

The Impact of Image Based Factors and Training on Threat Detection Performance in X-ray Screening

Adrian Schwaninger*, Anton Bolfing*, Tobias Halbherr[†], Shaun Helman[‡], Andrew Belyavin[‡] and Lawrence Hay[‡]

*Max Planck Institute for Biological Cybernetics, Tübingen, Germany, and
Department of Psychology, University of Zurich, Switzerland
e-mail: a.bolfing@psychologie.uzh.ch, a.schwaninger@psychologie.uzh.ch

[†]Department of Psychology, University of Zurich, Switzerland, Email: t.halbherr@psychologie.uzh.ch

[‡]QinetiQ Limited, Hampshire, United Kingdom

Abstract—In this study, two experiments are reported which investigated the relative importance of five different image based factors and one human factor (training) in mediating threat detection performance of human operators in airport security x-ray screening. Experiment 1 was based on a random sample of roughly 16'000 records of threat image projection (TIP) data. TIP is a software function available on state-of-the-art x-ray screening equipment that allows the projection of 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 opacity in the x-ray image of a passenger bag had a small although significant effect on detection performance. The two image based factors clutter and bag size did not have a significant effect.

Experiment 2 was conducted using an offline-test in order to provide controlled and more detailed data for analyzing the image based factors from Experiment 1, as well as the human factor of training. In particular the individual factors' main effects on detection performance, main effects of all factors taken together and factor interactions were analyzed. In the test design 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. For screener training all five sites use the same computer-based training system. 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 could be another image based factor that affects detection of threat items 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, with empirically based conclusions regarding the importance of the bag size variable, by itself as well as in relation with other performance relevant factors, this study provided the scientific basis for a political decision making process regarding the improvement of aviation security.

Two experiments are reported. Experiment 1 is based on threat image projection (TIP) data. Experiment 2 is based on an off-line computer based test, which allows investigating the combined effects of image-based factors, effects of training as well as factor interactions. The use of these two methods to answer the same research question will ensure that the overall approach is complementary. Both methods have their own strengths and weaknesses: TIP data give high ecological validity but low experimental control; off-line computer based tests using controlled stimuli allow more experimental control, but less ecological validity. If both methods provide the same answer to the research question, this can be taken as stronger evidence that the findings are genuine, and not simply an artefact of the particular method used.

The two experiments both follow the paradigm using computer algorithms to estimate image based factors that influence threat detection performance in x-ray screening. This paradigm was developed at University of Zurich and presented at ICRAT 2006 in Belgrade (Bolfing, Michel, & Schwaninger, 2006a) and published before (Schwaninger, Michel, & Bolfing, 2007; Bolfing, Michel, & Schwaninger, 2006b; Schwaninger, Michel, & Bolfing, 2005). None of these papers used TIP data for analysis, which ensures high ecological validity. Experiment 2 is based on a much larger data set than the previous studies augmenting reliability. The inclusion of bag size and training as additional factors is completely novel within this paradigm. Since threat detection performance in aviation security x-ray

screening depends on the x-ray images but also on the human screeners—the final decision makers—human factors should not be neglected in a comprehensive model whose goal is to explain the x-ray threat detection process.

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: view difficulty, superposition, and bag complexity (see figure 1).

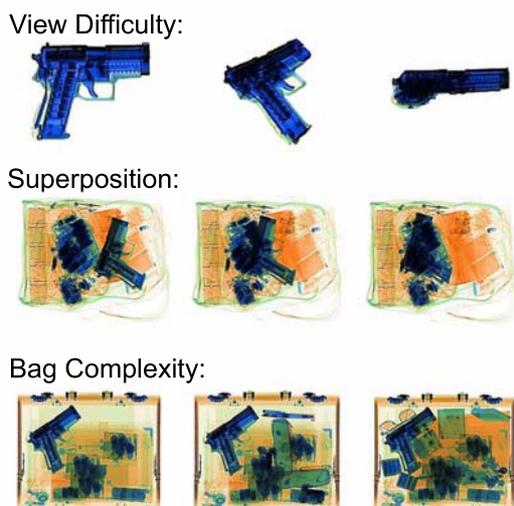


Fig. 1. Illustration of the three basic image based factors suggested by Schwaninger (2003b) and Schwaninger, Hardmeier, and Hofer (2005)

The concepts of these image based factors have been mathematically modeled (Schwaninger, Michel, & Bolfling, 2007, see Bolfling, & Schwaninger, 2007 for the latest version). View difficulty is modeled as a statistically calculable value between 0 and 1 named FTI view difficulty. Superposition and bag complexity are modeled as image processing measurements with bag complexity being 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 Bolfling, & Schwaninger (2007).

II. THREAT IMAGE PROJECTION (TIP) χ^2 ANALYSIS: EXPERIMENT 1

A. Method

1) *Threat Image Projection (TIP) Data:* In order to ensure high ecological validity, we 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 TIP distinguishes between cabin baggage screening (CBS) and hold baggage screening (HBS). In CBS, guns, knives, improvised explosive devices (IEDs) and other threats are subject to identification and

confiscation. In HBS, the focus rests mainly on IEDs and dangerous goods such as gasoline containers or diver lamps. The current investigation is confined to CBS. In CBS TIP, fictional threat items (FTIs) are occasionally projected into x-ray images of passenger bags during the routine baggage screening operation. A sufficiently large sample of TIP events allows statistically reliable measurements of 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 that have been routinely recorded on-the-job with approximately 700 professional x-ray screeners throughout the first half of 2007 at a large European airport. We decided to apply χ^2 analyses to each image based factor separately to measure its impact on detection performance in terms of hit rate (i.e. correctly judging a bag as being NOT OK).

2) χ^2 Analysis: To compare the effects on detection performance of the independent variables¹ FTI view difficulty, superposition, opacity, clutter and bag size, the following procedures were applied to the TIP data described above. A histogram was created for each independent variable (image based factor). For each variable the upper and lower 2.5% of the cases in the data were excluded to remove outlier data from the analysis. Furthermore this made possible the definition of five equidistant bins with at least 100 data points each (TIP events).

Hit rates were calculated for each of the five equidistant bins to run χ^2 tests with the null hypothesis H_0 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 χ^2 statistics see for example Coolican (2004).

B. Results

The results below are listed separately for each image based factor introduced above (see Bolfling, & Schwaninger, 2007 for further information and formulae). Each of the following subsections begins with a graphical illustration of the image based factors' effects on the threat detection performance measure hit rate. The x -axes show the five equidistant bins into which the whole data range was subdivided. Low values are on the left, high values on the right. The y -axes show the hit rates of the image based factors' bins. For reasons of confidentiality hit rates cannot be 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. χ^2 statistics can be interpreted as follows: the larger the $\chi^2(df, N)$ value the larger the effect. Additionally χ^2 effect sizes w are given.

¹The variables correspond to the continuously represented variables used in the multiple regression analysis in Experiment 2 (see figure 8)

Again, the larger the effect size, the larger the effect. However, please be aware that χ^2 and w values do not state the direction of the effect.

To summarize the χ^2 analysis results a bar plot graphic is provided at the end of this section illustrating the χ^2 effect sizes of the five image based factors on the hit rate (see figure 7). 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 contributing to this large impact are the threat category of the object and the training of human operators (see Experiment 2).

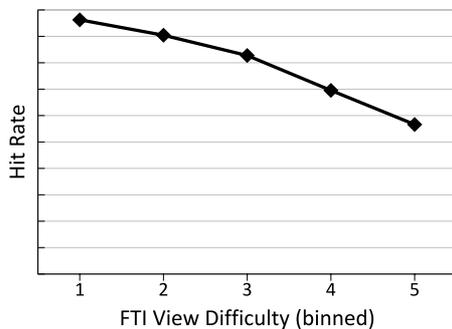


Fig. 2. Illustration of the impact of FTI view difficulty on hit rate.

TABLE I
 χ^2 ANALYSIS RESULTS: FTI VIEW DIFFICULTY

χ^2 value	$\chi^2(4, N = 13'541) = 198.04$
Significance	Highly significant: $p < .001$
χ^2 effect size	$w = .12$

2) *Superposition*: Figure 3 illustrates the large effect of superposition on detection performance.

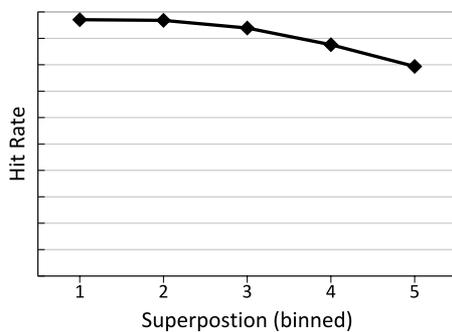


Fig. 3. Illustration of the impact of superposition on hit rate.

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

TABLE II
 χ^2 ANALYSIS RESULTS: SUPERPOSITION

χ^2 value	$\chi^2(4, N = 13'713) = 72.98$
Significance	Highly significant: $p < .001$
χ^2 effect size	$w = .07$

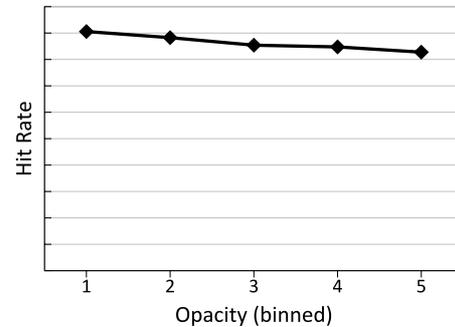


Fig. 4. Illustration of the impact of opacity on hit rate.

TABLE III
 χ^2 ANALYSIS RESULTS: OPACITY

χ^2 value	$\chi^2(4, N = 13'718) = 9.90$
Significance	Significant: $p < .05$
χ^2 effect size	$w = .03$

Here the question arises whether it is opacity as a perceptual concept that does not have much influence on threat detection performance, or whether the image measurement formula of opacity is not properly modeled.

4) *Clutter*: Figure 5 illustrates the hit rates of the five clutter bins. There is no significant effect of clutter on detection performance. As with opacity, the question arises whether it is the concept of clutter that does not influence hit rates in TIP, or whether the computational model of clutter needs to be improved.

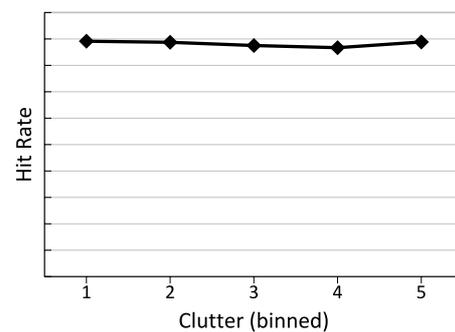


Fig. 5. Illustration of the impact of clutter on hit rate.

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

As with clutter, the effect of bag size on detection performance does not reach statistical significance.

TABLE IV
 χ^2 ANALYSIS RESULTS: CLUTTER

χ^2 value	$\chi^2(4, N = 13'726) = 0.98$
Significance	Not significant: $p = .913$
χ^2 effect size	$w = .01$

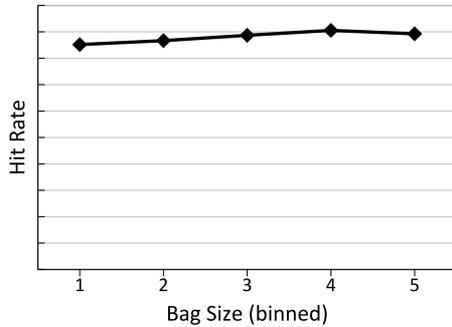


Fig. 6. Illustration of the impact of bag size on hit rate.

TABLE V
 χ^2 ANALYSIS RESULTS: BAG SIZE

χ^2 value	$\chi^2(4, N = 13'758) = 4.45$
Significance	Not significant: $p = .348$
χ^2 effect size	$w = .02$

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 with the specific operational situations encountered when threats are deliberately hidden in difficult bags. But 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*: 200 X-ray screeners from five European sites with varying amounts of training in x-ray image interpretation.

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 items: guns, knives, improvised explosive devices (IEDs), and 'other' threats. Images of cabin baggage were captured from x-ray machines at a 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 the 1024 bags needed for the study were drawn. 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 would be orthogonal in the stimulus set. Several full sets of 2048 images (the 1024 images containing the FTIs, and the same images without FTIs) were created. The one with the most desirable properties in terms of variable orthogonality was chosen for use in the study.

3) *Design*: The study employed a 4 (FTI category: guns, knives, IEDs, other) x2 (view difficulty: easy, difficult) x2 (superposition: low, high) x2 (bag complexity: low, high) x2 (bag size: small, large) x2 (image type: FTI, no FTI) within-participants design. Since there were 16 FTIs in each category, this design results in a total of $16 \times 4 \times 2 \times 2 \times 2 \times 2 = 2048$ images which were to be presented to the screeners. The images

6) *Comparison of the χ^2 Effect Sizes*: In figure 7, the effect sizes w are compared. The factor FTI view difficulty has the highest effect size with $w = .12$, while clutter shows the lowest effect size with $w = .01$. The factors opacity, bag size and clutter show small effect sizes. The effects of clutter and bag size did not reach statistical significance.

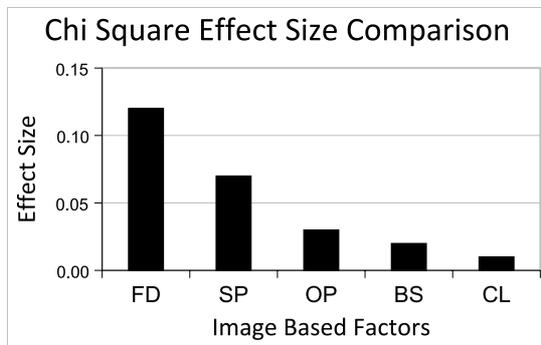


Fig. 7. Comparison of the effect sizes among the image based factor.

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 affect screener performance. Schwaninger, Michel, and Bolting (2007) replicated these results. Using similar image measurements as in Experiment 1, they measured similar effects for 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

were presented to the 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 than the hit rate alone because it takes the hit rate and the false alarm rate into account (see Hofer & Schwanger, 2004 for different measures of x-ray detection performance). Since the off-line test showed each bag once with a threat and once without one, accurate measurements of hit and false alarm rates could be obtained.

B. Results

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. A multiple linear regression was used to examine the main effects together. Additionally, we calculated a linear regression with hours of recurrent computer based training prior to testing as predictor. 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 the 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).

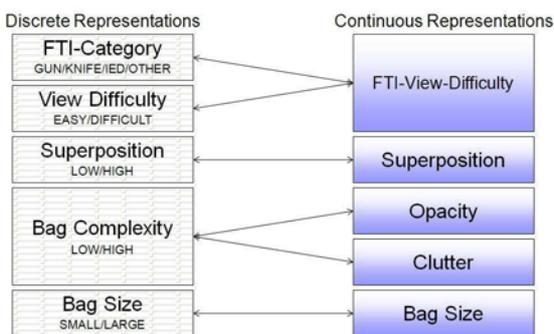


Fig. 8. Illustration of relationship between discrete and continuous representations of variables

1) *Linear Regression and Multiple Linear Regression:* The regression analyses will help us understand the direct 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. This transformation was necessary in order to achieve a linear relationship between superposition and 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.

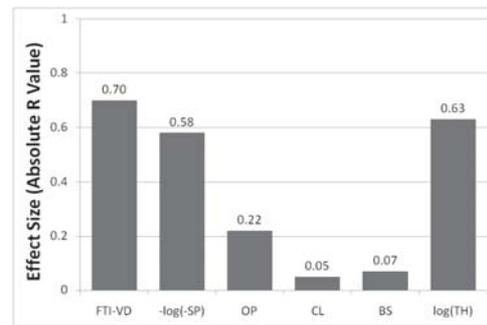


Fig. 9. Illustration of effect sizes R

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. It shows the overall effect size, again the absolute value of the correlation R, of all the image based factors taken together. With $R = 0.77$ the effect size is very high. The effect size of the only human factor analyzed (hours of recurrent computer based training), with $R = 0.63$, is also large. We can see that in the multiple linear regression model the factor bag size is the only one not reaching statistical significance. Put another way: In the presence of the other image based factors bag size did not lead to a statistically significant change in detection performance in our experiment. As shown in figure 12 adding bag size to the linear model only leads to a minimal increase of its effect size from $R = 0.772$ to $R = 0.773$.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.773 ^a	.597	.595	.690

^a. Predictors: (Constant), BS, FTIViewDifficulty, OP, -LOG(-SP), CL

Fig. 10. Multiple linear regression overview

2) *ANCOVA:* A repeated measures analysis of covariance (ANCOVA) was conducted to analyze the main effects of image based factors, their interactions and their interactions 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

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.232	.314		-7.104	.000
	FTI/ViewDifficulty	.755	.033	.568	23.163	.000
	-LOG(-SP)	-.822	.089	-.227	-9.195	.000
	OP	-10.243	.860	-.366	-11.917	.000
	CL	23.505	3.370	.223	6.975	.000
	BS	1.51E-006	.000	.030	1.304	.193

^a . Dependent Variable: d

Fig. 11. Multiple linear regression details

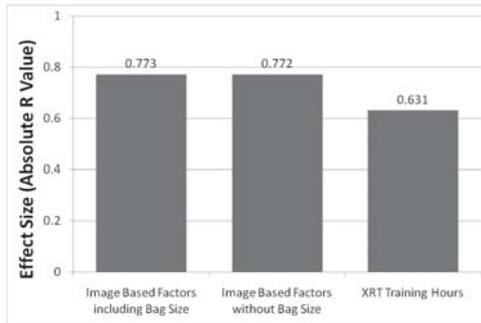


Fig. 12. Combined effect size of image based factors and effect size of training

ANCOVA analysis effects of the covariate training hours are isolated from the effects of image based factors. Furthermore, in the ANCOVA inter-individual differences between screening officers ('screener variance') are taken into account. Superposition shows the largest effect size (η^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 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 $\eta^2 \geq .07$. The interaction of view difficulty with threat category can at least partly be explained 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 with easy views. The interaction of superposition with threat category indicates that some threat item categories are more sensitive to superposition than others. For example, from the regression analysis above we know that superposition effects are higher with knives than with guns.

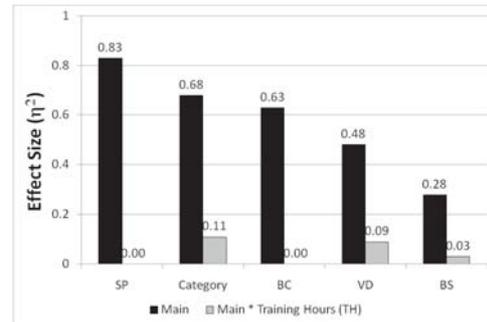


Fig. 13. Illustration of ANCOVA main effects and interactions with the covariate training hours

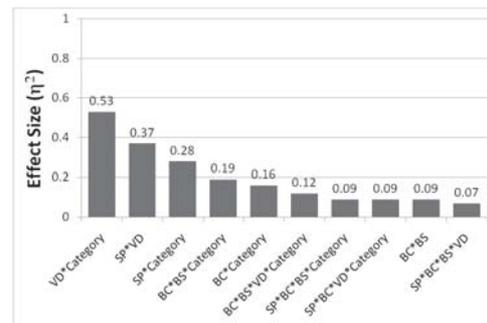


Fig. 14. Illustration of the the ten largest ANCOVA interactions

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, although 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 large effect size in the ANCOVA. Apart from this, the ANCOVA results reflect the regression analysis results closely, both in main effects and interactions. Threat category and view difficulty had considerable interactions with the covariate training hours. This shows that training is particularly effective in the case of difficult item categories such as IEDs and for difficult viewpoints. Bag size, although intuitively plausible as relevant factor, turned out to play only a minor role in determining threat detection performance. The same is true for 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 for superposition and complexity (to a bigger extent for opacity than for clutter). Clearly, these factors need to be taken account of in any future work on performance-relevant image based factors. When looking at the influence on detection performance of all image based factors together, there is no statistically significant effect of bag size. When using a more sophisticated model of data analysis including main effects of

FTI view difficulty, superposition, bag complexity, bag size and the interactions of these variables, there is a small effect of bag size. In Experiment 2 we were also 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 which 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 per passenger bag. More recent technology can provide 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 another one. Furthermore, training is an important tool in lessening detrimental effects on detection performance of difficult views. Our ANCOVA analysis has supported earlier findings that training leads to particularly large improvements in detection performance for difficult views (Koller, Hardmeier, Michel, & Schwaninger, in press).

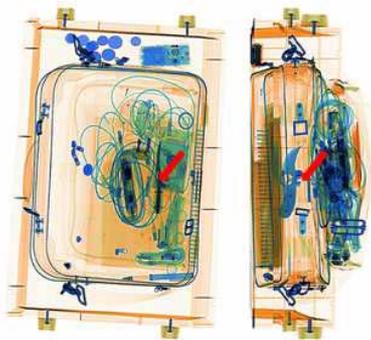


Fig. 15. Illustrative example of how multi-view systems can help improving detection performance in spite of undesirable view difficulty and superposition effects.

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 equipment that warn the

human operator (x-ray screener) with a "dark alarm", which would be triggered by opaque areas that are deemed too large or dense for unassisted human interpretation. Manual search would follow when a dark alarm was indicated.

C. Screener Selection and Training

A very important approach to face the problem of improving threat detection performance in x-ray screening consists in screener selection and screener training. The psychological literature provides evidence that figure ground segregation (related to superposition) as well as mental rotation (related to view difficulty) are visual abilities that 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 to increase detection performance of screeners substantially.

In addition to stable abilities, there are several aspects of visual knowledge relevant to x-ray image interpretation. Knowledge based factors such as knowing which objects are dangerous or prohibited and what they look like in x-ray images are trainable. Training also has beneficial effects on screeners' abilities to deal with certain image based factors. For example, training particularly improves the ability to deal with difficult views. 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/F/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

- [1] S. Koller, D. Hardmeier, F. Hofer, and A. Schwaninger, "Investigating training, transfer, and viewpoint effects resulting from recurrent cbt of x-ray image interpretation," in *Journal of Transportation Security*, in press.
- [2] A. Bolfling and A. Schwaninger, "Measurement formulae for image-based factors in x-ray imagery," in *VICOREG Technical Report, November 26, 2007*.
- [3] A. Schwaninger, S. Michel, and A. Bolfling, "A statistical approach for image difficulty estimation in x-ray screening using image measurements," in *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization*, ACM Press, New York, USA, 2007, pp. 123-130.
- [4] D. Hardmeier, F. Hofer, and A. Schwaninger, "The role of recurrent cbt for increasing aviation security screeners' visual knowledge and abilities needed in x-ray screening," in *Proceedings of the 4th International Aviation Security Technology Symposium, Washington, D.C., USA, November 27 - December 1, 2006b*, pp. 338-342.

- [5] —, "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, ICRAAT 2006, Belgrade, Serbia and Montenegro, June 24-28, 2006a*, pp. 393–397.
- [6] F. Hofer, D. Hardmeier, and A. Schwaninger, "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, 2006*, pp. 303–308.
- [7] A. Schwaninger, D. Hardmeier, and F. Hofer, "Aviation security screeners visual abilities & visual knowledge measurement," in *IEEE Aerospace and Electronic Systems*, vol. 20(6), 2005, pp. 29–35.
- [8] J. Bortz, *Statistik für Human- und Sozialwissenschaftler*, 6th ed. Wien, New York: Springer, 2004.
- [9] H. Coolican, *Research Methods and Statistics in Psychology*, 4th ed. London: Hodder & Stoughton, 2004.
- [10] A. Schwaninger, "Evaluation and selection of airport security screeners," in *AIRPORT*, vol. 2, 2003b.
- [11] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, New Jersey: Lawrence Earlbaum Associates, 1988.
- [12] F. Hofer and A. Schwaninger, "Using threat image projection data for assessing individual screener performance," in *AIRPORT*, vol. 2, 2003b, pp. 417–426.
- [13] —, "Reliable and valid measures of threat detection performance in x-ray screening," in *IEEE ICCST Proceedings*, vol. 38, 2004, pp. 303–308.
- [14] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.

APPENDIX

For more detailed information on the concepts and mathematical models of the image based factors as well as for examples for a better understanding of the formulae refer to the on-line technical documentation by Bolting, & Schwaninger (2007) made available at:

http://www.psychologie.uzh.ch/vicoreg/publications/index_byarea.htm

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 (indices N_{OV}) containing the same FTI object (indices O) in the same view (subindices 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.

$$FtiVD_{OVj} = \frac{\sum_{i=1, j \neq i}^{N_{OV}} (max(DetPerf) - DetPerf_{OV_i})}{N_{OV} - 1} \quad (1)$$

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.

$$FtiVD_{OV} = \frac{\sum_{i=1}^{N_{OV}} MissRate_{OV_i}}{N_{OV}} \quad (2)$$

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.

$$SP = C - \sqrt{\sum_{x,y} (I_{SN}(x,y) - I_N(x,y))^2} \quad (3)$$

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. \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 6). 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.

$$CL = \frac{\sum_{x,y} I_{hp}(x,y)}{BS} \quad (4)$$

$$\text{where } I_{hp}(x,y) = I_N * \mathcal{F}^{-1}(hp(f_x, f_y)) \\ = \mathcal{F}^{-1}(\mathcal{F}(I_N \cdot hp(f_x, f_y)))$$

$$\text{and } hp(f_x, f_y) = 1 - \frac{1}{1 + \left(\frac{\sqrt{f_x^2 + f_y^2}}{f}\right)^d}$$

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).

$$OP = \frac{\sum_{x,y} (I_N(x,y) < 64)}{BS} \quad (5)$$

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.

$$BS = \sum_{x,y} (I_N(x,y) < 254) \quad (6)$$