numerical ability in several contexts (e.g. working with fractions, percentages or calculating a 15% tip). The average of all eight items resulted in the numeracy score. We chose the SNS because subjective numeracy in a broader sense could be interpreted as a special type of self-efficacy (Bandura 1997) with regard to the handling of numbers. In some studies, self-efficacy measures have been shown to be relevant to information-seeking behaviours (Bass et al. 2006; Hong 2006; Satia, Galanko, and Neuhouser 2005). As looking for information in a graph like the PPS can be understood as information-seeking behaviour, the SNS seems to be the best instrument for this study because it provides information about both the ability and the self-efficacy aspects of numeracy.

Scenario and risk information

We used a risk scenario about prenatal screening, which we presented as slides on a 15.4″ monitor of a laptop computer. The hypothetical story on the first slide was about a 35-year-old woman ('Sandra'), pregnant with her first child, who had her blood tested for the probability of her unborn child having Down syndrome. The result of this test consisted of the probability that the child would be affected. This probability (1 in 112) was depicted graphically in a modified version of the PPS on the second slide (Figure 1; modified after Keller and Siegrist (2009), Stallings and Paling (2001)). We presented two reference risks in the graph: Down syndrome in women of Sandra's age (1 in 365) as well as Down syndrome in the general population (1 in 680). The actual numbers were not explicitly mentioned in the graph. They were only graphically depicted as data points and had to be derived by the participants using the logarithmic scale.

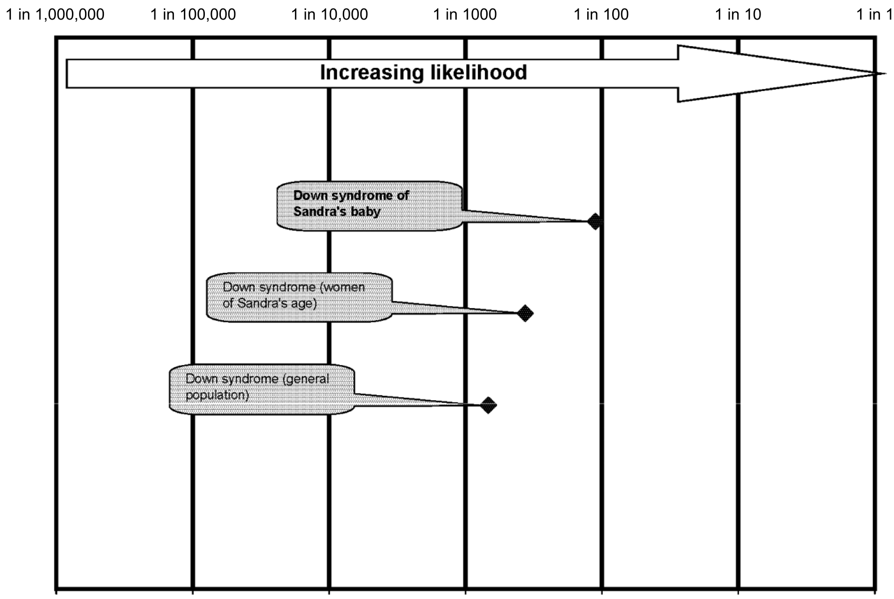

Figure 1. Modified Paling perspective scale; modified after Keller and Siegrist (2009), Stallings and Paling (2001).

Above and below the graph, we presented textual information. Above the graph, there was a short text instructing the participants to keep in mind the level of the depicted risk and the emotions caused by the depicted risk. Below the graph, there was information about how to proceed to the next slide.

Eye-tracker

We used a 50-Hz SMI HED4 eye-tracker to measure people's visual attention. This is a dual-Purkinje dark pupil eye-tracker system consisting of two cameras on a helmet, which the participant wears on the head. The scene camera records the scene a person is looking at (25 pictures per second). The eye camera films the eye itself (50 pictures per second). The data from both cameras are transmitted to a computer, which measures the centre of the pupil as well as Purkinje light reflections. Special software calculates the gaze direction and location based on these parameters. These are then presented in the scene video, whereby the gaze is depicted as a small circle in the scene.

Procedure

The study took place in a laboratory with constant light. First of all, the eye-tracker was explained, and all the participants provided written informed consent. Then, the experimenter installed and calibrated the eye-tracker. Calibration was checked and, if necessary, repeated.

The participants were then shown an exercise scenario about the risk of malaria infection while travelling (depicted in a pie chart). This scenario enabled the participants to get used to the eye-tracker and to the type of task, and was not analysed. The eye-tracker was re-calibrated, and then the Down syndrome scenario described above was shown. The eye-tracker recorded the eye movements for as long as respondents perceived the graph; there was no time limit. The participants then assessed their emotional response to the depicted risk (on a scale from 0, very negative, to 100, very positive) and the level of the depicted probability (on a scale from 1, very low, to 6, very high). These questions and the instructions focusing on the presented risk and emotions were intended to stimulate the respondents' evaluation of the presented risk. At the end of the session, the participants provided the demographic information (gender, age, education), and the subjective numeracy was measured with the SNS (Fagerlin et al. 2007; Zikmund-Fisher et al. 2007).

Data analysis

This procedure resulted in 47 videos of the PPS with overlaid gazes (in the form of a moving red circle called the 'gaze cursor'), which were coded by using video-coding software (Interact, version 8.4.4, Mangold International). Prior to the analysis, the PPS was subdivided into 22 so-called areas of interest (AOI), which were defined as regions in and around the graph that one could look at while watching and interpreting the PPS (see Figure 2).[1] We assigned a code to each AOI.

A code was assigned to an AOI if the gaze stayed on an AOI for at least three video pictures (120 ms). This time threshold was chosen because it approximately equalled the minimum time span of the fixation (see Salvucci and Anderson 2001). Each such time period of at least 120 ms within one AOI was called a 'gaze event'. From these gaze events, we calculated two types of gaze variables: cumulative duration of gazes in each AOI (in seconds) and frequency of gaze events in each AOI. Thus, the gaze duration variables provided information about how long participants looked at each AOI while they processed the graph, irrespective of how often they looked at this

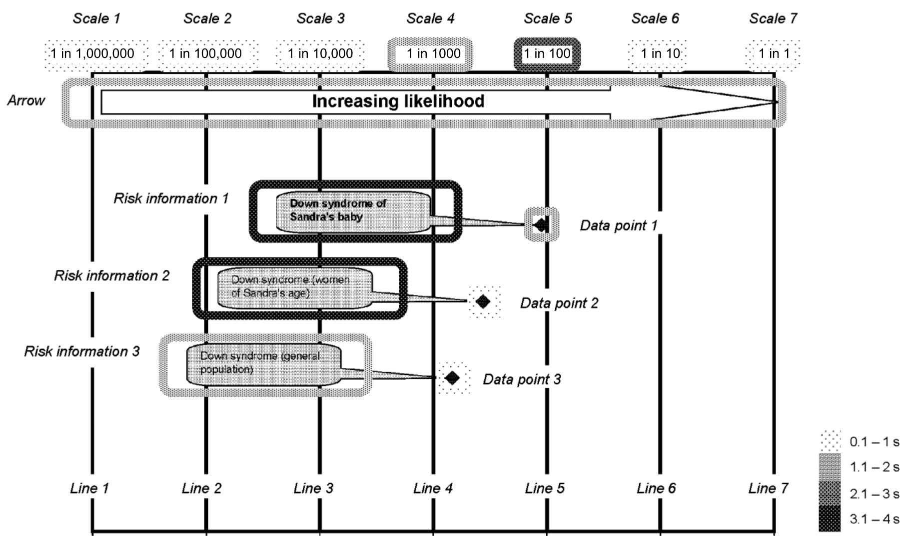

Figure 2. Duration of gazes in different AOIs (median values; the darker the pattern, the longer the gazes; all AOIs that are looked at for less than 0.1 seconds are not marked; names of AOIs in italics).

AOI. Gaze frequency variables, on the other hand, showed how many times respondents looked at the AOIs, irrespective of how long they did this.

To track gaze patterns between AOIs in the PPS, we also analysed certain sequences that were necessary to understand the graph. These crucial sequences consisted of all three risk information AOIs (indicating people's understanding of the nature of the risks in question), the corresponding data points (understanding of the position of the risks in question), the Scales 4 and 5 (understanding of the probability of the risks in question), and the Scales 3 and 6 (understanding of the logarithmic scale). The sequences were calculated as pairs of these AOIs that were looked at one after the other (in terms of frequencies of each sequence). For example, when someone looked first at Data Point 1 and then at Scale 5, this was coded as sequence Data Point 1/Scale 5 with a frequency of 1.

Our analyses showed that all line AOIs as well as the AOI 'scale undefined' were looked at very rarely. Therefore, we did not report any results related to these AOIs. The texts above and below the graph were coded as text variables. Missing information was coded when the gaze cursor was not visible for more than eight video pictures (320 ms, not in order to overrate blinks that should be shorter than this duration, see Caffier, Erdmann, and Ullsperger 2003), or if the gaze cursor was visible but absolutely not interpretable for at least three video pictures (120 ms, e.g. due to strong head movements). The videos contained, on average, 6% missing data (SD = 4%).

Fourteen videos were rated by an independent second rater. Based on the codes of each rater, we calculated gaze durations, gaze frequencies and sequences of these 14 participants for all AOIs (except lines and scale undefined), as well as for total variables (total duration, total number of gaze events, total number of AOIs considered). Pearson correlations were computed between the two ratings of these variables to assess the inter-rater reliability. This resulted in the very high average correlation coefficients for the duration variables ($M_r = 0.97$), the frequency variables ($M_r = 0.93$) and the sequence variables ($M_r = 0.85$). Both raters were blind to the participants' numeracy values in order to avoid over- or under-rating of certain AOIs.

Data analysis was conducted with SPSS (version 16.0.1, SPSS Inc.) and SYSTAT (version 12.00.08, SYSTAT Software Inc.). As gaze variables partially showed outliers and were not normally distributed, medians and 95% confidence intervals (CIs) of the medians were reported. These CIs were calculated with bootstrapping (Efron and Tibshirani 1993; Mooney and Duval 1993). With this procedure, the sample was considered as a population from which many random samples were drawn (with replacement). The median was then calculated for all of these hypothetical samples. The CI of the observed median could, in turn, be derived from these bootstrapped medians. We chose 1000 samples of 47 units each for this procedure. The associations between the different gaze variables and subjective numeracy were examined with Spearman rank correlations.

Results

Overall description of visual attention to the PPS

As we were interested in the perception of the graph and not in the processing of text, all subsequent analyses referred to the perception of the PPS without instructional text. The participants looked at the PPS for a median duration of 28.4 s (CI (95%) = 22.2–30.5). The median number of gaze events for looking at the PPS was 37 (CI (95%) = 31–43). The participants looked at 14 (median; CI (95%) = 13–15) of the 22

AOIs within the PPS. On average, the participants assessed the target risk depicted in the PPS (1 in 112) as rather high (median = 3.9, SD = 1.1; scale 1–6) and associated it with negative feelings (median = 21.2, SD = 12.9; scale 0–100).

Figure 2 shows how long the participants look at the various parts of the graph. Medians and CIs for gaze duration and gaze frequency are shown in Table 1. Results showed that the participants spent the most time looking at Risk Informations 1 and 2, whereas they spent significantly less time on Risk Information 3. Most participants looked at all three data points. In other words, they looked at the information that was needed to make risk comparisons. However, there were differences within the gaze durations of the three points. Data Point 1 was looked at for the longest, followed by Data Point 2 and Data Point 3. Results further suggested that the participants focused more on Scales 4 and 5 than on the other scale points.

Overall results of gaze duration and gaze frequency were very similar. Both suggested that the participants looked longer and more often at the parts of the graph that were important for understanding the target risk information, and that they looked for less time and less frequently at the areas that were less important. Furthermore, there seemed to be a decrease in attention from Risk Information 1/Data Point 1 (the target risk to be communicated), as well as Risk Information 2/Data Point 2 (the first reference information), to Risk Information 3/Data Point 3 (the second reference information).

In addition to the duration and the frequency of gazes in different AOIs, the frequency of gaze sequences between certain areas can shed light on the visual attention to the graph. Figure 3 shows the distribution of such sequences that are important for the understanding of the graph and appear with at least a median frequency of 1. The most frequent sequences are the ones between Risk Information 1 and Data Point 1, as well as between Data Point 1 and Scale 5. In addition, there seems to be a concentration on the target risk to be communicated in the first place. This is observable in

Table 1. Medians and 95% confidence intervals of gaze durations (cumulated, in seconds) and of gaze frequencies, $N = 47$.

	Gaze duration		Gaze frequency	
Variable	Median	CI (95%)	Median	CI (95%)
Scale 1	0.8	0.5–1.4	1	1–2
Scale 2	0.2	0.2–0.5	1	1–1
Scale 3	0.4	0.3–0.7	1	1–2
Scale 4	1.4	1.1–2.3	3	3–4
Scale 5	2.2	1.5–2.8	4	3–5
Scale 6	0.4	0.3–0.7	1	1–2
Scale 7	0.4	0.3–0.6	1	1–1
Arrow	1.5	1.1–1.9	2	2–3
Risk Information 1	3.3	2.7–4.2	5	4–5
Risk Information 2	3.2	2.7–4.7	2	2–3
Risk Information 3	2.0	1.4–2.6	2	1–2
Data Point 1	1.2	0.8–1.7	4	4–5
Data Point 2	0.6	0.6–0.8	2	2–3
Data Point 3	0.4	0.4–0.6	2	1–2

Note: Confidence intervals were calculated with bootstrapping by using the percentile method (see Efron and Tibshirani 1993).

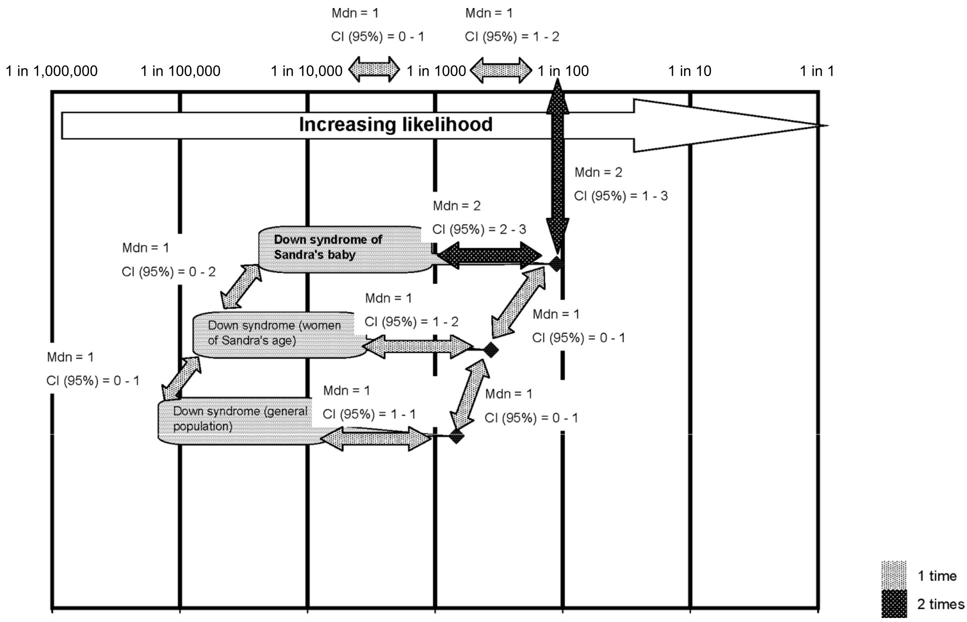

Figure 3. Frequencies of gaze sequences between important AOIs (median values; the darker the pattern, the more often the sequence occurs; all sequences that occur less than once are not marked).

regard to the sequences between the reference risk information and the corresponding data points. Again, these results reflect a greater visual attention to these important parts of the graph. However, it becomes clear that the participants might not have made all the risk comparisons that the graph offers. The results suggest that they compared Data Point 1 directly with Data Point 2, and Data Point 2 directly with Data Point 3. On the other hand, they probably did not make the direct comparison between Data Point 1 and Data Point 3 (median = 0, CI (95%) = 0–0). The same is true for the comparisons between the risk information AOIs. The sequence analysis indicates that the participants focused on the understanding of the target risk (Risk Information 1, Data Point 1, and Scale 5) and paid less attention to the other information presented in the graph.

In sum, the analysis of the overall visual attention to the PPS indicated that the participants focused on the parts of the PPS that were most relevant to the understanding of the target risk information. They paid less attention to the elements that were less important to make such an evaluation. Results further suggested that the participants paid more attention to Risk 2 than to Risk 3, and that they probably only compared the adjacent risk information AOIs.

Correlations between visual attention and subjective numeracy

Mean subjective numeracy was 4.4 (SD = 0.7) on a scale from 1 to 6. The eight items of the scale showed good internal consistency (Cronbach's $\alpha = 0.81$). To explore the relationship between visual attention to the PPS and subjective numeracy, we calculated rank correlations between the subjective numeracy and the total duration, the number of gaze events and the number of AOIs that were considered.

There was a significant negative correlation between the subjective numeracy and the time spent looking at the PPS ($r_S = -0.37$, $p = 0.01$), as well as between the subjective numeracy and the total number of gaze events (r_S = –0.32, p = 0.03). Results thus suggested that the participants with lower subjective numeracy needed more time and more gaze events to extract information from the graph. The absolute number of AOIs that the participants looked at was not significantly correlated with subjective numeracy (r_S = –0.20, ns). We computed a relative version of this variable, in which the number of AOIs that were totally considered was related to the total number of gaze events. Higher values of this relative variable indicated that fewer gaze events were needed to look at the considered AOIs, whereas lower values indicated that more gaze events were necessary to take the considered AOIs into account. Thus, this variable measured a kind of efficiency with which the graph was looked at. This relative variable was positively correlated with subjective numeracy (r_S = 0.34, p = 0.02). These results indicated that the lower subjective numeracy was associated with a longer period of time that was spent looking at the graph and more switching between the AOIs. On the other hand, higher subjective numeracy was related to more efficient processing of the PPS because less time and fewer gaze events were needed to take into account a similar number of elements in the graph.

Discussion

Overall visual attention to the PPS

The first aim of this study is to find out how people look at a risk communication graph in which a risk is placed in the context of other risks. Results show that participants focus on the parts of the graph that are important for the understanding of the target risk information. This shows that it was basically clear for the participants where the information that had to be extracted could be found in the graph. Furthermore, risk ladders provide reference risk information. Our results showed that many participants did look at all three risk information areas. However, they paid less attention to the third risk information area, and they might not have directly compared all three risk information areas. Participants most often compared the target risk information with the adjacent one but not with the second reference risk. There were two explanations for this concentration on the first reference risk. On the one hand, the mere fact that the second reference risk was spatially further away could have led to less attention to this information. On the other hand, the first reference risk provided information about a population similar to that of the target risk. This similarity could have increased the perceived relevance of the first reference risk and, therefore, might have led to increased attention to this information. These results may imply that people, when they were looking at the PPS, did not use the full potential for risk comparisons that this graph actually offered them. Instead, they seemed to focus on the target risk information and its comparison with the reference risk that was spatially closest and/or most similar.

In regard to the logarithmic scale, the participants focused on Scales 4 and 5. One reason for this result was probably the fact that the data points of the three risks were located between these two scale points. Participants had to compare the data points to the corresponding scale points to get an impression of the probabilities of the depicted risks. The remaining scale points were looked at only once and only for a very short time period. As the meaning of a logarithmic scale might not be intuitively easy to understand, we would expect that at least the scale points adjacent to Scales 4 and 5 would be looked at for longer and more often to integrate the meaning of the scale (for the association between longer and more frequent gazes and processing of a graph, see

Renshaw et al. 2004). This was not the case in our sample. It cannot be excluded, therefore, that some participants mistakenly perceived the distance between Scale 4 and Scale 5 not as logarithmic but as linear. Thus, further studies are needed to investigate the understanding of the logarithmic scale.

Our results suggest that two points should be considered when designing risk ladders for graphical risk communication. First, spatial closeness and/or similarity seem to influence which risk comparisons people make in a graphically depicted risk ladder. It is thus crucial for graph designers to choose this reference risk with care and according to the task at hand. Second, our study gives a preliminary suggestion that a logarithmic scale might not be the best choice for a risk communication graph. If the risks depicted in a risk ladder do not vary too much in probability, the comprehension of the graph might be enhanced if a linear scale is used.

Correlations between visual attention and subjective numeracy

The second aim of this study is to find out whether there is a relationship between visual attention to the PPS and subjective numeracy. Overall, lower subjective numeracy is associated with longer total gaze duration and with more gaze events. These results may suggest that it is more difficult for persons with lower subjective numeracy than for persons with higher subjective numeracy to understand the graph and extract information from it (for the association between longer and more frequent gazes and processing of a graph see Carpenter and Shah 1998; Ratwani, Trafton, and Boehm-Davis 2008; Renshaw et al. 2004). This finding is in line with our first hypothesis. Furthermore, subjective numeracy is positively correlated with the relative values of number of considered AOIs. Thus, participants with higher subjective numeracy look at about the same number of elements in a shorter amount of time than participants with lower subjective numeracy. According to Goldberg and Kotval (1999), such relative measures are an indication of higher search efficiency. Thus, this result supports our second hypothesis and adds to the assumptions from Peters and Levin (2008), who suggest that individuals with higher numeracy manage to integrate more information in their decision-making processes than persons with lower numeracy. Our results imply that this mechanism can be due to the efficiency of this integration process.

All in all, participants with lower subjective numeracy seem to have problems extracting and integrating information from the PPS. Thus, our results suggest that graphs might not be easy to process per se and should be customized to simplify the information integration process for persons with lower numeracy. This finding might be of crucial relevance to risk communication, as graphical displays are considered to be less complex than numbers and are therefore recommended for risk communication with persons of higher age or low numeracy (see, e.g., Finucane 2008). It can also explain why participants with lower numeracy in Keller and Siegrist's study (2009) are not able to differentiate between high and low risks when these risks are depicted in the PPS.

However, the modification of graphs alone might not be enough to enhance the understanding of risk communication graphs. The situation in the laboratory, with the eye-tracker on the head of the participant and a researcher in the same room who could see where participants were looking, was probably highly motivating for all participants to gaze at the graph until they had looked at most parts of it. In some situations, such an external motivation may be present, for example a medical doctor who presents such a graph and emphasizes the importance of understanding the presented probability. In other everyday situations, however, this kind of motivation might not be present. It is an open question whether people would then also take the same time and effort to look at a risk communication graph in the absence of external motivation. There are indications that information seeking behaviour is positively associated with context-specific self-efficacy (Bass et al. 2006; Hong 2006; Satia, Galanko, and Neuhouser 2005). It is, therefore, possible that, in some real-world situations, people with lower subjective numeracy look for less information in graphs because of their lower self-efficacy in understanding the numerical risk information depicted in these graphs. This mechanism might also lead to differences between persons with high and low numeracy in understanding risk information depicted in graphs as observed by Keller and Siegrist (2009). If this was the case, the modification of the graph alone would not enhance risk communication efforts, as the crucial element would not be the understanding of the graph but the attention that was paid to the graph.

In addition, one should keep in mind that we used hypothetical scenarios that focused on the processing of graphically depicted information and that did not include decision-making based on this information. In reality, however, such graphs would be used in situations that could have severe medical implications for the patients showing them. These special situations might influence the processing of risk graphs in many ways that we could not observe in this study. Worry about cancer, for example, has been found to be associated with more attention to health information (Beckjord et al. 2008). On the other hand, seeking information about medical conditions might be inhibited by negative affect when one has the feeling of less control over one's health (Lee et al. 2008).

In short, future studies are needed to show whether the skills used to extract information from a graph, as well as the complexity of the graph, the external motivation or an interaction of these factors influence the understanding of graphically depicted risks. Moreover, the role of specific aspects of real decision-making situations should be examined, such as motivating or inhibiting factors interplaying with numeracy in the processing of graphically depicted risks.

Limitations

This study has several limitations. First, we recruited our sample in a way to ensure a broad range of subjective numeracy. This led to some biases concerning other variables. The sample included more men than women, and the mean age as well as the mean level of education was rather high. Thus, the results may not be easily generalized to other populations and should first be verified in other samples. Second, there were some limitations related to the eye-tracking data. We used a mobile eye-tracker that could not give exact fixations due to movements of the head. Therefore, our data consisted of manually coded cumulative gaze duration and frequency variables. The studies on which we based our interpretations, on the other hand, usually took fixations or even saccades as units of measurement (Carpenter and Shah 1998; Goldberg and Kotval 1999; Ratwani, Trafton, and Boehm-Davis 2008; Renshaw et al. 2004). Furthermore, these previous studies were conducted on search tasks or on usability testing of computer interfaces. Thus, it was not perfectly clear whether these findings could be transferred one-to-one to gaze event data such as ours. Given these limitations, we tried to apply the interpretational framework with care to prevent overinterpretation of our data.

Conclusion

The results of the present study suggest that an eye-tracker is a useful tool for evaluating the perception of graphical risk communication, as it can indicate which information receives visual attention and, thus, which may be easily processed. The present research also demonstrates that an eye-tracker can be successfully used for examining individual differences. Higher subjective numeracy seems to be related to more efficient processing of the PPS. This result suggests that graphical risk communication, which some scholars explicitly recommend for persons with lower numeracy (Apter et al. 2008; Nelson et al. 2008), should be further improved to address the needs of these persons. Our findings imply that the reference information, and maybe also the logarithmic scale, can be starting points for the modification of the PPS. These implications can be of use in contexts other than medical risk communication (e.g. for environmental risks, where risk ladders are also common, see Connelly and Knuth 1998; Sandman, Weinstein, and Miller 1994), which is important for findings in risk research more generally (Loefstedt and Six 2008).

Note

1. Not shown in Figure 2 is an AOI labelled 'scale undefined', which means that the gaze is somewhere on the scale, but not clearly focused on one of the scale points.

References

Apter, A.J., M.K. Paasche-Orlow, J.T. Remillard, I.M. Bennett, E.P. Ben-Joseph, R.M. Batista, J. Hyde, and R.E. Rudd. 2008. Numeracy and communication with patients: They are counting on us. *Journal of General Internal Medicine* 23, no. 12: 2117–24.

Bandura, A. 1997. *Self-efficacy: The exercise of control.* New York: Freeman.

Bass, S.B., S.B. Ruzek, T.F. Gordon, L. Fleisher, N. McKeown-Conn, and D. Moore. 2006. Relationship of internet health information use with patient behavior and self-efficacy: Experiences of newly diagnosed cancer patients who contact the National Cancer Institute's Cancer Information Service. *Journal of Health Communication* 11, no. 2: 219–36.

Beckjord, E.B., L.J.F. Rutten, N.K. Arora, R.P. Moser, and B.W. Hesse. 2008. Information processing and negative affect: Evidence from the 2003 health information national trends survey. *Health Psychology* 27, no. 2: 249–57.

Bostrom, A. 2008. Lead is like mercury: Risk comparisons, analogies and mental models. *Journal of Risk Research* 11, nos. 1–2: 99–117.

Caffier, P.P., U. Erdmann, and P. Ullsperger. 2003. Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology* 89, nos. 3–4: 319–25.

Carpenter, P.A., and P. Shah. 1998. A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology-Applied* 4, no. 2: 75–100.

Connelly, N.A., and B.A. Knuth. 1998. Evaluating risk communication: Examining target audience perceptions about four presentation formats for fish consumption health advisory information. *Risk Analysis* 18, no. 5: 649–59.

Edwards, A., G. Elwyn, and A. Mulley. 2002. Explaining risks: Turning numerical data into meaningful pictures. *British Medical Journal* 324, no. 7341: 827–30.

Efron, B., and R.J. Tibshirani. 1993. *An introduction to the bootstrap.* New York: Chapman & Hall.

Fagerlin, A., B.J. Zikmund-Fisher, P.A. Ubel, A. Jankovic, H.A. Derry, and D.M. Smith. 2007. Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making* 27, no. 5: 672–80.

Feldman-Stewart, D., N. Kocovski, B.A. McConnell, M.D. Brundage, and W.J. Mackillop. 2000. Perception of quantitative information for treatment decisions. *Medical Decision Making* 20, no. 2: 228–38.

Finucane, M.L. 2008. Emotion, affect, and risk communication with older adults: Challenges and opportunities. *Journal of Risk Research* 11, no. 8: 983–97.
Gigerenzer, G., and A. Edwards. 2003. Simple tools for understanding risks: From innumeracy to insight. *British Medical Journal* 327, no. 7417: 741–4.
Goldberg, J.H., and X.P. Kotval. 1999. Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics* 24, no. 6: 631–45.
Grunert, K.G., and J. Wills, M. 2007. A review of European research on consumer response to nutrition information on food labels. *Journal of Public Health* 15, no. 5: 385–99.
Holsanova, J., N. Holmberg, and K. Holmqvist. 2009. Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology* 23, no. 9: 1215–26.
Hong, T. 2006. The Internet and tobacco cessation: The roles of Internet self-efficacy and search task on the information-seeking process. *Journal of Computer-Mediated Communication* 11, no. 2: 536–56.
Just, M.A., and P.A. Carpenter. 1976. Eye fixations and cognitive processes. *Cognitive Psychology* 8, no. 4: 441–80.
Keller, C., and M. Siegrist. 2009. Effect of risk communication formats on risk perception depending on numeracy. *Medical Decision Making* 29, no. 4: 483–90.
Korner, C. 2004. Sequential processing in comprehension of hierarchical graphs. *Applied Cognitive Psychology* 18, no. 4: 467–80.
Lee, S.Y., H. Hwang, R. Hawkins, and S. Pingree. 2008. Interplay of negative emotion and health self-efficacy on the use of health information and its outcomes. *Communication Research* 35, no. 3: 358–81.
Lipkus, I.M. 2007. Numeric, verbal, and visual formats of conveying health risks: Suggested best practices and future recommendations. *Medical Decision Making* 27, no. 5: 696–713.
Loefstedt, R.E., and P. Six. 2008. What environmental and technological risk communication research and health risk research can learn from each other. *Journal of Risk Research* 11, nos. 1–2: 141–67.
Marteau, T.M., E. Dormandy, and S. Michie. 2001. A measure of informed choice. *Health Expectations* 4, no. 2: 99–108.
Mooney, C.Z., and R.D. Duval. 1993. *Bootstrapping: A nonparametric approach to statistical inference (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07–095).* Newbury Park, CA: Sage.
Nelson, W., V.F. Reyna, A. Fagerlin, I. Lipkus, and E. Peters. 2008. Clinical implications of numeracy: Theory and practice. *Annals of Behavioral Medicine* 35, no. 3: 261–74.
Paling, J. 2003. Strategies to help patients understand risks. *British Medical Journal* 327, no. 7417: 745–8.
Peters, E., J. Hibbard, P. Slovic, and N. Dieckmann. 2007. Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Affairs* 26, no. 3: 741–8.
Peters, E., and I.P. Levin. 2008. Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. *Judgment and Decision Making* 3, no. 6: 435–48.
Peters, E., D. Vastfjall, P. Slovic, C.K. Mertz, K. Mazzocco, and S. Dickert. 2006. Numeracy and decision making. *Psychological Science* 17, no. 5: 407–13.
Ratwani, R.M., J.G. Trafton, and D.A. Boehm-Davis. 2008. Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology: Applied* 14, no. 1: 36–49.
Rayner, K., B. Miller, and C.M. Rotello. 2008. Eye movements when looking at print advertisements: The goal of the viewer matters. *Applied Cognitive Psychology* 22, no. 5: 697–707.
Renshaw, J.A., J.E. Finlay, D. Tyfa, and R.D. Ward. 2004. Understanding visual influence in graph design through temporal and spatial eye movement characteristics. *Interacting with Computers* 16, no. 3: 557–78.
Reyna, V.F., and C.J. Brainerd. 2007. The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences* 17, no. 2: 147–59.
Salvucci, D.D., and J.R. Anderson. 2001. Automated eye-movement protocol analysis. *Human-Computer Interaction* 16, no. 1: 39–86.

Sandman, P.M., N.D. Weinstein, and P. Miller. 1994. High risk or low: How location on a 'risk ladder' affects perceived risk. *Risk Analysis* 14, no. 1: 35–45.
Satia, J.A., J.A. Galanko, and M.L. Neuhouser. 2005. Food nutrition label use is associated with demographic, behavioral, and psychosocial factors and dietary intake among African Americans in North Carolina. *Journal of the American Dietetic Association* 105, no. 3: 392–402.
Schwartz, L.M., S. Woloshin, W.C. Black, and H.G. Welch. 1997. The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine* 127, no. 11: 966–72.
Siegrist, M., M.-E. Cousin, and C. Keller. 2008. Risk communication, prenatal screening, and prenatal diagnosis: The illusion of informed decision-making. *Journal of Risk Research* 11, no. 1–2: 87–97.
Siegrist, M., P. Orlow, and C. Keller. 2008. The effect of graphical and numerical presentation of hypothetical prenatal diagnosis results on risk perception. *Medical Decision Making* 28, no. 4: 567–74.
Stallings, S.P., and J.E. Paling. 2001. New tool for presenting risk in obstetrics and gynecology. *Obstetrics and Gynecology* 98, no. 2: 345–9.
Visschers, V.H.M., R.M. Meertens, W.F. Passchier, and N.K. de Vries. 2007. How does the general public evaluate risk information? The impact of associations with other risks. *Risk Analysis* 27, no. 3: 715–27.
Visschers, V.H.M., R.M. Meertens, W.F. Passchier, and N.K. de Vries. 2009. Probability information in risk communication: A review of the research literature. *Risk Analysis* 29, no. 2: 267–87.
Waters, E.A., N.D. Weinstein, G.A. Colditz, and K. Emmons. 2006. Formats for improving risk communication in medical tradeoff decisions. *Journal of Health Communication* 11, no. 2: 167–82.
Weinstein, N.D., P.M. Sandman, and W.K. Hallman. 1994. Testing a visual display to explain small probabilities. *Risk Analysis* 14, no. 6: 895–6.
Woloshin, S., L.M. Schwartz, S. Byram, B. Fischhoff, and H.G. Welch. 2000. A new scale for assessing perceptions of chance: A validation study. *Medical Decision Making* 20, no. 3: 298–307.
Zikmund-Fisher, B.J., D.M. Smith, P.A. Ubel, and A. Fagerlin. 2007. Validation of the subjective numeracy scale: Effects of low numeracy on comprehension of risk communications and utility elicitations. *Medical Decision Making* 27, no. 5: 663–71.