DOI: 10.1111/j.1539-6924.2011.01636.x

Reducing the Effort to Comprehend Risk Models: Text Labels Are Often Preferred Over Graphical Means

Ida Hogganvik Grøndahl,[1] Mass Soldal Lund,[2] and Ketil Stølen[2,3*]

Risk analysis involves people with different roles and competences. The validity of the outcome depends on that they are able to communicate; ideally between themselves, but at least with or via a risk analyst. The CORAS risk modeling language has been developed to facilitate communication between stakeholders involved in the various stages of risk analysis. This article reports the results from an empirical investigation among professionals, where the purpose was to investigate how graphical effects (size, color, shape) and text labels introduced in the CORAS risk modeling language affected the understanding. The results indicate that if graphical effects are used to illustrate important information, they should also be accompanied by informative textual labels.

KEY WORDS: Empirical study; modeling; risk; risk analysis; visualization

1. INTRODUCTION

The CORAS risk modeling language,[(1)] in the following referred to as the CORAS language, has been designed to document risks, facilitate analysis, and communicate risk-relevant information throughout the various phases of an asset-driven, defensive risk analysis process. By asset-driven we mean that the main assets identified initially by the customer are used to focus everything that happens thereafter in the analysis; by defensive we mean that the focus is on defending the assets as in security- or safety-oriented risk analysis, rather than building new assets in addition to defending the existing ones as in financial risk analysis. To facilitate communication between participants of diverse backgrounds, the CORAS language employs simple icons and relations that are easy to read. In particular, the CORAS language is meant to be used during brainstorming sessions where discussions are documented along the way.

The CORAS language along with its detailed user guidelines, the work on which was initiated in 2002, has been developed in an iterative manner driven by industrial field trials.[(2,3)] The major decisions regarding its underlying foundation, notation, and guidelines are supported by empirical investigations.[(4,5)] Some of these investigations, although motivated by the needs of CORAS, are of relevance for risk modeling in general. This article presents results of such general nature originating from a study of risk modeling preferences among professionals within the IT sector.

From earlier studies of risk analysis concepts and terminology, we know that some concepts are more difficult to understand than others.[(5)] These studies led to the hypothesis that the concepts that were found easy to understand would also be most suitable and appropriate to represent in a simple and straightforward manner, while the difficult concepts would need more sophisticated representations.

Relying on this hypothesis we decided in the prolongation of these studies to investigate how we best

[1] Scandpower Risk Management, Norway.
[2] SINTEF ICT, Norway.
[3] University of Oslo, Norway.
[*] Address correspondence to Ketil Stølen, SINTEF ICT, Norway; kst@sintef.no

0272-4332/11/0100-1813$22.00/1

could represent three such concepts, namely (1) likelihood of a threat scenario path, (2) vulnerability, and (3) severity of risks. This article reports on this investigation. The goal was to identify representations that would convey, in an intuitive manner, the intended meaning of risk models. The research questions for the study on which this article reports are presented in Section 2. The study tested various means for representation known from the field of information visualization,[6] in addition to textual information labels. As explained in the conclusion, the study had direct impact on the development of the CORAS language.

The article is structured as follows: in Section 2, we give the background and motivation for our investigations. In Section 3 we present related work. In Section 4, we introduce various techniques for information visualization that can be used in modeling, and then describe the alternatives we explored. Section 5 presents the design of the experiment. The results are given in Section 6. Our findings are discussed in Section 7 and the identified threats to validity are reported in Section 8. Section 9 presents the conclusions of our work.

2. BACKGROUND AND MOTIVATION

The motivation for the study had two sources: (1) our previous findings from a number of experiments regarding the conceptual foundation of asset-driven, defensive risk analysis,[4,5] and (2) experiences from the industrial field trials mentioned earlier, for which a summary is given by Hogganvik.[3] In the following, we describe this in more detail. Let us first explain the conceptual foundation on which our work is based, presented as a Unified Modeling Language (UML) class diagram[7] in Fig. 1. Its definitions are taken from the following international standards:

- Information Technology: Guidelines for Management of IT Security (ISO/IEC13335),[8,9]
- Australian/New Zealand Standard for Risk Management (AS/NZS4360),[10]
- Information Security Risk Management Guidelines (HB231).[11]

The model can be explained as follows: *stakeholders* are those people and organizations who may affect, be affected by, or perceive themselves to be affected by a decision or activity regarding the target of analysis.[10] An *asset* is something to which a stakeholder directly assigns value, and hence for which the stakeholder requires protection.[11] Assets are subject to *vulnerabilities*, which are weaknesses that can be exploited by one or more threats.[9] A *threat* is a potential cause of an unwanted incident.[9] A threat may be classified as human (with accidental origin or deliberate harmful intentions) or nonhuman (also called environmental).[9] An *unwanted incident* is an event that may harm or reduce the value of assets—something we want to prevent.[8] A *risk* is the chance of something happening that will have an impact upon objectives (assets).[10] Our model captures this interpretation by defining a risk to consist of an unwanted incident, a likelihood measure, and a consequence measure. The abstract concept *risk*, the more concrete *unwanted incident*, and their respective relationships to *asset* require a more in-depth explanation. In our definition, an unwanted incident that harms more than one asset gives rise to one distinct risk toward each of the assets. This allows us to quantify loss of asset value using different scales for different assets; for example, loss of investment is straightforwardly measured in terms of monetary value, while information leakage may be more easily measured by the number of information records leaked. It also enables us to handle conflicting consequence estimates for different stakeholders since an asset is always defined with respect to a single stakeholder. The level of risk is measured by a *risk value*[10] (e.g., low, medium, high, or other scales) that is based upon the estimated *likelihood* (a general description of frequency or probability[10]) of an unwanted incident occurring and its *consequence* in terms of damage to an asset. A *treatment* is the

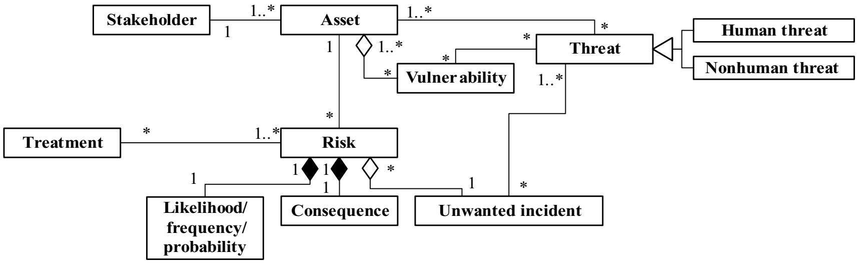


Fig. 1. The conceptual foundation.

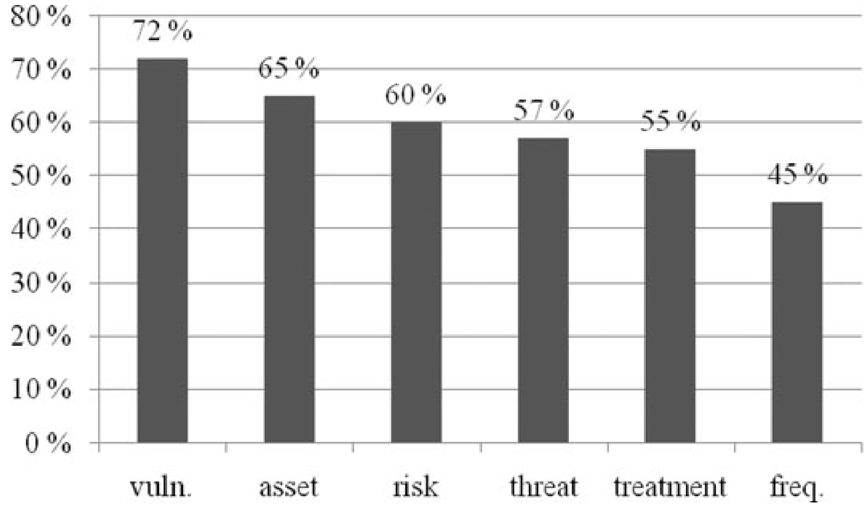


Fig. 2. Percentage of correct answers for each category (mean).

selection and implementation of appropriate options for dealing with risk.[10]

An experiment focusing on the understanding of this conceptual foundation[5] was conducted prior to the study reported in this article. This experiment included 57 subjects who were professionals within the IT sector as well as IT students. They were given a questionnaire with questions tailored to investigate the respondents' understanding of risk-analysis-related concepts: *asset*, *treatment*, *frequency measures (freq.)*, *risk*, *vulnerability (vuln.)*, and *threat*. Their answers were then evaluated as *correct*, *wrong*, or *uncertain.* The percentage of correct answers for each of the concepts is presented in Fig. 2. As can be seen from the figure, the questions related to vulnerabilities and assets received the most correct answers, while questions related to frequencies (likelihoods) received a much lower score. The latter corresponds well with the experiences from the field trials, which identified the need for a good way of specifying the likelihood of the *paths leading to unwanted incidents that constitute risks*; in particular, visualizing the *most likely* paths.

Although the concept *vulnerability* seemed to be well understood, being able to specify vulnerabilities explicitly in scenarios leading to unwanted incidents was identified as important in several of the industrial field trials. However, how this should be done was unclear. In the empirical investigations[4,5] as well as in the field trials,[2,3] we experienced that although the concept *risk* belongs to everyday vocabulary, most people find it difficult to specify exactly what a risk is as well as its severity. The concept of risk is different from the other concepts studied in the sense that it is more abstract and is used to cover other, more concrete concepts.

The CORAS risk modeling language has been developed over several years.[12–14] The abovementioned field trials providing continuous feedback have been very important by providing a test-bed for different solutions and for identifying needs. Among the important suggestions from these field trials is that the modeling of causal scenarios leading to risks provided by the CORAS language is useful, but that it would be beneficial to include *vulnerabilities* in these scenarios.[12] In an empirical investigation conducted on an earlier version of the CORAS language, we concluded that visualizing the various concepts with meaningful graphical icons increased the readability of the diagrams.[4]

The findings and considerations summarized motivated us to investigate the following:

(1) How should we visualize likelihood measures in relation to graph navigation (reading and understanding the diagrams), especially the likelihood of threat scenario paths?
(2) How should we visualize vulnerabilities?
(3) How can we visualize risks in the models to improve the understanding of this concept and in particular its severity?

It may be argued that some of the concepts of risk analysis, and especially the concept of risk, are inherently difficult to understand. This makes it, however, even more important to find good ways to convey their correct interpretation. Our concern is purely pragmatic, namely, improved communication with stakeholders and participants in risk analyses.

While visualization is not the only way of achieving this, it is an important means requiring careful consideration.

In the following, we deal with each of the three issues listed in separate subsections. The first is referred to as "representing graph navigation" (Sections 4.1, 7.1, and 9.1), the second as "representing vulnerabilities" (Sections 4.2, 7.2, and 9.2), and the third as "representing risk" (Sections 4.3, 7.3, and 9.3).

3. RELATED WORK

The study of the effectiveness of visual or graphical communication of uncertainty and risk goes back a couple of decades. Ibrekk and Morgan[15] present an important study where a group of nontechnical people was subjected to a number of graphical means for visualizing uncertainty. Although their results have no direct relevance to our study, they show that how uncertainty is presented by graphical means is certainly not irrelevant for how it is perceived. A further interesting result of their study is that some knowledge of statistics did not significantly improve the subjects' performance. In a study of communication of health risks, Connelly and Knuth[16] found that their respondents felt that a presentation combining a short text and an illustration was clearer and easier to understand than a longer piece of text.

According to a survey from 1999 by Lipkus and Hollands,[17] studies testing visual aids in risk communication until that point in time were few and did not in a satisfactory manner explain *why* particular visual aids should enhance risk communication or how the tasks at hand (what is the purpose of the communication) affected the results. Still they conclude that the evidence available in 1999 points in the direction that visual aids are useful for communicating risk, but that the tasks of the reader (the purpose of the communication) always must be considered when choosing what aids to apply. Of interest to our study is their recommendation that areas or volumes should be avoided when visualizing magnitudes as readers tend to get their estimates of the magnitudes wrong when subjected to these visual aids.

Winn[18] points out that phrases such as "a picture is worth a thousand words" are too simplistic since the usefulness of pictures is highly dependent on the situation at hand and the skills of the reader. Nevertheless, software engineering studies have shown that applying graphical means to program and component specifications increases the understanding of the system,[19] and graphical means have been applied successfully for visualizing program interdependencies.[20] According to Larkin and Simon,[21] the key activities of a reader of a diagram is searching and recognizing relevant information, and then using this to draw conclusions.

Table I. Symbols Used in the Diagrams

Symbol	Meaning
Stick figure with bomb	Threat
Oval with bomb	Threat scenario
Rectangle with explosion	Unwanted incident
Stacks of coins	Asset
Padlock	Vulnerability
Logical gate	Logical gate
Danger sign (red triangle)	Risk

4. MEANS AND MECHANISMS TO EASE COMPREHENSION

In the following, we motivate and present the means and mechanisms that were selected as subjects for investigation, structured according to the three main issues identified at the end of Section 2: *representing graph navigation, vulnerabilities*, and *risks*. At the end of each subsection, we present the alternatives for the representation investigated in the experiment. The tasks of the experiment, all requiring the subjects to compare diagrams, were numbered Task 1 through Task 7 according to the order that the subjects were asked to perform them. The symbols of the diagrams are explained in Table I.

We refer to the different tasks of the experiment using the following convention: "TE2" means Task 2 in the experiment, "D1" means Diagram Alternative 1, "TE2D1D2" means Diagram Alternative 1 compared to Diagram Alternative 2 in Task 2. In the following the tasks are presented thematically. The order in which the tasks are presented therefore diverts from their numbering within the experiment. The theme of Task 1 falls outside of what is treated in this article, and this task is therefore not part of the presentation and the results.

4.1. Representing Graph Navigation

There will usually be a number of paths through a threat diagram, and each of these represents a scenario describing how a threat may harm an asset. In our experience, sorting the important information

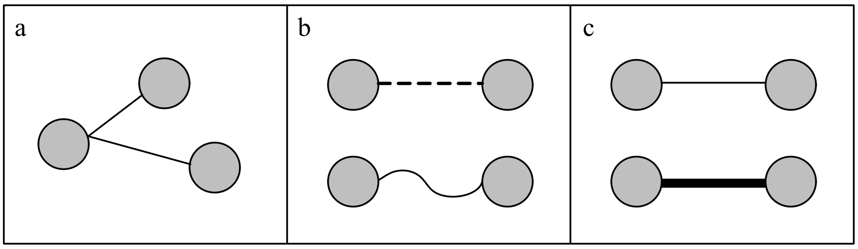


Fig. 3. Node-link configurations.

from the less important is essential in a risk analysis, and also a major challenge. We therefore believe it to be useful to draw special attention to the most likely paths. It should be noted that we did not test the hypothesis that this is useful, but rather how we best can represent graphically the most likely path under the assumption that this is useful.

We focus on the node-link diagram type, which in its simplest form consists of nodes connected via edges. A closed contour in a node-link diagram generally represents a concept of some kind, and a linking line between concepts represents some kind of relationship between them. These lines may be manipulated to represent different meanings (Fig. 3).

Lines linking closed contours may have different colors, be straight or wavy, or have other graphical qualities that represent an attribute or type of relationship.(6) The thickness of a connecting line may be used to represent the magnitude of a relationship (a scalar attribute). A line with different thickness or color may "pop-out" perceptually and thereby call for the reader's attention.(22) We chose to test thick, thin, and dashed lines to illustrate more or less likely paths through the threat diagrams. The use of line variations was also inspired by network traffic maps that often use node-link diagrams to visualize network traffic. The load or type of link data can typically be illustrated using different line colors or varying the thickness of the lines.(23) Variation in appearance of lines is also inspired by the gestalt principle *similarity*, meaning something that is different from its surroundings will be easier identified by the reader (e.g., *italic style* vs. normal style in text). The gestalt principles(24) are some of the most recognized sets of rules for good graphical displays. Implementing the gestalt principles may reduce the effort needed to understand illustrations, program interfaces, websites, etc.

Winn(25,26) has found that people whose languages are written from right to left also process graphics from right to left. A related finding is that the item to the left was always interpreted as the cause, while the item to the right was the effect.(27) This is in accordance with one of the node-link rules(6) stating that "placing closed contours spatially in an ordered sequence can represent conceptual ordering of some kind." We decided to indicate the direction with arrows since this has been found helpful in understanding similar notations like flow charts (used in schematic representations of processes).(28) If the number of lines increases, there is a danger of losing track of the path. Repeated arrows may help distinguish between cross-overs and connection points.

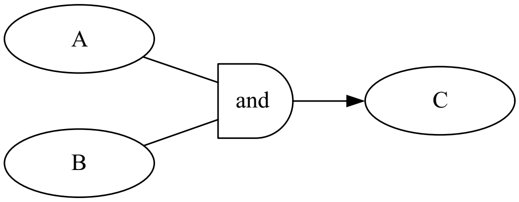


Fig. 4. The AND-gate symbol.

We needed a way of saying both "If A and B occurs, C must occur" and "If either A or B occurs, C must occur." The challenge was to find the preferred way of illustrating this. When looking for an AND operator symbol, it was natural to turn to one of the most frequently used techniques in risk analysis: fault tree analysis (FTA).(29) Fault trees let you specify dependencies between events in a tree structure using logical AND/OR operators called *gates*, adopted from electrical circuit design.(30) We decided to test the logical gate for "AND" and added an information label with the text "and" to the symbol to ease comprehension for people without background in FTA or circuit design (Fig. 4). There is a similar symbol for OR gates, but we chose to test only one of them as they are very similar.

In addition to the AND-gate symbol, we also tested dashed lines (*cf.* the above discussion on appearance of lines) and "information notes" (inspired by the UML)(7)) as means for expressing "AND." In UML, notes may be attached to a diagram element to convey additional information about the specific element. The idea was to use the notes as an alternative to logical gates.

4.1.1. Set-Up for "Representing Graph Navigation"

The effect of two different line styles was tested in TE2 for visualizing paths that are more likely to be chosen by a threat (Fig. 5):

- D1: arrows with dashed lines are used to illustrate the less likely paths and arrows with solid lines are used to illustrate the more likely paths.
- D2: arrows with thick lines are used to illustrate the more likely paths and arrows with thin lines are used to illustrate the less likely paths.

Three modeling alternatives for logical AND were tested in TE4 (Fig. 6):

- D1: the special AND-gate symbol from fault tree notation and electrical circuit design labeled with the text "and" is used to illustrate AND dependency.
- D2: arrows with dashed lines are used to illustrate AND dependency.
- D3: UML-like notes consisting of a dashed line with the label "and" between two arrows are used to illustrate AND dependency.

4.2. Representing Vulnerabilities

In research on the quality of modeling languages, pictures are claimed to have a much higher perceptibility, and information conveyed in pictures is emphasized at the cost of textual information.[31] This is in accordance with the results from our experiment with graphical symbols where the textual stereotyping (tagging) seemed to be overlooked.[4] In the same experiment, we tested diagrams with general UML symbols tagged by text labels versus diagrams where also graphical icons were attached to the symbols. In the experiment we found that the representation using special risk-related icons performed better than the one with conventional UML symbols. Representing vulnerabilities by special symbols was not part of this experiment, but the findings suggest that carefully designed symbols may be useful, and that vulnerabilities could be represented by a graphical symbol in addition to the text.

We chose to use an *open padlock symbol*, which is often used to symbolize a lack of security within the computer security domain. It is a simple and internationally known symbol. To achieve what by Goodman[32] is called "syntactic disjointness," the symbol is unique and it is neither too small nor too large to cause misinterpretations (large elements often attract attention at the cost of the smaller elements). Syntactic disjointness makes it easier to separate the different elements in a model. We decided to compare this representation to an approach where vulnerabilities are listed only textually and assigned to the assets of relevance.

The latter may be seen as an example of the gestalt principle *proximity*, stating that things that logically belong together should be placed close to each other (commonly used in program interfaces and websites). This view is very suitable when the

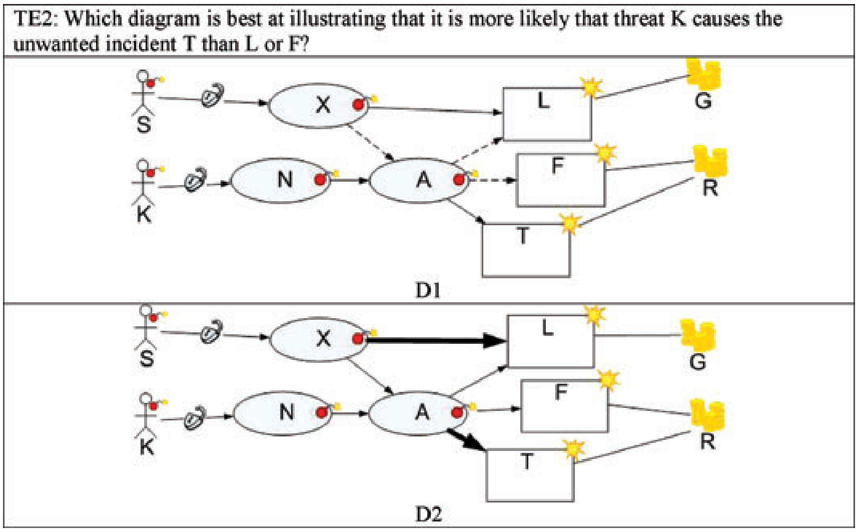
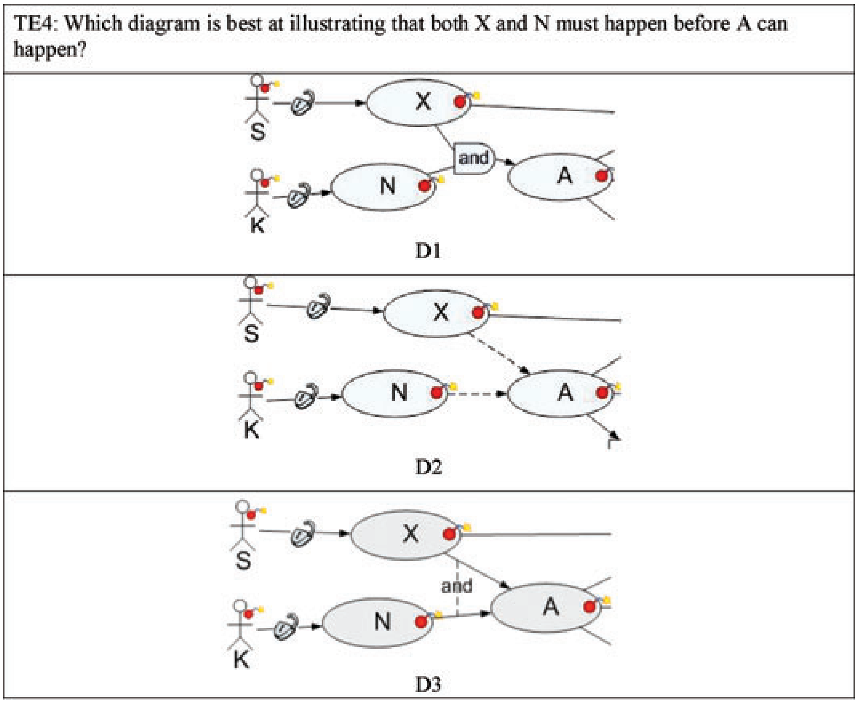

TE2: Which diagram is best at illustrating that it is more likely that threat K causes the unwanted incident T than L or F?

Fig. 5. The diagram alternatives for "most likely path through a graph."



TE4: Which diagram is best at illustrating that both X and N must happen before A can happen?

Fig. 6. The diagram alternatives for "logical AND in graph paths."

focus is on the assets and their vulnerabilities, but when it comes to addressing *which* threats exploit *which* vulnerabilities and especially *where* in the chain of events the exploitation takes place, the situation is different. This kind of information is especially valuable in relation to treatment identification. As a consequence of this, we decided to try modeling the vulnerabilities where they logically belong in the chain of events. The experiment is therefore concerned with the placement of vulnerabilities, as symbols in the paths representing chains of events or as text attached to the assets.

4.2.1. Set-Up for "Representing Vulnerabilities"

Two alternative representations of shared vulnerabilities (common for two or more assets) were tested in TE5 (Fig. 7):

- D1: vulnerabilities are represented by an open padlock symbol and a vulnerability description. The vulnerabilities are located where they logically belonged in the sequence of events.
- D2: vulnerabilities are represented by specifying them as textual attributes of each asset. Since the same vulnerability may be relevant for more than one asset, it may be listed under several assets.

The same alternatives were also tested in Task 6 (TE6, Fig. 8), but in the opposite order (i.e., TE6D1 = TE5D2 and TE6D2 = TE5D1). However, the question asked was changed to ask about exploited vulnerabilities instead of shared vulnerabilities. This was done in order to investigate whether or not the respondents had different preferences toward the alternatives with respect to different aspects of vulnerabilities.

4.3. Representing Risks

Representing risks is a great challenge due to their abstract nature. Illustrating the severity or magnitude of risks and the impact of unwanted incidents helps keep focus on the most critical risks, something that is especially useful during the final risk treatment phase. To find a way to visualize the severity of risks, we investigated means like line thickness, textual information labels, shapes that varied in size or color, and more. The color of an enclosed region may represent a concept type, and the size may be used to represent the magnitude of a concept. This is also inspired by the gestalt principle of breaking the pattern to attract attention. We want the reader to quickly discover the most serious unwanted incidents since they often represent major risks. We therefore tested the effect of marking the most serious unwanted

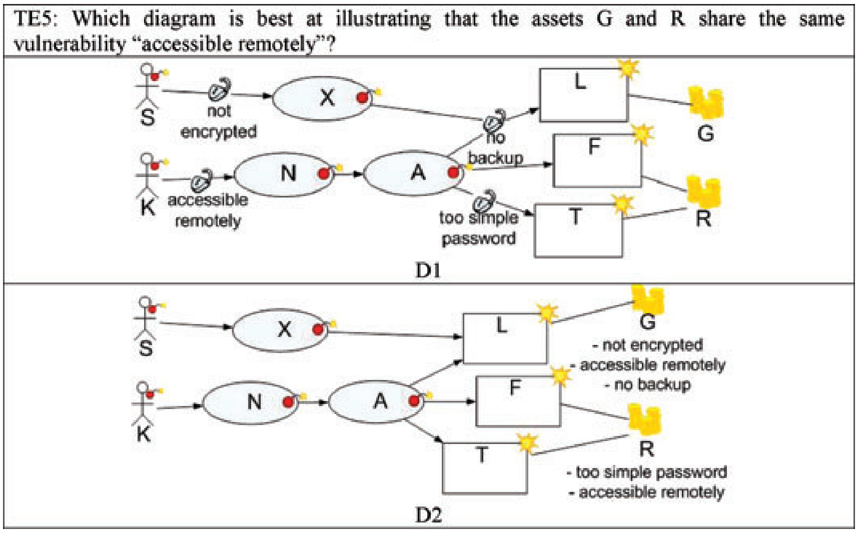

TE5: Which diagram is best at illustrating that the assets G and R share the same vulnerability "accessible remotely"?

Fig. 7. The diagram alternatives for "shared vulnerabilities."

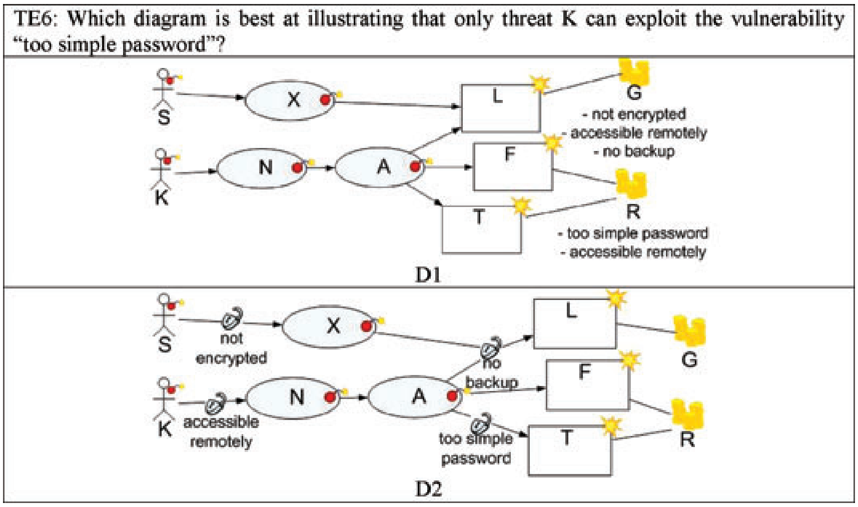
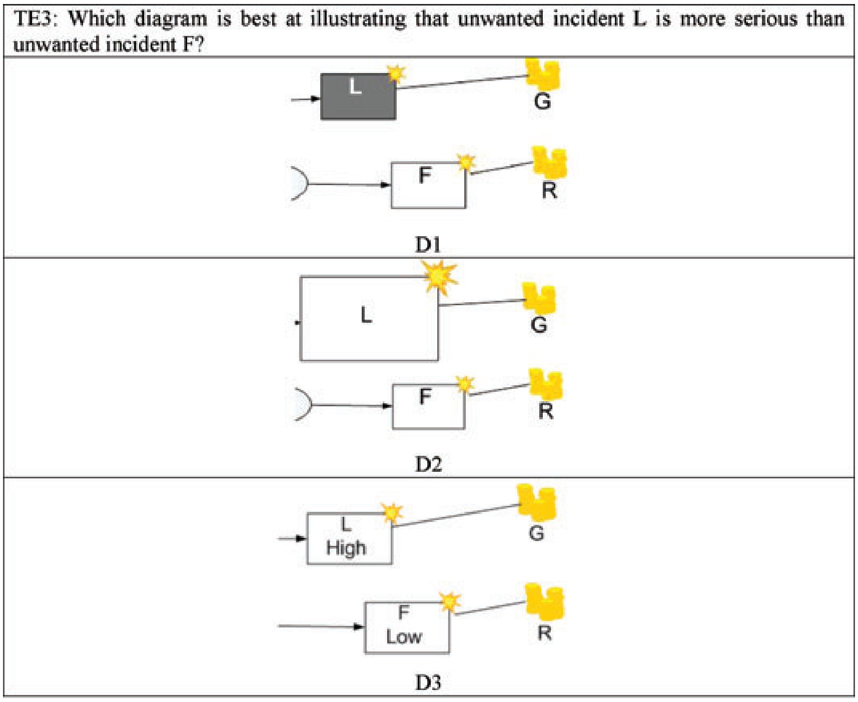

TE6: Which diagram is best at illustrating that only threat K can exploit the vulnerability "too simple password"?

Fig. 8. The diagram alternatives for "which vulnerabilities are exploited by which threats."

incidents with a darker color since dark colors are more easily noticed against white background than light colors.

There are some aspects to be aware of when using color coding. The number of different colors one may use is limited by the reader's ability to remember and distinguish the colors. Much research has been conducted to find the optimal number of colors, and the suggestions vary from 5 to 8.(6,33–37) In our case, we only used gray and white. It is also important to keep in mind how a model with colored symbols will look when printed in black and white and whether this may affect the interpretation of the symbols.(32) Different types of coding in statistical graphs have been investigated with respect to display search time. The coding that gave the best performance was color, the second best coding was shape, and third came letters/digits.(38) This top ranking of color has been confirmed by many other studies.(39) According to these results, we should benefit from using color to emphasize the most serious incidents, meaning that the reader should identify them more quickly as compared to using other means. The color alternative was compared to using *size* to symbolize the most serious incident since large elements are more easily noticed than small ones (D2, Fig. 9). As in the case of colors, the number of different size categories is not indefinite. The number of different size steps that can be distinguished from each other at a glance is as low as four.(6) An important aspect when using shapes is to avoid symbols that are too similar. Shapes that are too similar have been found to increase search time and are therefore not recommended.(40) The number of risk levels used in a risk



TE3: Which diagram is best at illustrating that unwanted incident L is more serious than unwanted incident F?

Fig. 9. The diagram alternatives for "the magnitude of unwanted incidents.'

analysis can vary, but usually the number of levels is between two and five. In our experiment material, we compared only two size categories as we were mainly interested in testing the concept of using different sizes.

We also explored the use of textual information in the unwanted incident symbol and as annotations to lines. Our motivation for doing this was that studies of rapid processing of information[(41,42)] have found letters and digits to be some of the best forms of coding. These investigations involved studying short-term memory and how graphical means are memorized using Sternberg tasks.[1] Digits, colors, and letters were found to be the top three coding forms for rapid processing of information. Another investigation found digits, words, and letters to be superior to colors and shapes when it comes to processing speed,[(43)] and digits, letters, and words were found to represent less subjective workload than colors and shapes.[(43)] These findings support the modeling alternatives that use textual information labels to illustrate the seriousness of unwanted incidents and risks.

Ware[(6)] describes some general rules regarding the use of shapes. For instance, the shape of a closed contour can be used to represent a type of concept. We investigated the possibility of representing risks with a symbol and using size to represent their severity. We selected a traditional red and white road sign that symbolizes "danger" and that varies in size according to the severity of the risk.

4.3.1. Set-Up for "Representing Risks"

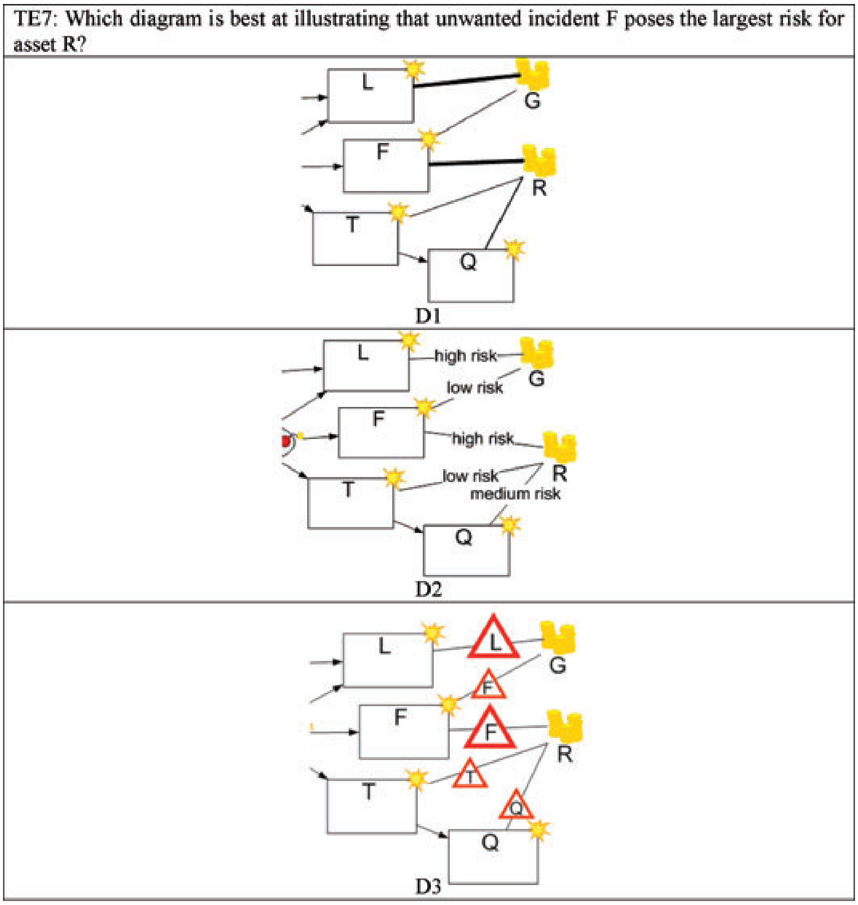
We investigated whether size, color, or textual information is better suited to represent the magnitude (severity) of an unwanted incident in TE3 (Fig. 9).

- D1: the most sever unwanted incidents have darker color than the less so.
- D2: the most sever unwanted incidents have larger symbols (boxes) than the less so.
- D3: the severity of the unwanted incidents is represented by textual information labels such as "High" and "Low."

In Task 7 (TE7), we investigated the following modeling alternatives (Fig. 10):

- D1: thick lines between unwanted incidents and assets are used to illustrate severe risks.
- D2: the severity of risks is illustrated by placing textual information labels such as "high risk" and "low risk" on the lines between unwanted incidents and assets.

[1] For a limited time the subject is shown a set of objects that must be memorized.



TE7: Which diagram is best at illustrating that unwanted incident F poses the largest risk for asset R?

Fig. 10. The diagram alternatives for "the magnitude of risks."

- D3: the size of a warning sign attached to the lines between unwanted incidents and assets is used to illustrate the severity of risks (the bigger the sign, the higher the severity).

5. EXPERIMENT DESIGN

We conducted an empirical study using professionals within the IT sector as subjects in order to decide upon the preferred notation. Similar representations had been tested in a prestudy using IT students as subjects.

5.1. Subjects

The survey was distributed via e-mail to people in various parts of the IT industry. The people receiving the survey were personal contacts of the main author of this article; partly professional contacts and partly contacts from her period as a student at the Norwegian University of Science and Technology. In all, 41 persons received the survey via e-mail and of these, 33 persons responded.

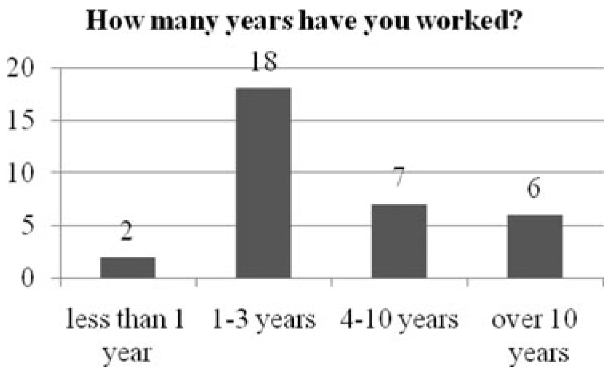


Fig. 11. The subjects' years of work experience.

The 33 subjects received no payment and participated voluntarily. The majority of the subjects were males, between the ages of 26 and 35 and with 1–3 years of work experience. The work experience of the subjects is shown in Fig. 11. The subjects were asked about their experience with tasks within the following categories: *system design, development,*

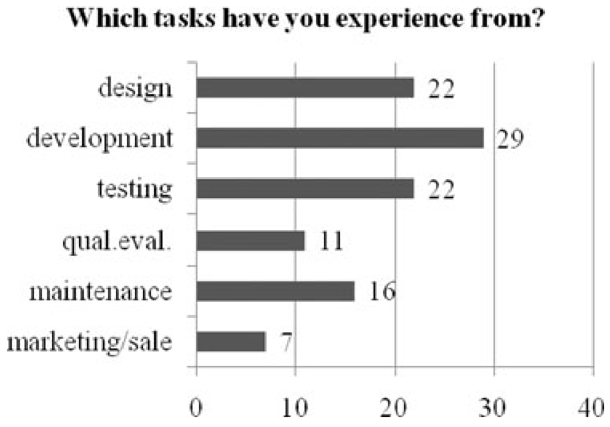


Fig. 12. The subjects' task experience.

testing, quality evaluation, maintenance, and *marketing/sales*. The subjects' experience in these categories of tasks is shown in Fig. 12.

5.2. Material

The material was in the form of a questionnaire where the subjects were asked to prioritize the different modeling alternatives. It was similar to the prestudy using IT students. (Complete versions are provided in a technical report[12] and the results are compared in Section 7.) It consisted of eight tasks and was estimated to take about 15–20 minutes to complete. The tasks were related to variations of the basic threat diagram shown in Fig. 13. The focus of each task varied, but the main issues were how to simplify graph navigation, how to best represent vulnerabilities, and how to highlight major risks and unwanted incidents. The explanation given in the material is shown in Fig. 14.

5.3. Hypotheses

The null hypothesis and the alternative hypothesis for the purpose of testing preference in each of the tests were:

- H_0: the modeling alternatives are equally preferred by the subjects.
- H_1: the modeling alternatives are not equally preferred.

In addition, we tested differences in preference between the subjects with short experience and the subjects with long experience. For these tests, we used the following null hypothesis and alternative hypothesis:

- H_0: the subjects with short and long experience have the same preferences.
- H_1: the subjects with short and long experience do not have the same preferences.

5.4. Analysis Method

The data were coded on a scale from –3 to +3, as shown in Fig. 15. The average score for each of the tasks could then be calculated, and the null hypothesis with respect to preference is then that the average equals zero. In order to test this hypothesis, we conducted a two-tailed Student's *t*-test[44] with significance level 0.95 for each task.[2]

In order to test the hypotheses related to differences in preference, we divided the subjects into two groups: the subjects with short work experience (0–3 years) and the subjects with long work experience (4+ years). For each task, we then conducted a two-tailed Mann-Whitney test[44] to see if there were any significant differences between the preferences of the subjects with short work experience and the subjects with long work experience.

All tests were made using the statistical package SPSS.[46]

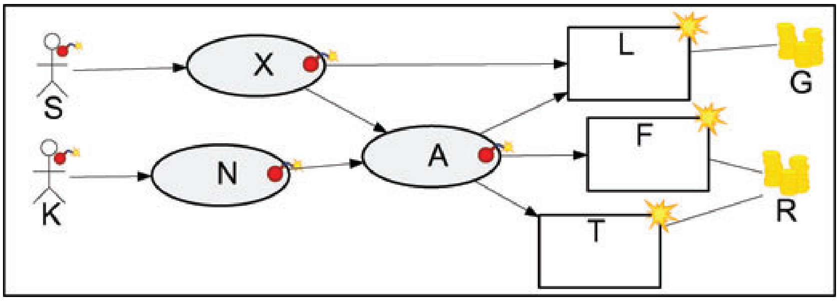
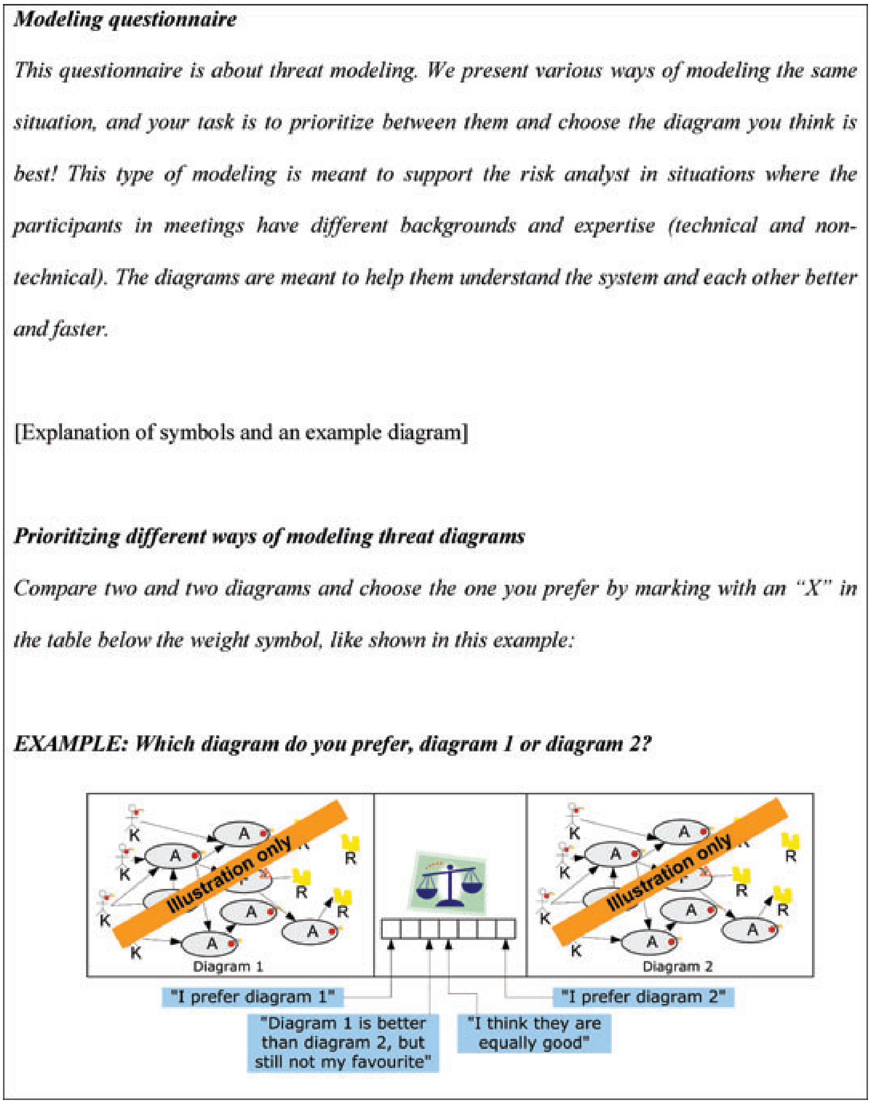


Fig. 13. The basic threat diagram.

[2] The same data were tested using Pearson's chi-square test[45] with significance level 0.95 in a technical report.[12] In these tests, the scale from –3 to +3 was not used, but instead a count of negative and positive numbers. However, these tests had the same results with respect to the testing of the hypotheses as the *t*-test reported in this article.



Modeling questionnaire

This questionnaire is about threat modeling. We present various ways of modeling the same situation, and your task is to prioritize between them and choose the diagram you think is best! This type of modeling is meant to support the risk analyst in situations where the participants in meetings have different backgrounds and expertise (technical and non-technical). The diagrams are meant to help them understand the system and each other better and faster.

[Explanation of symbols and an example diagram]

Prioritizing different ways of modeling threat diagrams

Compare two and two diagrams and choose the one you prefer by marking with an "X" in the table below the weight symbol, like shown in this example:

EXAMPLE: Which diagram do you prefer, diagram 1 or diagram 2?

Illustration only
Diagram 1
Illustration only
Diagram 2
K
A
R
"I prefer diagram 1"
"Diagram 1 is better than diagram 2, but still not my favourite"
"I think they are equally good"
"I prefer diagram 2"

Fig. 14. Task explanation.

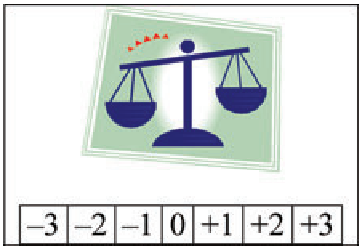


Fig. 15. Coding data.

6. RESULTS

The results from the statistical tests are presented below. The abbreviations used to refer to tasks and comparisons in the results are summarized in Table II. The *t*-tests are grouped in accordance to the three research questions posed at the end of Section 2. As we will see, there was a significant difference between the diagram alternatives in four of the six tasks. The Mann-Whitney tests are only presented in the cases where there was a significant difference between the subjects with short- and long-work experience. This is the case for two of the tests.

Tables III and IV show the results from representing graph navigation: the likelihood of paths (TE2) and logical AND (TE4).

The conclusion from the test of the likelihood of paths (TE2) was that neither of the modeling

Table II. Summary of Tasks and Comparisons

Task/Comparison	Summary
TE2	Task 2: Graph navigation
TE2D1D2	Dashed lines vs. line thickness
TE3	Task 3: Representing risk
TE3D1D2	Color vs. size
TE3D1D3	Color vs. text
TE3D2D3	Size vs. text
TE4	Task 4: Graph navigation
TE4D1D2	Gate vs. dashed lines
TE4D1D3	Gates vs. notes
TE4D2D3	Dashed lines vs. notes
TE5	Task 5: Representing vulnerabilities
TE5D1D2	Symbol vs. textual attributes
TE6	Task 6: Representing vulnerabilities
TE6D1D2	Textual attributes vs. symbol
TE7	Task 7: Representing risk
TE7D1D2	Line thickness vs. text
TE7D1D3	Line thickness vs. symbol and size
TE7D2D3	Text vs. symbol and size

Table III. Statistics for Graph Navigation (TE2, TE4)

t	*N*	Mean	Std. Deviation	Std. Error Mean
TE2D1D2	33	–3.3	2.3	0.40
TE4D1D2	33	–2.4	1.2	0.21
TE4D3D1	33	2.2	1.4	0.24
TE4D2D3	33	1.8	1.4	0.25

Table IV. Test Results for Graph Navigation (TE2, TE4)

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	*df*	Sig. (2-tailed)	Mean Difference	Lower	Upper
TE2D1D2	−0.84	32	0.41	−0.33	−1.1	0.47
TE4D1D2	−11.4	32	0.00	−2.4	−2.9	−2.0
TE4D3D1	9.2	32	0.00	2.2	1.7	2.7
TE4D2D3	7.3	32	0.00	1.8	1.3	2.3

alternatives was significantly preferred to the other (keep H_0). For logical AND (TE4), we found significant differences between all modeling alternatives: D1 is preferred over both D2 and D3, and D3 is preferred over D2 (reject H_0).

From Tables V and VI, we can see that in test TE4D2D3 there was a significant difference in the preferences of the subjects with short experience and the subjects with long experience (reject H_0). Those with short experience had a higher preference of D3 over D2 than those with long experience.

Table V. Statistics for Experience (TE4D2D3)

	Experience	*N*	Mean	Mean Rank	Sum of Ranks
TE4D2D3	Short	20	2.3	21	410
	Long	13	1.2	12	151
	Total	33	1.8		

Table VI. Test Results for Experience (TE4D2D3)

	TE4D2D3
Mann-Whitney U	60
Asymp. Sig. (2-tailed)	0.007

Table VII. Statistics for Vulnerabilities (TE5, TE6)

	N	Mean	Std. Deviation	Std. Error Mean
TE5D1D2	33	0.27	2.0	0.35
TE6D1D2	33	1.4	1.6	0.29

Table VIII. Test Results for Vulnerabilities (TE5, TE6)

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	*df*	Sig. (2-tailed)	Mean Difference	Lower	Upper
TE5D1D2	0.77	32	0.45	0.27	−0.45	0.99
TE6D1D2	4.8	32	0.00	1.4	0.78	1.9

Tables VII and VIII show the results from representing vulnerabilities: shared vulnerabilities (TE5) and threats & vulnerabilities (TE6).

The tests for representing vulnerabilities show that neither of the diagram alternatives for shared vulnerabilities (TE5) were significantly preferred to the other (keep H_0), while for threats & vulnerabilities (TE6), the representation in D2 is significantly preferred (reject H_0).

Tables IX and X show the results from representing the magnitude of risks (TE7) and unwanted incidents (TE3).

From test TE3 we can conclude that the modeling alternative in D3 is preferred to that of D1, but not significantly to that of D2 (reject H_0). The result

Table IX. Statistics for Unwanted Incident/Risk (TE3, TE7)

	N	Mean	Std. Deviation	Std. Error Mean
TE3D1D2	33	0.06	2.3	0.39
TE3D3D1	33	−0.94	2.0	0.35
TE3D2D3	33	0.42	2.2	0.38
TE7D1D2	33	1.0	1.9	0.33
TE7D3D1	33	0.12	2.1	0.36
TE7D2D3	33	−1.4	1.8	0.31

Table X. Test Results for Unwanted Incident/Risk (TE3, TE7)

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	*df*	Sig. (2-tailed)	Mean Difference	Lower	Upper
TE3D1D2	0.15	32	0.88	0.06	−0.74	0.86
TE3D3D1	−2.7	32	0.012	−0.94	−1.7	−0.23
TE3D2D3	1.1	32	0.27	0.42	−0.35	1.2
TE7D1D2	3.2	32	0.003	1.1	0.39	1.7
TE7D3D1	0.34	32	0.74	0.12	−0.61	0.86
TE7D2D3	−4.7	32	0.00	−1.4	−2.0	−0.80

Table XI. Statistics for Experience (TE7D3D1)

	Experience	*N*	Mean	Mean Rank	Sum of Ranks
TE7D3D1	Short	20	0.80	20	399
	Long	13	−0.92	12	162
	Total	33	0.12		

Table XII. Test Results for Experience (TE7D3D1)

	TE7D3D1
Mann-Whitney U	71
Asymp. Sig. (2-tailed)	0.027

for TE7 shows that D2 is preferred over both D1 and D3 (reject H_0).

From Tables XI and XII, we can see that in test TE7D3D1 there was a significant difference in the preferences of the subjects with short experience and the subjects with long experience (reject H_0). What we can see is that the subjects with short experience had a weak preference for D1, while the subjects with long experience had a weak preference for D3.

7. DISCUSSION OF MODELING PREFERENCES

In the following, the results for each task are discussed. The diagram alternatives are referred to as D1, D2, D3, meaning "diagram 1," "diagram 2," etc. in the relevant figures. Whenever the results from the prestudy are relevant, they are mentioned. However, since the material used in the experiment was quite different from that which was used in the prestudy, we cannot compare the results directly.

7.1. Representing Graph Navigation

The results from evaluating TE4 (Fig. 6) showed that D1 was preferred to D3, which again was preferred to D2. Both alternatives with text labels were preferred to the nontext alternative. The same result was found in the prestudy using similar representations. The conclusion is that the subjects prefer a combination of a closed contour and a textual information label. Besides the possibility of being known from other notations, the AND-gate shape used in D1 is clearer and more easily noticed than the other two alternatives. Using the AND symbol will also make the diagrams less cluttered compared to alternative D3. The results also show that the subjects with short experience had a stronger preference for D3 over D2 than the subjects with long experience. This may support the conclusion since it indicates that the less experienced subjects have a stronger preference toward text labels.

In TE2 (Fig. 5), the result showed that neither of the alternatives was preferred for this purpose. The same result was also obtained in the prestudy. It was a bit surprising that the thick line in D2 was not preferred over the thinner line in D1 since the modeling alternative is comparable to the use of thick lines in network maps to illustrate heavy traffic. A possible explanation for this is that the solidity of a line, in contrast to text, does not convey a unique interpretation. Even though this study cannot give a definite explanation of this result, we have, during field trials, found it more helpful to show likelihood of paths by annotating the threat scenarios with likelihood estimates (numbers or text). The unwanted incident likelihood can then be estimated on the basis of the likelihood estimates of the threat scenarios that cause the incident. This has led to the conclusion that we should use textual information labels to show likelihood of paths, rather than relying on graphical means only.

7.2. Representing Vulnerabilities

Vulnerabilities are weaknesses, flaws, or deficiencies of the analysis object that expose assets to threats. During field trials, we have experienced a need for representing vulnerabilities explicitly in the diagrams. Vulnerabilities may be modeled from several perspectives: in some situations, we are interested in specifying that several assets are subject to the same vulnerabilities, while in other cases, we like to see which vulnerabilities a given threat may exploit.

In TE5 (Fig. 7), the experimental results show that neither of the two representations was preferred to illustrate shared vulnerabilities. The prestudy showed a preference for D2. This task only dealt with assets and their vulnerabilities; therefore, we find it surprising that D2 was not preferred since this alternative had grouped all information concerning vulnerabilities below each asset. With respect to solving a task like TE5, it can be assumed that more effort is required to use D1 where the reader must deduce which vulnerabilities are relevant for which asset.

The same two diagram alternatives were also used in TE6 (Fig. 8), which aimed to identify the vulnerabilities a threat may exploit. This task was new in the experiment and therefore not tested in the prestudy. The result from TE6 showed that the alternative using the padlock symbol (D1) was significantly preferred over the alternative representation.

7.3. Representing Risks

We often experience that the concept *risk* can be difficult to understand due to its abstract nature. A risk is an unwanted incident that has been given specific likelihood and consequence estimates. To make the risk-concept less abstract, we investigated how it can be graphically represented in threat diagrams. In risk evaluation and treatment identification, we find it useful to be able to specify the magnitude of both risks and unwanted incidents.

The results for TE3 (Fig. 9) show a significant difference between the representations, and that D3 (textual information) was preferred to the other two alternatives. This result was also found in the prestudy. A possible explanation is that the textual information in this case had a unique interpretation, while size or color could have many interpretations.

It was quite surprising to see that the risk symbol alternative in TE7 (Fig. 10, D3) received the lowest score. The textual information alternative was preferred and was probably the representation that required the least effort to understand. It is important to consider the effects of giving size and color a specific semantic interpretation. Such a decision would mean that every time an element diverges from the standard size or color, the reader will expect it to mean something special, even if it was not the intention of the modeler. The degree of difference will also play an important role. In order for an element to stand out as different and to avoid confusion, the color or size of the element has to be sufficiently different from the other elements. The results show that the subjects with short experience had a weak preference for D1 over D3, while the subjects with long experience had a weak preference for D3 over D1. It is, however, difficult to find any good explanation for this.

7.4. Summary

The overall findings suggest that textual information in graphical models seems to be preferred over purely graphical means. In the already mentioned study of icons and text labels in the initial version of the CORAS language,[4] we compared a set of UML profile models to the standard UML models, both stereotyped with text labels (Fig. 16). We found that the subjects receiving the special symbol version of the material completed more tasks given limited time than the other group and concluded that the text label stereotyping was not particularly significant. In other words, the symbols helped the subjects identify the elements of the modeling language.

In the investigation presented in this article, on the other hand, we found that text labels were preferred over graphical means (size, color, line thickness) when it came to representing magnitudes (likelihood of path, severity of risk). Comparing the two experiments, we hypothesize that text labels used as categorization labels may quickly be ignored by the reader because they do not contain the amount of information needed to understand the meaning. Rather than categorizing elements with text labels, text labels should be used to communicate information that the reader otherwise has to seek in other documents.

8. THREATS TO VALIDITY

The main threats to validity of the empirical results from this investigation are described in this section. First of all, the diagrams we used were less

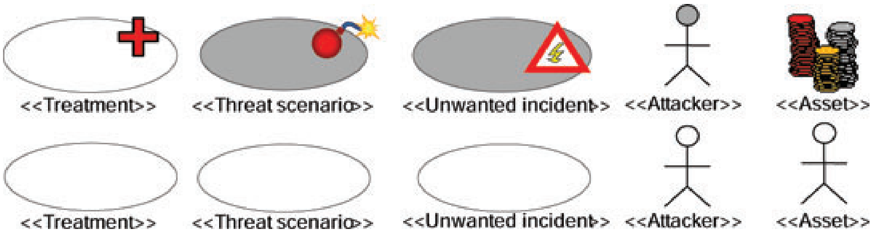


Fig. 16. Comparing stereotyping alternatives.

detailed and complex than real diagrams, but more than just examples of notation. Real risk and threat diagrams contain more descriptive text, and this additional text might make the text labels less visible. This weakness is difficult to avoid. On the other hand, our practical experience with risk analysis indicates that the parts of diagrams under focus at any given point in time are not very large. There are two reasons for this. First, large and complex diagrams become too difficult to work with and are usually divided into smaller and simpler diagrams. Second, discussions during a risk analysis session are usually concentrated on only a few scenarios at a time. As a result, the attention is most of the time on fragments of diagrams for which our results should be significant.

A further threat to validity may be that the symbols (padlock, warning sign, and logical gate) and color (gray) used in the experiment are not optimal for the information we were trying to visualize; in other words, that text labels were preferred because of bad choice of symbols and color. On the other hand, the symbols have not been randomly chosen, but are alternatives that have materialized through experience from our industrial field trials.

We might have obtained a different result if we had used a color signalizing "danger" to a larger extent than gray. It would have been interesting to see the result if we had tested colors like red, orange, and green; colors that are often used to illustrate high, medium, and low risk. Another threat to validity is that people may prefer what they are used to over things that are new, even if they might prefer the new representation after getting used to it. In this respect, text labels are probably more familiar than newly introduced symbols, and this may give the text labels an immediate higher preference.

The experiment addresses a selection of risk analysis modeling challenges. It would be both interesting and useful to also test other aspects and modeling alternatives, but in this case, we had to limit the size of the material to increase the likelihood of it being completed by the subjects.

The experiment had 33 respondents. With the sample size at hand we can only expect to identify medium to large effect sizes as statistically significant.[47] Even though smaller effect sizes may be of possible theoretical interest, they are less relevant for the development of the CORAS language. For the purpose of making design decisions about the language, we are mainly interested in differences associated with large effect sizes.

The various modeling alternatives were tested on a sample of subjects that is quite homogenous with respect to background, age, and sex. The subjects received no introduction to risk analysis other than the short background provided with the material. We do not know whether the results would have been different if the subjects had been familiar with risk analysis. To some degree this does prevent us from generalizing. However, the technical background of our subjects and their lack of experience in risk analysis are not uncommon for participants in IT-related risk analyses. For exactly this reason, targeting IT professionals was one of the goals of this study. Although we cannot know whether our results are valid for the general population, our subjects should be considered representative of large parts of the relevant population, that is, the kind of people that are likely to be involved in IT-related risk analyses.

Since the material of the experiment was distributed by e-mail, we could not control how the subjects chose to fill in the questionnaire (helping facilities, time spent, etc.). We do not believe this affects the overall validity of the results since the subjects were asked to choose the alternatives they *preferred*, that is, a completely subjective measure.

9. CONCLUSIONS

In risk analyses, it is often helpful to draw simple diagrams to communicate or discuss risk scenarios. It is essential that these diagrams are easily understood by people without special training and experience in modeling or risk analysis. A typical situation is a structured brainstorming session involving stakeholders with different backgrounds and competences. This article addresses issues related to the following question: To what extent and in what way may we improve the reading and comprehension of such diagrams using text labels and simple graphical means?

Our empirical investigation indicates that mechanisms like size and color coding used to convey particular information, such as the severity of risks, in graphical models are less preferred by the subjects as compared to textual information labels. The size or color of an element does not in general convey a unique interpretation in a model, while textual information is more specific. The subjects tend to prefer the representations where they get the most information without having to interpret any additional graphical means. An earlier study(4) showed that graphical symbols help users parse diagrams. Taken together, these studies points in the direction that graphical means are useful for categorization of elements, while the magnitude of elements, which may be seen as additional information, are best conveyed using textual labels.

The results of the investigation reported on in this article had a direct impact on the development of the CORAS language.(48) In particular we discarded any ideas about using size, color, or dashing, and thickness of lines to convey information in our diagrams, and decided to use text labels to specify likelihoods, consequences, and risk levels. Further, we decided to use the padlock symbol to represent vulnerabilities instead of listing them as attributes to assets.

9.1. Conclusions Regarding "Representing Graph Navigation"

To assess the various threats in risk analyses, it is useful to describe the paths via which they are most likely to harm the assets. Our study showed no significant benefit of using either thick or dashed lines for highlighting (attracting attention to) or deemphasizing (detracting attention from) the more or less likely paths.

We investigated various alternatives to represent the logical AND gate. The conclusion was clear: the classical logical AND-gate symbol annotated with the text label "and" was preferred.

To summarize: the only alternative that showed significant preference was the gate symbol.

9.2. Conclusions Regarding "Representing Vulnerabilities"

In our study, we saw that representing vulnerabilities is most simply done by placing them in the paths from threats to assets. However, when looking at one particular asset and its vulnerabilities, there is a tendency toward a preference of having the vulnerabilities listed under the asset. Since it may be interesting to analyze vulnerabilities and assets independently of the threat scenarios, we suggest that a specialized *asset diagram* where the vulnerabilities are attached to assets may be useful.

To summarize, if the purpose is to show the relation between assets and vulnerabilities, the threats, threat scenarios, and paths between these are not as important as when the threat's preferred strategy is to be analyzed. Since we are interested in both views (threats and which vulnerabilities they may exploit, as well as assets and their vulnerabilities) it might be useful to employ different kinds of diagrams allowing the vulnerabilities to be viewed from two different perspectives.

9.3. Conclusions Regarding "Representing Risk"

An important aspect of a risk analysis is to identify the risks with the highest risk levels. It should therefore be possible to express the magnitude, or severity, of both risks and unwanted incidents. We found textual information to be preferred over both size and color. We believe this is because text can carry a more precise and unique interpretation than an element of varying size and color. Text labels should not be used as categorization labels in a diagram if they do not convey important information needed to understand the diagram. Rather than categorizing elements with text labels, text labels should be used to communicate diagram-specific information that the reader otherwise has to seek in other documents.

To summarize, textual information labels should be used to indicate the severity of a risk or an unwanted incident since an element's size or color is too ambiguous.

9.4. Future Work

As explained earlier, the investigation on which this article reports has provided important input to the development of the CORAS language. After a period with experimentation, the language was stabilized and formalized. The language has also been extended with rules for calculating likelihoods, support for hierarchical diagrams, legal risk analysis, and dependency analysis. In addition, a diagram editor has been developed.[49]

A larger challenge, however, is the study of how the language works in practice. Controlled experiments like the one reported on in this article are useful for getting reliable information, but are limited with respect to the questions we are able to ask. Field trials, on the other hand, provide useful experience from the real-life use of the language, but the evidence from field experiments tends to be anecdotal in nature. An important line of further work would therefore be to design and carry out field trials in the form of case studies from which reliable evidence on the practical use of the language can be derived.

ACKNOWLEDGMENTS

The work on which this article reports has been funded by the Research Council of Norway through the projects SECURIS (152839/220) and DIGIT (180052/S10). We are grateful to Jan Heim and Asbjørn Følstad for their help and advice on the statistical analysis.

REFERENCES

1. den Braber F, Hogganvik I, Lund MS, Stølen K, Vraalsen F. Model-based security analysis in seven steps: A guided tour to the CORAS method. BT Technology Journal, 2007; 25(1):101–117.
2. den Braber F, Mildal AB, Nes J, Stølen K, Vraalsen F. Experiences from using the CORAS methodology to analyze a web application. Journal of Cases on Information Technology, 2005; 7(3):110–130.
3. Hogganvik I. A Graphical Approach to Security Risk Analysis [dissertation]. University of Oslo, 2007.
4. Hogganvik I, Stølen K. On the comprehension of security risk scenarios. Pp. 115–124 in: Proceedings of the 13th International Workshop on Program Comprehension (IWPC'05), 2005.
5. Hogganvik I, Stølen K. Risk analysis terminology for IT-systems: Does it match intuition? Pp. 13–23 in: Proceedings of the International Symposium on Empirical Software Engineering (ISESE'05), 2005.
6. Ware C. Information Visualization: Perception for Design (2nd Ed). San Francisco, CA: Elsevier, 2004.
7. Object Management Group. Unified Modeling Language (UML): Superstructure, version 2.0, 2005.
8. International Organization for Standardization, International Electrotechnical Commission. ISO/IEC13335: Information Technology: Guidelines for the Management of IT Security (Part 3), 1998.
9. International Organization for Standardization, International Electrotechnical Commission. ISO/IEC13335: Information Technology: Security Techniques: Management of Information and Communications Technology Security (Part 1), 2004.
10. Standards Australia, Standards New Zealand. AS/NZS4360: Australian/New Zealand Standard for Risk Management, 2004.
11. Standards Australia, Standards New Zealand. HB231: Information Security Risk Management Guidelines, 2004.
12. Hogganvik I, Stølen K. Investigating Preferences in Graphical Risk Modeling. Oslo: SINTEF ICT, 2007. Report No.: A57.
13. Houmb SH, den Braber F, Lund MS, Stølen K. Towards a UML profile for model-based risk assessment. Pp. 79–91 in Proceedings of UML 2002 Satellite Workshop on Critical Systems Development with UML (CSD-UML'02), Munich University of Technology, 2002.
14. Lund MS, den Braber F, Stølen K, Vraalsen F. A UML Profile for the Identification and Analysis of Security Risks During Structured Brainstorming. Oslo: SINTEF ICT, 2004. Report No.: STF40 A03067.
15. Ibrekk H, Morgan G. Graphical communication of uncertain quantities to nontechnical people. Risk Analysis, 1987; 7(4):519–529.
16. Connelly AN, Knuth BA. Evaluating risk communication: Examining target audience perception about four presentation formats for fish consumption health advisory information. Risk Analysis, 1998; 18(5):649–659.
17. Lipkus MI, Hollands JG. The visual communication of risk. Journal of the National Cancer Institute. Monographs, 1999; 25:149–163.
18. Winn W. An account of how readers search for information in diagrams. Contempory Education Psychology, 1993; 18:162–185.
19. Bratthall L, Wohlin C. Is it possible to decorate graphical software design and architecture models with qualitative information? An experiment. IEEE Transactions on Software Engineering, 2002; 28(12):1181–1193.
20. Linos PK, Aubet P, Dumas L, Helleboid Y, Lejeune D, Tulula P. Visualizing program dependencies: An experimental study. Software Practice and Experience, 1994; 24(4):387–403.
21. Larkin JH, Simon HA. Why a diagram is (sometimes) worth ten thousand words. Cognitive Science, 1987; 11:65–99.
22. Treisman A, Gormican S. Feature analysis in early vision: Evidence from search asymmetries. Psychological Review, 1988; 95(1):15–48.
23. Becker RA, Eick SG, Wilks AR. Visualizing network data. IEEE Transactions on Visual Computer Graphics, 1995; 1(1):16–21.
24. Wertheimer M. Laws of organization in perceptual forms. [English translation of: Untersuchungen zur Lehre von der Gestalt, II. Psychol Forsch. 1923; 4: 301–350]. Pp. 71–88 in Ellis WD (ed). A Source Book of Gestalt Psychology. London: Routledge & Kegan Paul, 1938.
25. Winn W. Perceptual strategies used with flow diagrams having normal and unanticipated formats. Perceptual and Motor Skills, 1983; 57:751–762.
26. Winn W. The role of diagrammatic representation in learning sequences, identification, and classification as a function of verbal and spatial ability. Journal of Research in Science Teaching, 1982; 19:79–89.
27. Winn W, Solomon C. The effect of the rhetorical structure of diagrams on the interpretation of simple sentences. Unpublished manuscript, 1991. University of Washington.
28. Chattratichart J, Kuljis J. An assessment of visual representations for the "flow of control." Pp. 45–48 in Proceedings of the 12th Workshop of the Psychology of Programming Interest Group (PPIG'00). Cosenza, Italy, 2000.
29. International Electrotechnical Commission. IEC61025: Fault Tree Analysis (FTA), 1990.
30. Sedra AS, Smith KC. Microelectronic Circuits. New York: Oxford University Press, 2003.

31. Krogstie J. Conceptual Modeling for Computerized Information Systems Support in Organizations [dissertation]. Norwegian Institute of Technology. University of Trondheim, 1995.
32. Goodman N. Languages of Art: An Approach to a Theory of Symbols. Indianapolis, IN: Hackett, 1976.
33. Cahill MC, Carter RCJ. Color code size for searching displays of different density. Human Factors, 1976; 18(3):273–280.
34. Christ RE. Review and analysis of color coding research for visual displays. Human Factors, 1975; 17(6):542–570.
35. Cleveland WS, McGill R. Graphical perception and graphical methods for analyzing scientific data. Science, 1985; 229:828–833.
36. Shneiderman B. Designing the User Interface. Reading, MA: Addison-Wesley, 1992.
37. Wickens CD. Engineering Psychology and Human Performance (2nd Ed). New York: HarperCollins, 1992.
38. Christ RE. Research for evaluating visual display codes: An emphasis on colour coding. Pp. 209–228 in Easterby R, Zwaga H (eds). Information Design: The Design and Evaluation of Signs and Printed Material. Chichester: John Wiley and Sons, 1984.
39. Jubis RMT. Coding effects on performance in a process control task with uniparameter and multiparameter displays. Human Factors, 1990; 32(3):287–297.
40. Smith L, Thomas D. Color versus shape coding in information displays. Journal of Applied Psychology, 1964; 48(3):137–146.
41. Cavanagh JP. Relationship between the immediate memory span and the memory search rate. Psychological Review, 1972; 79:525–530.
42. Schneider W, Shiffren RM. Controlled and automatic human information processing I: Detection, search, and attention. Psychological Review, 1977; 84:1–66.
43. Tan KC. Effects of Stimulus Class on Short-Term Memory Workload in Complex Information Displays [disseration]. Virginia Technical University, 1990.
44. Bhattacharyya GK, Johnson RA. Statistical Concepts and Methods. New York: John Wiley & Sons, 1977.
45. Siegel S, Castellan J. Non-Parametric Statistics for the Behavioural Sciences (2nd Ed). New York: McGraw-Hill International Editions, 1988.
46. SPSS Statistics. Available from: http://www.spss.com/statistics/, Accessed September 24, 2010.
47. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd Ed). New York: Psychology Press Taylor & Francis Group, 1988.
48. Hogganvik I, Stølen K. A graphical approach to risk identification, motivated by empirical investigations. Pp. 574–588 in Proceedings of the 9th International Conference on Model Driven Engineering Languages and Systems (MoDELS'06). Springer, 2006. (Lecture Notes in Computer Science; vol. 4199).
49. Lund MS, Solhaug B, Stølen K. Model-driven risk analysis. In The CORAS Approach. Berlin/Heidelberg: Springer, 2011.