Abstract—In this study, two experiments are reported which investigated the relative importance of five image based factors and training on threat detection performance of human operators in x-ray screening. Experiment 1 was based on a random sample of about 16,000 records of threat image projection (TIP) data. TIP is a software function available on state-of-the-art x-ray screening equipment which allows projecting fictional threat images (FTIs) into x-ray images of passenger bags during the routine baggage screening operation. Analysis of main effects showed that image-based factors can substantially affect screener detection performance in terms of the hit rate (identification of FTIs). There were strong effects of FTI view difficulty (rotation of FTIs) and superposition of FTIs by other objects in the x-ray image of a passenger bag. The amount of opaque areas in the x-ray image of a passenger bag (Opacity) had a small although significant effect on detection performance. Clutter and bag size were not significant.

Experiment 2 was conducted using an offline-test in order to examine main effects in combination, their interactions and effects of training. The following image-based factors were varied systematically: Threat (FTI) category (guns, knives, improvised explosive devices, other threats), view difficulty, superposition, bag complexity (a combination of opacity and clutter) and bag size. Data were collected from 200 screening officers at five sites across Europe that have a specific computer-based training package installed. Consistent with the results obtained in Experiment 1, there were large main effects of threat (FTI) category, view difficulty, and superposition. Again consistent with Experiment 1, effects of bag complexity (opacity and clutter) and bag size were much smaller. In addition to Experiment 1, the number of computer based training (CBT) hours was available for each security officer participating in the study. Training turned out to be a key driver to improving threat detection performance in x-ray screening and seemed to mediate the effects of some image based factors.

Possible implications regarding the enhancement of human-machine interaction in x-ray screening are discussed.

I. INTRODUCTION

Screening passenger bags for threat items using state-of-the-art x-ray machines is an essential component of airport security. Previous work (Schwaninger, 2003b, Schwaninger, Hardmeier, & Hofer, 2005, and Schwaninger, Michel, & Bolfing, 2007) has identified image-based factors that affect human performance in x-ray screening tasks: object view difficulty, superposition by other objects and bag complexity (opacity and clutter). Recently, the question has been raised whether bag size would be another image-based factor that affects human detection performance when visually inspecting x-ray images of passenger bags. In this study we determined effects and interactions of image based factors and human factors (amount of recurrent computer-based training). In addition, this study provided the basis for scientifically based conclusions about the significance of the bag size variable by itself and in relation with other variables.

Two experiments are reported. Experiment 1 is based on threat image projection (TIP) data ensuring high ecological validity. Experiment 2 is based on an off-line computer based test, which allows investigating the effects of image-based factors in combination, effects of training and their interactions.

A. Image Based Factors

Schwaninger (2003b) and Schwaninger, Hardmeier, and Hofer (2005) have identified three image based factors which affect threat detection by x-ray screeners significantly: view difficulty, superposition, and bag complexity (see figure 1).

![Fig. 1. Illustration of the three basic image based factors suggested by Schwaninger (2003b) and Schwaninger, Hardmeier, and Hofer (2005)](image-url)
ematically implemented (Schwaninger, Michel, & Bolfing, 2007, see Bolfing, & Schwaninger, 2007 for the latest version). View difficulty is implemented as a statistically calculable value between 0 an 1 named FTI view difficulty. Superposition and bag complexity are implemented as image processing measurements whereby bag complexity is split up into clutter and opacity. The introduction of the image based factor bag size in this study necessitated normalization of earlier implementations of clutter and opacity regarding bag size. Formulae and short descriptions of the underlying concepts are specified in Bolfing, & Schwaninger (2007).

II. THREAT IMAGE PROJECTION (TIP) χ² ANALYSIS: EXPERIMENT 1

A. Method

1) Threat Image Projection (TIP) Data: In order to ensure a high ecological validity of the results, we have decided to analyze data from threat image projection (TIP). TIP is a software function of state-of-the-art x-ray screening equipment used at security checkpoints in airports, nuclear power plants, navigation docks etc. In aviation security it is distinguished between cabin baggage screening (CBS) and hold baggage screening (HBS). In CBS, guns, knives and improvised explosive devices (IEDs) and other threats are subject of identification and confiscation. In HBS, the focus rests mainly on IEDs and dangerous goods such for example gasoline containers, diver lamps, etc. The current investigation is confined to cabin baggage screening. In CBS TIP, fictional threat items (FTIs) are projected into x-ray images of passenger bags during the routine baggage screening operation. Given a minimum number of TIP events in order to have available reliable data, this allows measuring detection performance of human operators (x-ray screeners) on-the-job (Hofer & Schwaninger, 2005) and thus with high ecological validity.

The data basis of this study consists of a random sample of 16'329 TIP events which have been recorded at a large European airport with about 700 professional x-ray screeners of 16'329 TIP events which have been recorded at a large airport with about 700 professional x-ray screeners.

2) χ² Analysis: To compare the effects of the independent variables FTI view difficulty, superposition, opacity, clutter and bag size on detection performance, the following procedures were applied to the TIP data described above. A histogram was created for each independent variable (image based factor) with the independent variable on the x-axis and its frequency on the y-axis. The upper and lower 2.5% each of the cases in the data were excluded to remove outlier data from the analysis as well as to be able to build five equidistant bins with at least 100 data points each (TIP events).

Relative frequencies of hits were calculated for each of the five equidistant bins to run χ² tests with the null hypothesis H0 that the hit rates are equal across bins. Effect size analysis based on Cohen (1988) was used to compare the effect sizes of the different independent variables. For detailed information on χ² statistics see for example Coolican (2004).

B. Results

The results below are listed separately for each image based factor introduced above (see Bolfing, & Schwaninger, 2007 for further information and formulae). Each of the following subsections opens up with a graphical illustration of the image based factors’ effects on the threat detection performance measure hit rate. The x-axes show the five bins into which the whole data range was subdivided, starting with low values on the left. The y-axes show the hit rates against the image based factors’ bins. For reasons of confidentiality hit rates are not given explicitly, but the hit rate scales are reasonably chosen and kept constant throughout the whole document.

Following the graphical illustrations (figures 2-6), statistical test values are given in tables I-V. Basically, χ² statistics can be interpreted as follows: The larger the χ² (df, N) value the larger the effect. Additionally also χ² effect sizes w are given. Again, the larger the effect size, the larger the effect. Nevertheless, be aware that χ² and w values do not state the direction of the effect.

A summary bar plot graphics illustrating the impact of the five image based factors on the hit rate relative to each other is provided at the end of this results section (see figure 7). The bar plots show the χ² statistics effect sizes. The image based factors are arranged such that their effects decrease in size.

1) FTI View Difficulty: Figure 2 illustrates the large impact of FTI view difficulty on human detection performance in terms of hit rate. This is partly due to the fact that objects are more difficult when depicted from an unusual viewpoint (see figure 1). Other factors are related to the threat category and training of human operators (see Experiment 2).

Fig. 2. Illustration of the impact of FTI view difficulty on hit rate.
2) **Superposition:** Figure 3 illustrates the large effect of superposition on detection performance.

3) **Opacity:** Figure 4 shows the significant but relatively small influence of opacity on detection performance in terms of hit rate.

Here the question arises whether opacity as a perceptual concept does not explain much variance in threat detection performance or whether the image measurement formula of opacity does not measure what the concept is about.

4) **Clutter:** Figure 5 illustrates the small influence of clutter on hit rate. As for opacity, the question arises whether it is the concept of clutter that fails in predicting hit rates in TIP or whether the computational implementation needs to be improved.

5) **Bag Size:** Figure 6 shows the effect of bag size on hit rate in TIP.

The effect of bag size is not linear, i.e. based on these first results it can not be concluded that the larger bags are the lower the hit rate is.
6) **Comparison of the \( \chi^2 \) Effect Sizes:** In figure 7, the effect sizes \( w \) are compared. The factor FTI view difficulty has the highest effect size with \( w = .12 \), whereas clutter shows the lowest effect size with \( w = .01 \). The factors opacity, bag size and clutter show small effect sizes, while the effects of clutter and bag size did not reach statistical significance.

![Chi Square Effect Size Comparison](image)

**Fig. 7.** Comparison of the effect sizes among the image based factor.

### C. Discussion

The results obtained in Experiment 1 are consistent with earlier findings. Schwaninger, Hardmeier, and Hofer (2005) found that viewpoint, superposition and bag complexity affects screener performance. Schwaninger, Michel, and Bolfing (2007) could replicate these results. Using similar image measurements as in Experiment 1, they revealed similar effects sizes regarding the effects of FTI view difficulty, superposition, opacity (negatively correlated with transparency in Schwaninger et al., 2007) and clutter. However, several caveats are necessary to qualify the appropriateness of the results obtained in Experiment 1. Firstly, an analysis of auto-archive bags indicated that as would be anticipated, it is likely that TIP aborts are selectively eliminating certain bags (e.g. small bags rather than large bags) from the TIP image set, and thus reducing their presence. Secondly, it is not always clear how closely aligned TIP scores are to the specific operational situations encountered when threats are deliberately hidden in difficult bags. Most importantly, in Experiment 1 only main effects were analyzed. In order to gain a more complete picture, it is important to conduct a more controlled experiment, in which main effects in combination and their interactions can be measured reliably. This was conducted in Experiment 2.

### III. Off-line Computer Based Test: Experiment 2

#### A. Method

1) **Participants:** In Experiment 2, 200 X-ray screeners from five European sites with varying amount of training in x-ray image interpretation participated.

2) **Stimuli:** The stimuli were 1024 complete threat images (CTIs) and 1024 complete non-threat images (CNTIs). CTIs were created by projecting fictional threat items (FTIs) into 1024 X-ray images of bags. FTIs for the study were eight visually similar pairs of each of four types of threat item: guns, knives, improvised explosive devices (IEDs), and ‘other’ threats. Images of cabin baggage were captured from x-ray machines at a single European airport using the auto-archive function. The images were revised by three airport security supervisors to remove inappropriate images (e.g. images containing more than one bag, images containing incomplete bags, bags containing prohibited items or liquids, etcetera). This procedure resulted in 7606 bag images. Additional review by the QinetiQ team resulted in a total of 6659 bag images from which to draw the 1024 bags needed for the study. The final 1024 bags used for the study were chosen through a process of projecting the relevant FTIs into the bags such that the variables of interest were orthogonal in the stimulus set. Several of the full image sets of 2048 images (the 1024 images containing the FTIs, and also the same images without FTIs) were created, and the one with the most desirable properties in terms of variable orthogonality was chosen as the one to use in the study.

3) **Design:** The study employed a 4 (FTI category: guns, knives, IEDs, other) x 2 (view difficulty: easy, difficult) x 2 (superposition: low, high) x 2 (bag complexity: low, high) x 2 (bag size: small, large) x 2 (image type: FTI, no FTI) within-participants design. Since there were 16 FTIs in each category, this design results in 16x4x2x2x2x2 = 2048 images in total to be presented to the screeners. These images were shown to screeners in a random order in multiple testing sessions of 20 minutes each. As dependent variable the detection performance measure \( d^* \) (Green & Swets, 1966) was used. This measure provides a more valid estimate of detection performance because it takes the hit rate and the false alarm rate into account (see Hofer & Schwaninger, 2004 for different measures of x-ray detection performance). Since in the off-line test each bag was shown once with a threat and once without, accurate measurements of hit and false alarm rates could be obtained.

#### B. Results

Data form from the Type 1 sites were analysed in two ways. Firstly, by treating the variables FTI/view difficulty, superposition, opacity, clutter and bag size as continuous, a linear regression was employed to assess the main effects of each variable separately, and a multiple linear regression was used to examine the main effects together. Then, in order to examine interactions between the variables, and also to examine the role of an additional human-factor (amount of hours training on CBT), the discrete variables defined in Section 2.1.3 were used in an analysis of covariance (ANCOVA), with training hours as a covariate. Figure 3.2 shows the way in which these continuous and discrete variables are related to each other.

Data were analyzed in two ways. Firstly, by treating the
variables FTI view difficulty, superposition, opacity, clutter, and bag size as continuous, a linear regression was employed to assess the main effects of each image based factor on threat detection performance separately, and a multiple linear regression was used to examine the main effects together. Additionally, we applied another linear regression with hours of recurrent computer based training prior to the testing as its predictor. Then, in order to examine main effects as well as interactions between the variables, the discrete variables FTI category, view difficulty, superposition, bag complexity and bag size were used in an analysis of covariance (ANCOVA). Training hours served as covariate in the ANCOVA. Figure 8 shows the way in which these continuous and discrete variables are related to each other. Due to a high inter-correlation and a test design that demands independence of its variables, opacity and clutter were encoded into the single discrete variable bag complexity. FTI category and view difficulty were encoded into a single continuous variable because it is not sensible to encode either variable directly into a continuous variable. Instead we defined the variable FTI view difficulty as the difficulty - as measured in threat detection performance (\(d'\)) - screening officers had in solving a specific threat item in a specific view (easy or difficult) across all other conditions (i.e. superposition, bag complexity and bag size).

![Image of variable encoding](image)

**Fig. 8. Illustration of relationship between discrete and continuous representations of variables**

1) **Linear Regression and Multiple Linear Regression:** The regression analyses will help us understand the relationship between image based factors and \(d'\), as well as training hours and \(d'\). Figure 9 shows the relative effect sizes, the absolute values of the correlations with the dependent variable \(d'\), for the individual variables. For superposition and training hours a logarithmic transformation was applied, so as to achieve linearity in their relationship with detection performance \(d'\). With .70, .63 and .58, FTI view difficulty, training hours and superposition all have very high effect sizes. Opacity has a moderate to small effect size with .22, clutter and bag size have very small effect sizes with .05 and .07, respectively. Except for clutter, all correlations are statistically significant.

Figures 10 and 11 show the results of the multiple linear regression with all image based factors: FTI view difficulty, superposition (logarithmically transformed), opacity, clutter and bag size. The absolute values of the Beta coefficients in figure 11 indicate the relative importance of the different image-based factors when acting in combination. The column Sig. indicates whether a factor is significant (values < .05). As can be seen in figure 10 all image-based factors together can explain human detection performance with an effect size of \(R = 0.77\). Interestingly, the effect size of the only human factor analyzed (hours of recurrent computer based training), with \(R = 0.63\), is almost as large as the effect of all image-based factors together. As shown in figure 10 adding bag size to the linear model only leads to a minimal increase of its effect size from \(R = 0.772\) to \(R = 0.773\).

![Image of multiple linear regression](image)

**Fig. 9. Illustration of effect sizes R**

**Fig. 10. Multiple linear regression overview**

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.77</td>
<td>.597</td>
<td>.595</td>
<td>690</td>
</tr>
</tbody>
</table>

2) **ANCOVA:** A repeated measures analysis of covariance (ANCOVA) was conducted to analyze the main effects of image based factors, their interactions and the interaction with training. As can be seen in the main effects summary of figure 13 the repeated measures ANCOVA leads to only a slightly different pattern with regards to effect sizes than the linear regression analyses. These differences are due to the fact that, in contrast to the linear regression models, in the ANCOVA analysis effects of the covariate training hours are isolated from the effects of image based factors, and inter-individual differences between screening officers ('screener variance') is taken into account. Superposition shows the largest effect size (\(\eta^2\)), followed by FTI category, bag complexity and view difficulty. The main effect of bag size is
clearly smaller than the main effect of any other image based factor. Training hours has noteworthy interactions with FTI category and view difficulty. These interactions make sense, since we know from other studies that training can lead to comparatively larger performance increases for items that are comparatively difficult for novices (Koller, Hardmeier, Michel, & Schwaninger, in press). For example such is the case for improvised explosive devices (threat item category) or difficult views (view difficulty). There is also a small interaction of training with bag size, indicating that well trained screening officers are less affected by effects of bag size. Figure 14 gives an overview of the 10 largest interactions in the ANCOVA. All in all over 30 interactions reached statistical significance. Since the effect sizes of most interactions are very small we decided only to report interactions $\eta^2 \geq 0.07$. The interaction of view difficulty with threat category is explicable by the fact that detection performance of improvised explosive devices - unlike guns or knives - is largely independent of viewpoint. The interaction of superposition with view difficulty indicates that with difficult viewpoints superposition plays a larger role in determining detection performance than in easy views. The interaction of superposition with threat category indicates that some threat item categories are more sensitive to superposition than others. For example we know that superposition effects are higher with knives than with guns.

C. Discussion

With an overall correlation of .77 the linear modeling of detection performance with image based factors has a very high explanatory power. Superposition - though not always with the largest effect size - has shown the most robust effects on detection performance. Interestingly and in contrast to what one might have expected based on the results of the regression analyses, the variable bag complexity - a combination of opacity and clutter - showed a quite high effect size in the ANCOVA. Apart from that, the ANCOVA results reflect the regression analysis results nicely - both in main effects and interactions. Threat category and view difficulty had considerable interactions with the covariate training hours, testament to the fact that training is particularly effective for difficult items in those cases (i.e. for IEDs and for difficult views). Bag size, although intuitively plausible, turned out to play only a minor role in determining threat detection performance, together with clutter.

IV. General Discussion

There were large main effects of view difficulty and of FTI difficulty in all of the analyses, as expected. The same was true of superposition and complexity (opacity to a bigger extent than clutter). Clearly, any future work on image based factors that predict threat detection performance need to take account of these factors. When looking at detection performance overall (i.e. a multiple linear regression including the image based factors in combination but without their interactions), there is no effect of bag size. When using a more sophisticated model of data analysis including main effects of FTI view difficulty, superposition, opacity, clutter, bag size and the interactions of these variables, there is a small effect of bag size. In Experiment 2 we were able to examine the effect of the number of CBT training hours on threat detection performance. The key finding from the study is that the effect size for this variable was large, and seemed to mediate the effect of some image based factors on threat detection. Clearly, training is a key driver to improving threat detection performance in X-ray screening, and more work needs to be done to establish exactly what image based factors screeners need to be trained in to give the best improvements in threat detection accuracy.
V. RECOMMENDATIONS FOR IMPROVING HUMAN-MACHINE INTERACTION IN X-RAY SCREENING

A. FTI View Difficulty and Superposition

The factor FTI view difficulty refers to the fact that the identification of threat objects, as objects in general, is highly dependent on their viewpoint as well as on properties of the very object itself. Current x-ray screening equipment provides only one x-ray image of a passenger bag. More recent technology provides multiple views of a bag. Figure 15 illustrates how such new systems might be able to reduce the detection problems due to view difficulty and superposition. Objects that are superimposed by other objects from one perspective may be clearly visible from others. Additionally, our ANCOVA analysis has supported earlier findings showing that training improves detection performance particularly for difficult views (Koller, Hardmeier, Michel, & Schwaninger, in press).

Fig. 15. Illustrative example of how multi-view systems can help improving detection performance in spite of undesirable view difficulty and superposition effects.

B. Opacity

The image based factor Opacity refers to the amount of opaque areas in an x-ray image. X-ray systems with higher penetration have the potential to reduce detection problems due to opacity. In addition, it is possible to implement image measurement algorithms in x-ray equipments that warn the human operator (x-ray screener) with a “dark alarm” triggered by opaque areas. Manual search would follow when a dark alarm was indicated.

C. Screener Selection and Training

A very important approach to face these problems consists in screener selection and screener training. Psychological literature provides evidence that figure ground segregation (related to superposition) as well as mental rotation (related to view difficulty) are visual abilities which are fairly stable within a person. For example Hofer, Hardmeier, & Schwaninger (2006) and Hardmeier, Hofer, and Schwaninger (2006a) have shown that using computer-based object recognition tests in a pre-employment assessment procedure can help increasing detection performance of screeners substantially. In addition to stable abilities and aptitudes, there are several aspects of visual knowledge relevant to x-ray image interpretation. Knowledge based factors have been referred to knowing which objects are prohibited and what they look like in x-ray images of passenger bags. Computer-based training can be a powerful tool to improve x-ray image interpretation competency of screeners (e.g. Koller, Michel, Hardmeier, & Schwaninger, in press; Schwaninger, Hofer & Wetter, 2007; Ghylin, Drury, & Schwaninger, 2006).

ACKNOWLEDGMENT

This research was supported by the UK Department for Transport, the European Civil Aviation Conference Technical Task Force (ECAC TTF), and by the European Commission Leonardo da Vinci Programme (VIA Project, DE/06/C/7/TH-80403). Thanks to Zurich State Police, Airport Division and all other security organizations, companies and airports that supported this study by supplying screeners and data. Special thanks to Olive Emil Wetter and Jonas Sourlier of VICOREG for their valuable contributions in test construction, data analysis, and reporting.

REFERENCES


FTI View Difficulty

The general FTI view difficulty equation 1 describes a slight modification of the mean of the inverted detection performance value (DetPerf) over all items (index $N_{OV}$) containing the same FTI object (subindex $O$) in the same view (subindex $V$) as does the item in question. Inverted means here, that the measured detection performance is subtracted from the theoretical maximum detection performance. The slight modification refers to the exclusion of the item in question from averaging.

$$\text{FTI}_{OVj} = \frac{\sum_{i=1}^{N_{OV}} (\max(\text{DetPerf}) - \text{DetPerf}_{OVi})}{N_{OV} - 1}$$

For analyzing TIP data the inverted detection performance is the miss rate because usually only bag images containing threat items are recorded. If a large TIP data set is used, the exclusion of the item in question from the averaging can be abandoned due to its very small weight.

$$\text{FTI}_{OV} = \frac{\sum_{i=1}^{N_{OV}} \text{MissRate}_{OVi}}{N_{OV} - 1}$$

Superposition

Superposition equals the inverted Euclidean distance between the SN images (signal plus noise or threat) and N images (noise or non-threat images) regarding pixel intensity values.

$$SP = C - \sqrt{\sum_{x,y} (I_{SN}(x,y) - I_{N}(x,y))^2}$$

Clutter

This image based factor is designed to express bag item properties like textural unsteadiness, disarrangement, chaos or just clutter.

The method used in this study is based on the assumption, that such textural unsteadiness can be described mathematically in terms of the amount of high frequency regions.

Equation 4 represents a convolution of the empty bag image (N for noise) with the convolution kernel derived from a high-pass filter in the Fourier space. $I_N$ denotes the pixel intensities of the harmless bag image. $F^{-1}$ denotes the inverse Fourier transformation. $hp(f_x, f_y)$ represents a high-pass filter in the Fourier space. BS represents bag size (see equation 6). Cutoff frequency $f$ and transition $d$ (the filter’s order) were set to $f = 0.03$ and $d = 11$. The pixel summation on the high-pass filtered image was restricted to the bag’s area.

$$CL = \frac{\sum_{x,y} I_{hp}(x,y)}{BS}$$

where $I_{hp}(x,y) = I_N * F^{-1}(hp(f_x, f_y)) = F^{-1}(F(I_N * hp(f_x, f_y))$ and $hp(f_x, f_y) = 1 - \frac{1}{1 + \left(\frac{\sqrt{f_x^2 + f_y^2}}{f}\right)^d}$

Opacity

Opacity reflects the extent to which x-rays are able to penetrate objects in a bag. These attributes are represented in x-ray images as different degrees of luminosity. Equation 5 simply implements the number of pixels being darker than a certain threshold (e.g. 64) in the numerator relative to the bag’s overall size (denominator). BS represents the formula of the image based factor bag size (see equation 6).

$$OP = \frac{\sum_{x,y} (I_N(x,y) < 64)}{BS}$$

Bag Size

The bag size formula below is applicable to grayscale images represented by pixel luminosity values between 0 (black) and 255 (white). All pixels with luminosity lower than 254 (near white) are counted and summed up.

$$BS = \sum_{x,y} (I_N(x,y) < 254)$$