How Image Based Factors and Human Factors Contribute to Threat Detection Performance in X-Ray Aviation Security Screening

Anton Bolfing, Tobias Halbherr, and Adrian Schwaninger

Department of Psychology, University of Zurich, Switzerland
Max Planck Institute for Biological Cybernetics,
Tübingen, Germany
a.bolfing@psychologie.uzh.ch, t.halbherr@psychologie.uzh.ch,
a.schwaninger@psychologie.uzh.ch

Abstract. The present study examines the relative importance of a series of known and expected factors that highly affect threat detection performance in aviation security X-ray screening. Examined image based factors were threat item view difficulty, threat item superposition, bag complexity and bag size. Further, also the two human/demographic factors X-ray image interpretation training and age were examined. Image measurements and performance statistics for factors estimation are introduced. Three statistical approaches were applied in order to examine the impact of the introduced factors on threat detection performance and revealed consistent results. Bivariate correlations between detection performance and predictors/factors were analysed to estimate the isolated impact of each single factor independently of any other. Multiple linear regression analyses were applied for estimating the overall impact of all image based factors and human/demographic factors respectively. And analyses of covariance were applied in order to check for possible interaction effects between all factors of the models. All analyses were applied separately for the four threat item categories guns, knives, improvised explosive devices and other.

Keywords: aviation security, human factors, image based factors, X-ray imaging.

1 Introduction

Together with the tremendous growth of civil aviation the importance of aviation security and its public perception has dramatically increased in the last few decades [1]. The security checkpoints at the gates for x-raying passenger bags are the key element in aviation security all over the world. Despite great improvements in technical equipment, including high resolution X-ray machines, the decision whether a piece of luggage can enter an airplane or not is still made by human screeners. Therefore aviation security officers and their activity in screening passenger bags are a critical link of utmost importance in aviation security.

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 419–438, 2008.

In 2007 Schwaninger, Michel, and Bolfing [2] contributed an article on X-ray image difficulty estimation based on a set of image based factors. The study revealed that it is possible to predict average detection performance (across a sample of participants) on a single image quite well solely based on computationally accessible image properties. The image based factors used in that model were View Difficulty, Superposition, Clutter and Transparency. All image based factors can be automatically calculated. Multiple linear regression was used for statistical modeling. A comparison between the model based on automatically computed predictors (image based factors) and the same model based on human ratings (of the image based factors) revealed that our image measurements and statistics can predict human performance as well as human raters can. The study was based on a participants sample of 12 undergraduate students and on a X-ray object recognition test consisting of 256 images. The study reported in the following is an extensive amplification of the 2007 article. We were able to replicate the results of the earlier study with professional screeners and additional extensions in terms of data set size, additional factors and additional statistical analyses. The new test consists of 2048 test items and results are based on a participants sample of 90 screeners. Furthermore, three additional factors have been included: Human factors, namely Training and Age as well as the image based factor Bag Size. The number of threat categories was doubled by adding improvised explosive devices (IEDs) and 'other' to guns and knives.

In this study we analyze the effects on detection performance of prohibited items in passenger bags of two different groups of factors: 'Human factors' and 'image based factors'. The concept of image based factors subsumes all properties of the passenger bags' X-ray images that are relevant in mediating performance in detecting prohibited items. The concept of human factors subsumes available properties of the persons performing the screening task relevant in mediating threat detection performance. The aim of the reported study in this article is to investigate the role of image based and human factors on the threat detection performance in passenger bag screening tasks. For this purpose the effects of the different factors on threat detection performance, as well as their interactions will be assessed as comprehensively as possible. Previous work [3,4,2] has identified the following performance relevant image based factors: Threat Object View Difficulty, Superposition by other objects and Bag Complexity (represented in the following by Opacity and Clutter). The experiment is based on an off-line computer based test consisting of 2048 trials. The test is designed with the four image based factors View, Superposition, Bag Complexity and Bag Size systematically varied in order to avoid confounded variables. This design allows analysis of individual and combined effects of the image based factors, as well as analysis of their interactions. Furthermore we will analyze data of the human factors Training and Age [5]. Training was operationalized as the amount of hours spent on training using the 'X-Ray Tutor' computer based training system prior to testing.

1.1 Image Based Factors

Schwaninger (2003) [3] and Schwaninger, Hardmeier, and Hofer (2004) [4] have identified three image based factors which affect human threat detection performance

Fig. 1. Image based factors

significantly: View Difficulty, Superposition, and Bag Complexity. These image based factors have been modeled into mathematical formulae [2,6]. View Difficulty is implemented as an a posteriori calculable value named FTI View Difficulty. The abbreviation FTI represents fictional threat item, X-ray images of threat objects being artificially projected into X-ray images of passenger bags. 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.

FTI View Difficulty. The general formula for FTI View Difficulty can be seen in Equation 1. It is 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 the item in question. 'Inverted' refers to the fact that the measured detection performance is subtracted from a theoretical maximum detection performance (max(DetPerf)), in order to ensure that high values of FTI View Difficulty correspond with high difficulties. The slight modification refers to the exclusion of the item in question from averaging.

$$\mathrm{FtiVD}_{OVj} = \frac{\sum_{i=1, j\neq i}^{N_{OV}} (max(\mathrm{DetPerf}) - \mathrm{DetPerf}_{OVi})}{N_{OV} - 1} \tag{1}$$

Superposition. Superposition is modeled as the inverted Euclidean distance in pixel intensity between an SN image (signal plus noise or image containing a threat item) and its corresponding N image (noise or non-threat image).

$$SP = C - \sqrt{\sum_{x,y} \left(I_{SN}(x,y) - I_N(x,y)\right)^2} \tag{2}$$

Please note that in all reported analyses we used logarithmically transformed Superposition values. After inspecting the scatterplots of all our factors with the detection performance d' in order to check for non-linear relationships (a violation of the requirements of multiple linear regressions) we decided to linearize Superposition by this log-transform. This way heteroscedasticity (another violation of the requirements of MLR) can be avoided and the explained variance of the relationship between the factor and detection performance can be increased.

Clutter. Clutter is designed to capture bag image properties like disarrangement, textural noise, chaos or just plain clutter. We modeled the Clutter variable based on the assumption that image properties like the ones mentioned above correspond with larger amounts of high spatial frequencies in the image. Equation 3 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. $\mathcal{F}^{-1}$ 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 5). Cut-off 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.

$$\mathrm{CL} = \frac{\sum_{x,y} I_{hp}(x,y)}{BS} \tag{3}$$

$$\begin{aligned} \text{where} \quad I_{hp}(x,y) &= I_N * \mathcal{F}^{-1}(hp(f_x,f_y)) \\ &= \mathcal{F}^{-1}(\mathcal{F}(I_N \cdot hp(f_x,f_y)) \\ \text{and} \quad hp(f_x,f_y) &= 1 - \frac{1}{1 + \left(\frac{\sqrt{f_x^2+f_y^2}}{f}\right)^d} \end{aligned}$$

Opacity. Opacity represents how well X-rays are able to penetrate an object. High Opacity values represent low penetrability. In X-ray images this property is represented by pixel color and brightness. Opacity represents the total size of areas with pixels being darker than a certain threshold relative to the bag's size. In Equation 4 all pixels being darker than a certain threshold (e.g. 64) are summed up and divided by the bag's size (Bag Size as denominator).

$$\mathrm{OP} = \frac{\sum_{x,y} \big(I_N(x,y) < 64\big)}{\mathrm{BS}} \tag{4}$$

Bag Size. The Bag Size formula below is applicable to grayscale images with pixel brightness values ranging from 0 (black) to 255 (white). All pixels with brightness values lower than 254 (almost white) are considered as part of the bag. Bag Size is then defined as the size of the bag in number of pixels

$$\mathrm{BS} = \sum_{x,y} (I_N(x,y) < 254) \tag{5}$$

2 Methods and Procedures

2.1 Participants

The participants sample consists of 90 professional aviation security X-ray screening officers from two European airports (48 females). On average females are 40.6 and males 35.9 years old with standard deviations of 17.8 and 13.6 years respectively.

2.2 Stimuli

The 2048 test stimuli were created automatically using the image measurement algorithms described above. The number of trial images is determined by the following test design: The test consists of eight threat exemplar pairs per threat category. Given the categories Guns, Knives, IEDs and Other this results in 64 different exemplars of threat items. Each of these threat items is presented with each possible factor combination. Each of the image based factors introduced above is implemented in the design with two dichotomous parameter values representing low and high values. For View Difficulty, Superposition, Bag Complexity and Bag Size this results in 2x2x2x2= 16 factor combinations. The 64 threat exemplars in 16 factor combinations result in 1024 images. In order to apply signal detection theory [7] in the analysis all 1024 bag images containing fictional threat items (FTIs) are also presented in the test not containing any threat items. This results in the total of 2048 images.

The construction process of the test stimuli was partly manual and partly automated. In a first step the 64 threat exemplars were chosen manually such that the diversity of threat items in each of the four categories is well represented. In a second step a set of 1024 bag images was chosen based on the image measurements introduced above. In total 6659 bag images were analyzed regarding Clutter, Opacity and Bag Size. The 1024 bag images that were finally used for the test were all independently checked for credibility by X-ray screening experts. Subsequently we determined the membership of each image regarding high or low parameter values by applying median splits on each of the three image based factor distributions Opacity, Clutter and Bag Size. Opacity and Clutter are very highly intercorrelated. Thus it did not make sense to define and vary high and low parameter values for Opacity and Clutter independently. Instead the dichotomous variable Bag Complexity was defined based on Opacity and Clutter: For Bag Complexity high and low parameter values were defined as bags with both high or low Opacity and Clutter values, respectively. Bags with high Opacity and low Clutter values or vice versa were discarded. For each of the resulting factor combinations - low/high Bag Complexity x small/large Bag Size - 256 images were chosen manually. In the last step fictional threat items were automatically merged with the harmless bags. Each of the 64 threat exemplars exists in two different views. The easy views were depicted in frontal view and the difficult ones were depicted in a 85° rotation relative to the frontal view, either horizontally or vertically. This results in a total of 128 fictional threat items (FTIs). All FTIs were recorded by professional X-ray screening experts for credibility. Each of these 128 FTIs was finally merged with the 256 harmless bags with two different levels of Superposition - low and high. This procedure was applied four times for each combination of harmless bags. As already mentioned, this process was fully automatic. The underlying merging algorithm merges the images and calculates the Superposition value. If the Superposition value lies in the low or high superposition level range it is being saved as such. If not, the process is repeated until the FTI can be merged within the desired superposition value range.

2.3 Procedure

Since the test contains a very large amount of items participants completed it over multiple sessions. The test presentation was implemented within the computer based training system X-Ray Tutor 2.0 which can be run as a testing and a training environment. X-Ray Tutor is a well-established training tool designed to effectively improve and reliably test X-ray image interpretation competency. Customer airports are recommended to advise their security screeners to practice at least 20 minutes per week using X-Ray Tutor. The current test was inserted into the familiar training sequences and continued after each login until completion. After that, normal training continued. The shared basis of training and testing allows extracting training data of each screener prior to testing.

The 2048 images are presented in random order. The participants' task is to decide whether a piece of luggage would be OK or not OK to hypothetically enter an airplane by pressing buttons OK or NOT OK with a computer mouse. Unlike in training mode, in testing mode participants receive no feedback with regards to the correctness of their answers. Data are recorded in a database.

2.4 Statistics

The data are analyzed in three ways. First we present all bivariate correlations between the independent variables (image based factors and human factors) and detection performance d'. This gives a good estimation of how well detection performance can be predicted on the basis of our predictors. The second statistics we present are two separate linear regression models, one for image based factors and one for human factors, respectively. Regression analyses allow estimating the combined impact of the respective predictors together. Regression analyses allow estimating to what extent a certain set of predictors is able to predict the measured values [8], in this case the measured detection performance d'. The regression analysis using the image based factors as predictors is a replication and refinement of an earlier study by Schwaninger et al. [2] which was based on the X-Ray ORT [9]. The present analysis is based on a much larger item- and subject sample since we used a new test for the present study. The most notable differences between the X-Ray ORT and this test are the inclusion of the new image based factor Bag Size and the extension of the threat item categories by IEDs (Improvised Explosive Devices) and "Other". Since linear regression analyses do not take into account any kind of interaction effects between the predictors we report a third type of analysis. The analysis of covariance (ANCOVA), with our image based factors operationalized as repeated measures variables and the human factors as covariates. It is important to be aware of potential interaction effects, since the presence of large interactions would limit the amount of variance explained by the multiple linear regression models.

3 Results

In conformity with the Statistics section the results are reported in the following order: First we report the findings of the bivariate and partial correlations. In a second step we report the results of the multiple linear regression analyses per threat item category as well as for all categories combined. One set of multiple linear regression analyses will be based on image based factors and the other on human factors. Finally we report an ANCOVA with the image based factors as repeated measures and with the human factors Age and Training as covariates.

3.1 Bivariate Correlations

Figure 2 shows the bivariate correlations with d' of each image based factor and the partial correlations with d' of each human factor - with the respective other human factor serving as a control variable. The reason why we decided to treat image based factors and human factors differently is the following: The image based factors have been implemented within the computer based test in order to obtain orthogonal relationships between them. In other words, image based factor values vary independently across test items. Since we could not ensure independence of the human factors Age and Training through test design or sample selection, orthogonal relationships between human factors cannot be assumed. The data reveal that indeed Age and Training are confounded, with people tending to train more the older they are. Therefore we decided to additionally report partial correlations to avoid false conclusions regarding the effect of age and training on visual search tasks. Furthermore we decided to graphically report R^2 values instead of plain Rs. The great advantage of R^2 over R is that it can be directly interpreted as the amount of variance in the dependent variables (d') that can be explained by the independent variable (single factors). The disadvantage is the loss of information on the sign of the relationship due to squaring.

Figure 2 illustrates the relationship of the individual factors on detection performance d' by threat category for image based factors and human factors separately. The graphs clearly reveal that three factors explain a substantially higher amount of variance than all the others, i. e. FTI View Difficulty, Superposition and Training (log-transformed training hours). Age also shows a notable effect which is much smaller, but remains stable across all threat categories. Exact values are reported in Table 1. A detailed discussion on the data patterns is given in the Discussions section at the end of this article.

3.2 Multiple Linear Regression Analysis

Figures 3 - 5 all show scatterplots illustrating the statistical relationship between the observed (empirically measured) detection performance values d' (ordinate) and the

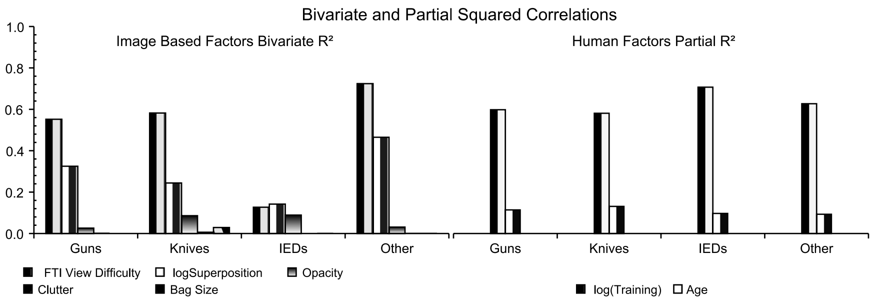

Fig. 2. The bivariate R^2 values for the image based factors and the partial R^2 values for the human factors are estimates for the amount of variance in detection performance d' that can be explained by the single factors

Table 1. Tabulation of the correlations between the single factors (human and image based) and the detection performance measure d' separately for each threat category. For human factors, also partial correlations are given, the respective other human factor taken as the control variable.

	Image based factors bivariate correlations with d'					**Human factors** partial correlations with d'		**Human factors** bivariate correlations with d'	
	FTI-VD	logSP	OP	CL	BS	logTR	Age	logTR	Age
Guns	-.74	-.57	-.16	-.06	-.02	.77	-.34	.75	.19
Knives	-.76	-.49	-.29	-.09	-.17	.76	-.36	.73	.16
IEDs	-.36	-.38	-.30	-.02	-.05	.84	-.31	.84	.28
Other	-.85	-.68	-.18	-.05	-.05	.79	-.31	.78	.24

standardized predicted values estimated by the respective multiple linear regression models (abscissa). For each model R^2 and R values are displayed in the bottom right corner of the scatterplot as a measure for the closeness of the relationship between model prediction and empirical measurements.

Models Across Categories. Figure 3 shows the scatterplots of the multiple linear regression models for the image based factors and human factors respectively. Differences concerning threat categories are not taken into account here. Both models can explain nearly 70% of the observed variance in d'. In the image based factors model 1024 data points are estimated. Each data point represents one signal-noise/noise item pair with its hit rate and false alarm across all 90 screeners. In the human factors model there are only 90 data points because d' values are calculated per subject across all 1024 item pairs.

Table 2 shows the most important statistical values of the multiple linear regression analyses for both models.

Models by Category. Figures 4 and 5 show the corresponding linear regression model scatterplots to Figure 3 but separately for each threat category. This allows us to

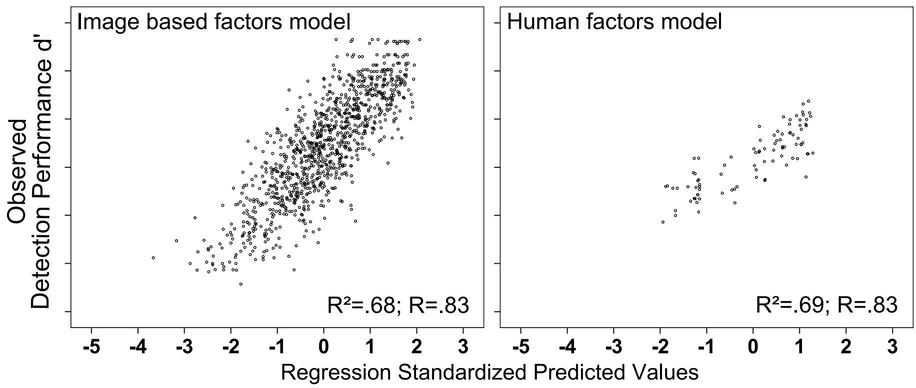

Fig. 3. General regression models across all threat item categories. Scatterplots with standardized predicted values of the image based and human factors multiple linear regression models on the x-axis and observed detection performance d' on the y-axis.

Table 2. Tabular summary of the general multiple linear regression models for all threat item categories. Standardized beta weights and p-values.

Model Summaries (All Categories)			
	Predictors	Beta weights β	Significance p
Image Based Factors	FTI View Difficulty	-.70	.000
	logSuperposition	-.127	.000
	Opacity	-.329	.000
	Clutter	.198	.000
	Bag Size	.021	.288
	$R^2 = .68$, adjusted $R^2 = .68$, $F(5, 1018) = 441$, $p < .000$		
Human Factors	logTrainingHours	.93	.000
	Age	-.26	.000
	$R^2 = .69$, adjusted $R^2 = .68$, $F(2, 87) = 98$, $p < .000$		

compare the relationship between image based factors and threat detection performance with the relationship between human factors and threat detection performance separately for each threat category. Table 3 shows the most important statistical values for each of the reported models.

Figure 4 shows the four scatterplots illustrating the predictive power of the image based factors regression model separately for each threat category.

Figure 5 shows the four scatterplots illustrating the predictive power of the human factors regression model separately for each threat category.

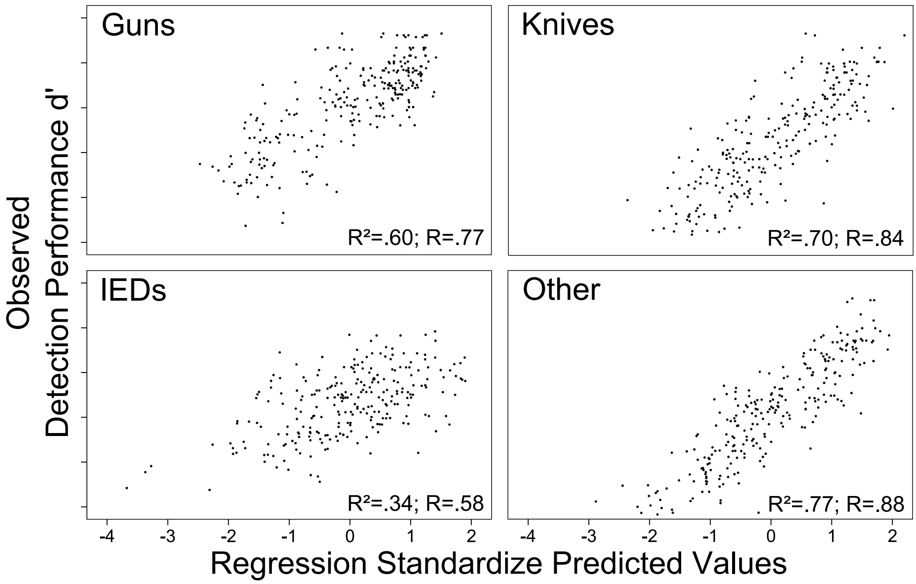

Fig. 4. Separate image based factors regression models for each of the four threat item categories

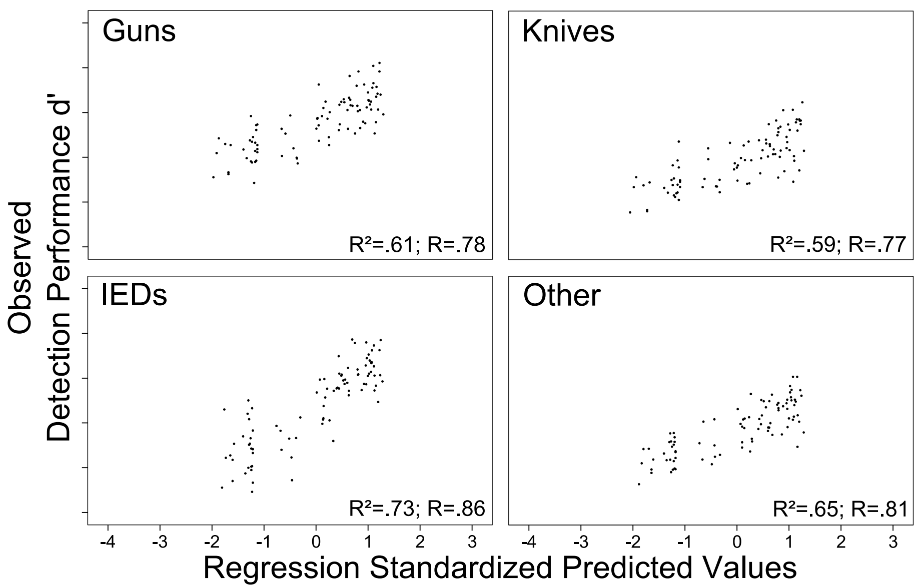

Fig. 5. Separate human factors regression models for each of the four threat item categories

Table 3. Tabular summary of seperate multiple linear regression models for each of the four threat item categories. Standardized beta weights and p-values.

Model Summaries (Per Category)									
	Predictors	Guns		Knives		IEDs		Other	
		β	p	β	p	β	p	β	p
Image Based Factors	FTI View Difficulty	-.67	.000	-.67	.000	-.29	.000	-.79	.000
	logSuperposition	-.12	.025	-.18	.000	-.25	.000	-.09	.061
	Opacity	-.27	.000	-.36	.000	-.51	.000	-.31	.000
	Clutter	.18	.005	.17	.003	.35	.000	.20	.000
	Bag Size	.04	.375	-.05	.256	.09	.105	.05	.187
	R^2	.60		.70		.38		.77	
	$adjustedR^2$	.59		.70		.32		.77	
	$F(5, 250)$	75		117		25		169	
	p	.000		.000		.000		.000	
Human Factors	logTraining	.88	.000	.88	.000	.94	.000	.90	.000
	Age	-.26	.001	-.29	.000	-.20	.003	-.22	.004
	R^2	.61		.59		.73		.65	
	$adjustedR^2$	.60		.58		.72		.64	
	$F(2, 87)$	69		63		118		80	
	p	.000		.000		.000		.000	

3.3 ANCOVA

ANCOVA with Category Treated as a Repeated Measures Factor. Figure 6 shows a short overview of the ANCOVA results. The ANCOVA allows us to integrate human factors as covariates into a repeated measures ANOVA of image based factors (including threat category) and thus allows us to explore interaction effects and dependencies among not just the image based factors, but also between the image based factors in combination with the human factors. On the left, Figure 6 illustrates the importance of the image based factors in terms of their effect size values η^2 (eta square) and their interactions with the human factors. The main effects of each image based factor are reported together with their interaction effects with Training (log-transformed training hours) and Age, respectively. On the right, Figure 6 additionally illustrates the ten largest significant interactions in terms of η^2 values. For details on the data, their patterns and conclusions please consider Table 4 and the Discussions section at the end of this article.

Table 4 shows all η^2 values and the significance levels of the main effects and the interaction effects illustrated in Figure 6.

ANCOVAs by Category. Figure 7 shows the results of the ANCOVAs applied separately for each threat item category. Only image based factors main effects and interactions with the human factors are given here. Refer to the Discussions section for detailed interpretation.

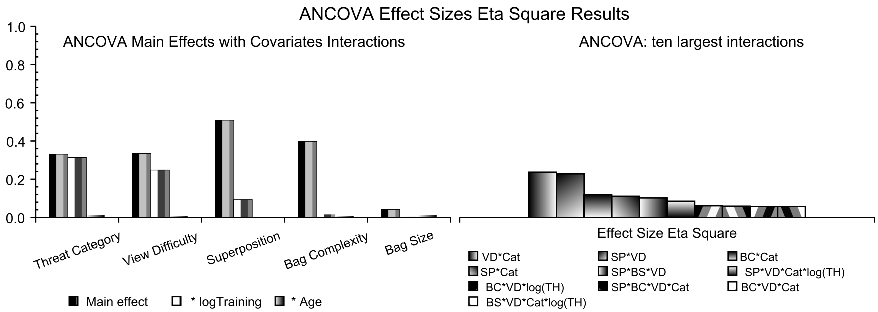

Fig. 6. Summary of ANCOVA main effect and interaction effect sizes. All covariate interactions and the ten largest remaining interactions are reported.

Table 4. Summary of ANCOVA main effects and covariate interactions

ANCOVA effect sizes η^2					
	Category	View Difficulty	Superposition	Bag Complexity	Bag Size
Main effects	$.33^{***}$	$.34^{***}$	$.51^{***}$	$.40^{***}$	.04
* logTraining	$.31^{***}$	$.25^{***}$	$.09^{**}$	.02	.00
* Age	.01	.01	.00	.01	.01

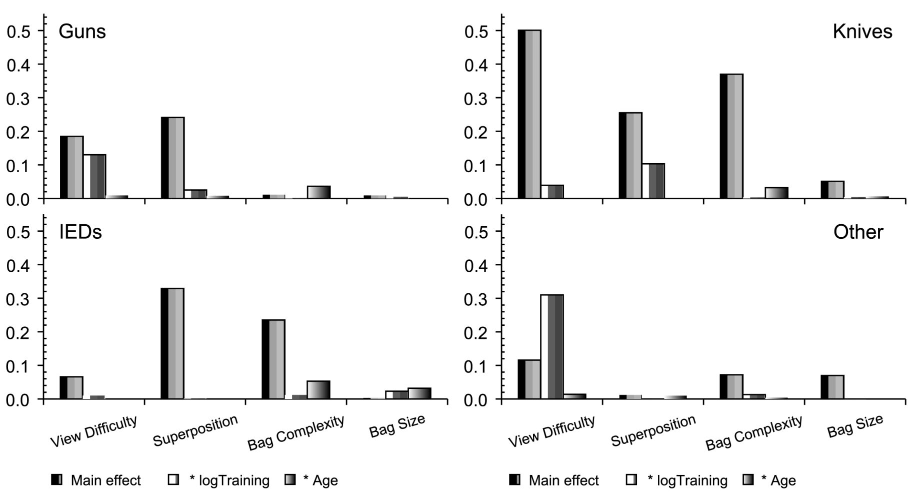

Fig. 7. ANCOVA by Category

Table 5. Tabular presentation of the ANCOVA results by threat item category

ANCOVA effect sizes η^2						
	Category	View Difficulty	Superposition	Bag Complexity	Bag Size
Main effects	$.33^{***}$	$.34^{***}$	$.51^{***}$	$.40^{***}$	$.04$
* logTraining	$.31^{***}$	$.25^{***}$	$.09^{**}$	$.02$	$.00$
* Age	$.01$	$.01$	$.00$	$.01$	$.01$
Main effects	Guns	$.19^{***}$	$.24^{***}$	$.01$	$.01$
* logTraining		$.13^{**}$	$.03$	$.00$	$.00$
* Age		$.01$	$.01$	$.04$	$.00$
Main effects	Knives	$.50^{***}$	$.26^{***}$	$.37^{***}$	$.05^{*}$
* logTraining		$.04$	$.10^{**}$	$.00$	$.00$
* Age		$.00$	$.00$	$.03$	$.01$
Main effects	IEDs	$.07^{*}$	$.33^{***}$	$.24^{***}$	$.00$
* logTraining		$.01$	$.00$	$.01$	$.02$
* Age		$.00$	$.00$	$.05^{*}$	$.03$
Main effects	Other	$.12^{**}$	$.01$	$.07^{*}$	$.07$
* logTraining		$.31^{***}$	$.00$	$.01$	$.00$
* Age		$.01$	$.01$	$.00$	$.00$

Table 5 shows all η^2 values and the significance levels of the main effects and the interaction effects illustrated in Figure 7.

4 Discussions

For the discussion of our findings we retain the same presentation order as we did in the Methods and Results sections. Starting with the findings of the bivariate correlations we continue discussing the multiple linear regression models ending with the discussion of ANCOVA results.

4.1 Bivariate Correlations

The correlations between our factors and d' can be interpreted as the observed relationships between our predictors and d' observations. This gives us a first impression of how much explained variance we can expect from a single predictor. Figure 2 and Table 1 reveal that there are close relationships between d' and the three predictors FTI View Difficulty, Superposition (log-transformed) and Training (log-transformed training hours). Also we can report a notable (partial) correlation between Age and d'. The remaining three predictors Opacity, Clutter and Bag Size show poor correlations. Clutter and Bag Size do not even reach the level of statistical significance of $p = .05$, Bag Size in knives being the only exception ($p < .01$). These findings are very surprising since we know from visual search research literature [10] that detection performance should decrease with growing set size. If Bag Size, the size of the area to search in, does not correlate with d' we would expect that Clutter, the amount of distractors in the set, does. Table 1 shows both the bivariate and partial correlations of the human factors with d' in order to allow direct comparisons. There are only very slight changes between bivariate and partial correlations with Training, but note that the signs of the correlations with Age all change from positive to negative when calculating partial correlation. The bivariate correlations reveal that the participants improve their detection performance with increasing age. This finding contradicts earlier studies on visual search tasks that revealed a deterioration of performance with age [11]. The partial correlations put this in perspective: Participants compensated age with more training, and indeed when controlling for training there is a small negative correlation between age and detection performance.

A very interesting aspect of this analysis is the comparison of the correlations' patterns for the different threat categories. The data reveal that different factors are important for being able to identify FTIs belonging to different threat categories. Figure 2 illustrates impressively how View Difficulty has a similar amount of influence on detection performance of guns and knives, but a much smaller amount on IEDs and a notably larger one on Other. Interpreting the pattern of Other is very difficult because Other is just the rest category for all the threat objects that do not fit into any of the other three categories. Thus the category Other includes as diverse objects as throwing stars (shuriken), tasers, hand grenades or gas tanks. Even though the IED stimulus set contains all sorts of hand made bombs, the stimuli are still comparatively homogenous making an interpretation of the image based factors much simpler than with Other: IEDs are generally made up of multiple essential parts (explosive material, fuse, cables, energy source, timer, etc.). Each of these has its own rotation (View) and its own Superposition value. Therefore it is not too surprising that the effects of View Difficulty and Superposition are highly diminished. A very interesting complementary finding is that while for IEDs image based factors in general show the lowest impact on detection performance compared to the other threat categories, human factors - especially Training - show the strongest effects on d' with IEDs.

4.2 Multiple Linear Regression Models

As already anticipated in the Results section we are very happy to report the achieved explained variances of nearly 70% for both the image based factors regression model as well as for the human factors regression model. The fact that we are able to explain such a big portion of variance from two distinct sets of predictors independently makes us very confident to get a grip on the process of X-ray threat detection. We are very confident that the image based factors together with human factors constitute the key aspects to cover for a better general understanding of the cognitive processes involved in this kind of visual search task. Nevertheless we still see some potential to further augment our model fits. This applies to the image based factors model as well as to the human factors model. Particularly with regards to the implementation of Clutter we see great potential to elicit larger predictive power - though to date we have not yet found a

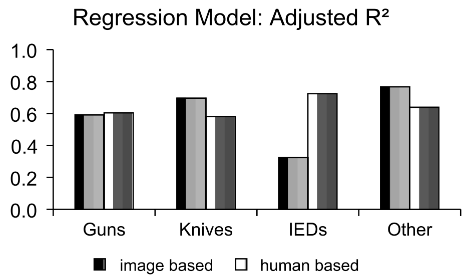

Fig. 8. Comparison of adjusted R^2 values of image based factor and human factor linear regression models for each threat item category

mathematical formula that rendered better results than the one currently in use. As for human factors: Until now we have merely investigated Training and Age. We see great potential in enlarging the set of human factors. Besides the undoubtedly important factor Training we expect visual abilities such as mental rotation, figure-ground segregation or visual search for highly specific patterns to be another important factor that should not be disregarded in a comprehensive model. Unfortunately for this study appropriate data were not available. As another human factor Gender should also be taken into account in future research.

Regarding the differences in terms of the explained variances and how they are made up of in terms of the beta weights among categories some remarkable notes must be made. The first pattern to catch one's eye is the different behavior of IEDs compared to the other threat categories. In the case of IEDs the image based factors' model regarding d' shows a comparatively low amount of explained variance. For all other categories the image based factors have a very high predictive power. A closer look at the beta weights and correlations reveals that it is particularly the effect of FTI View Difficulty which lies far below what would be expected based on the results from the other threat categories. The complementary finding is that compared to the other categories IEDs show the largest amount of explained variance with the human factors model. The beta weights and correlations reveal that this can overwhelmingly be attributed to Training. We can conclude that IEDs depend largely on human factors, especially on Training. We assume that detection performance with IEDs depends on knowledge as opposed to visual abilities to a larger extent than is the case with the other categories. In general the human factor models show much lower differences in predictive power among threat categories than the image based factor models do. We assume that the explanation for the comparatively large variation of the overall predictive power of the image based models as well as the comparatively large amount of variation between the correlations of the individual image based factors with d' lies in detection performance being based on several distinct cognitive processes dealing with the different image based factors, with their relative importance varying between categories.

4.3 ANCOVA/ Interactions

There are some essential differences concerning data types. The factor View Difficulty as it is operationalized for the ANCOVA cannot be directly compared with the variable FTI View Difficulty used in the correlation analyses and the regression models. FTI View Difficulty is a proportionally scaled measure derived statistically from performance data whereas the ANCOVA factor View Difficulty is a dichotomous nominal variable differentiating between easy and difficult views only. As a matter of fact, all image based factors - except of course for the factor 'Threat Category' - are dichotomous in the ANCOVA (refer to the paragraph on stimuli in the Methods section). Bag Complexity replaces the compound of Opacity and Clutter (refer to the Stimuli section).

ANCOVA with Category Treated as a Repeated Measures Factor. The main effects of the ANCOVA on all categories give a similar picture as did the correlations. In analogy to what we found in the correlation analyses, effects of Superposition are larger than effects of Bag Complexity and Bag Size. Furthermore the results show that there are considerable interactions with Training regarding the different threat categories. For example, as discussed above the effect of Training on detection performance is clearly larger in IEDs than in knives.

View Difficulty and Superposition also show interaction effects with Training. The improvement of the detection performance caused by training is clearly larger regarding the difficult views compared to the easy ones. The interaction between Superposition and Training on the other hand is fairly small. This could indicate that dealing with superposition is difficult to improve with training. For Bag Complexity and Bag Size interactions with Training are not significant. No evidence could be provided for interaction effects of any of our image based factors (including threat category) with Age.

In the following itemization we give a short overview of the four largest reported first order interactions and plausible interpretations with examples.

- VD * Cat: reflects the fact that effects of View Difficulty differ between Threat Categories e.g.: compare correlations of View Difficulty and d' between IEDs and Other
- VD * SP: difficult views are more affected by high Superpositions that easy ones and vice versa
- BC * Cat: reflects the fact that effects of Bag Complexity differ between Threat Categories
- SP * Cat: reflects the fact that effects of Superposition differ between Threat Categories

ANCOVAs by Category. Performing the ANCOVA individually per category makes several interesting effects visible. According to ANCOVA and in correspondence with the results of the correlation analysis, detection performance with guns depends on View Difficulty and Superposition. The interaction of View Difficulty with Training indicates that for guns training is particularly effective with difficult views. If we compare the relative effect size of View Difficulty in the ANCOVA with the relative effect size (R^2) of FTI-View-Difficulty in the correlation analysis we can see that the effect of View Difficulty in the ANCOVA is comparatively smaller. Since FTI-View-Difficulty also measures effects of properties of the threat item itself - i.e. the specific gun - we can conclude that for guns it is not only the View Difficulty that is important in mediating detection performance but also the kind of gun in question. For Knives View Difficulty, Superposition and Bag Complexity have large main effects. In contrast to the results for Guns, View Difficulty does not have a comparatively smaller effect size in the ANCOVA than in the correlation analysis, indicating that for Knives View Difficulty is indeed important in mediating detection performance, whereas the type of knife is not. Furthermore, for Knives the interaction between Training and View Difficulty does not reach statistical significance. In other words: dealing with View Difficulty in Knives is not improved through training in Knives. There is however a significant interaction between Superposition and Training. Finally, Knives are the only category with a (barely) significant main effect of Bag Size. The ANCOVA results for IEDs confirm the negligible influence of View Difficulty on detection performance for this threat category. Performance is mediated by Superposition and Bag Complexity alone. There is a small but statistically significant interaction between Bag Complexity and Age with IEDs. The ANCOVA for the category Other shows only two significant and still small main effects, namely the effects of View Difficulty and Bag Complexity. However there is a large interaction between View Difficulty and Training. The ANCOVA results for Other are in stark contrast to the results of the correlation analyses - especially with regards to Superposition. Superposition had a large effect size (R^2) in the correlation analysis whereas in the ANCOVA the main effect of Superposition does not even reach statistical significance. As has been mentioned above, the category Other is very heterogeneous which makes any interpretation of these seemingly contradictory results very difficult and speculative.

5 Conclusions and Recommendations for Improving Human-Machine Interaction in X-Ray Screening

5.1 FTI View Difficulty and Superposition

To date, X-ray screening technology delivering only one image per bag have been common in aviation security. More recent technology is capable of providing multiple views of a bag. Our research has shown that the image based factor View Difficulty can be addressed effectively with computer based training. The ANCOVA analysis has supported earlier findings showing that training improves detection performance particularly for difficult views [12]. The results for Superposition were not as promising, however. Figure 9 illustrates how new multi-view systems might be able to reduce the detection problems due to View Difficulty and Superposition. Objects that are very much superimposed by other objects from one perspective may be clearly visible from another.

5.2 Opacity

Opacity refers to the extent to which a bag is penetrated by X-rays. X-ray systems with higher penetration have the potential to reduce detection problems due to Opacity. Even

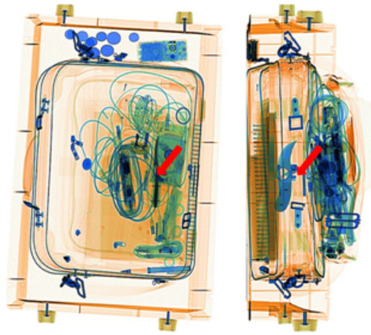

Fig. 9. Illustrative example of how multi-view systems can help improving detection performance in spite of undesirable View Difficulty and Superposition effects

so, the problem of dealing with baggage that render too opaque images remains. A solution is the implementation of a 'dark alarm' in X-ray equipments. This dark alarm would warn the human operator if a maximum amount of opaque areas in a bag were exceeded. Manual search would follow when a dark alarm was indicated.

5.3 Screener Selection and Training

A very important approach in improving X-ray screening efficiency and increasing threat detection performance is 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. Hofer, Hardmeier, and Schwaninger [13] and Hardmeier, Hofer, and Schwaninger [9] have shown that pre-employment assessment procedures based on computer based object recognition tests can help increase detection performance substantially.

The present study shows, that in addition to stable abilities there are also trainable skills that play a very important role in mediating detection performance. Knowledge based factors such as knowing which objects are prohibited and what they look like in X-ray images are learned by screeners in the computer based training. Computer-based training can be a powerful tool to improve X-ray image interpretation competency of screeners [12,14,15].

The list of human factors analysed in this study is far from being exhaustive. Additional studies investigating human and contextual factors such as vigilance, stress, heat and time pressure are in preparation but would go far beyond the scope of the study at this point.

5.4 Educational and Usability Aspects

Despite massive advances in the development of technological equipment it is still the human operator who decides whether a bag can enter an aircraft or not. As can be seen from the results regarding the importance of X-ray image interpretation training one must not forget that all technological equipment is of limited value if its operators are not trained thoroughly in their task. Therefore it is very important for engineers to keep in mind the human operators behind the technological equipments. The present study gives a series of key aspects that must be taken into account when developing training systems such as the image based factors introduced. On the other hand some of these aspects can indeed be directly addressed by technological means. For example the use of multi-view systems for dealing with view difficulty and superposition.

Acknowledgments

This research was supported by the European Commission Leonardo da Vinci Programme (VIA Project, DE/06/C/F/TH-80403). Special thanks go to the two airports that supported this study by supplying screeners and data.

References

1. Annual review of civil aviation 2005. International Civil Aviation Organization 61, 9 (2006)
2. Schwaninger, A., Michel, S., Bolfing, A.: A statistical approach for image difficulty estimationin x-ray screening using image measurements. In: Proceedings of the 4th Symposiumon Applied Perception in Graphics and Visualization, pp. 123–130. ACM Press, New York (2007)
3. Schwaninger, A.: Evaluation and selection of airport security screeners. Airport 2, 14–15 (2003b)
4. Schwaninger, A., Hardmeier, D., Hofer, F.: Aviation security screeners visual abilities & visual knowledge measurement. IEEE Aerospace and Electronic Systems 20(6), 29–35 (2005)
5. Riegelnig, J., Schwaninger, A.: The influence of age and gender on detection performance and the criterion in x-ray screening. In: Proceedings of the 2nd International Conference on Research in Air Transportation, ICRAT 2006, Belgrade, Serbia, Montenegro, June 24-28,2006, pp. 403–407 (2006)
6. Bolfing, A., Schwaninger, A.: Measurement formulae for image-based factors in x-ray imagery. In: VICOREG Technical Report, November 26 (2007)
7. Green, D.M., Swets, J.A.: Signal detection theory and psychophysics, pp. 187–194. Wiley, New York (1966)
8. Coolican, H.: Research Methods and Statistics in Psychology, 4th edn., Hooder & Stoughton, London (2004)
9. Hardmeier, D., Hofer, F., Schwaninger, A.: Increased detection performance in airport security screening using the x-ray ort as pre-employment assessment tool. In: Proceedings of the 2nd International Conference on Research in Air Transportation, ICRAT 2006, Belgrade, Serbia, Montenegro, June 24-28, pp. 393–397 (2006b)
10. Wolfe, J.: Guided search 2.0 a revised model of visual search. Psychonomic Bulletin & Review 1(2), 202–238 (1994)
11. Madden, D.J., Gottlob, L.R., Allen, P.A.: Adult age differences in visual search accuracy: Attentional guidance and target detectability. In: Psychology and Aging, pp. 683–694 (1999)
12. Koller, S., Hardmeier, D., Hofer, F., Schwaninger, A.: Investigating training, transfer, and viewpoint effects resulting from recurrent cbt of x-ray image interpretation. Journal of Transportation Security (2008)

13. Hofer, F., Hardmeier, D., Schwaninger, A.: Increasing airport security using the x-ray ort as effective pre-employment assessment tool. In: Proceedings of the 4th International Aviation Security Technology Symposium, Washington, D.C., USA, November 27 - December 1, pp. 303–308 (2006)
14. Schwaninger, A., Hofer, F., Wetter, O.E.: Adaptive computer-based training increases on the job performance of x-ray screeners. In: Proceedings of the 41st Carnahan Conference on Security Technology, Ottawa, October 8-11 (2007)
15. Ghylin, K.M., Drury, C.G., Schwaninger, A.: Two-component model of security inspection: application and findings. In: 16th World Congress of Ergonomics, IEA 2006, Maastricht, The Netherlands, July, pp. 10–14 (2006)