Aviation Security Screeners
Visual Abilities & Visual Knowledge Measurement

Adrian Schwaninger, Diana Hardmeier & Franziska Hofer
University of Zurich, Switzerland

ABSTRACT

A central aspect of airport security is reliable detection of forbidden objects in passenger's bags using X-ray screening equipment. Human recognition involves visual processing of the X-ray image and matching items with object representations stored in visual memory. Thus, without knowing which objects are forbidden and what they look like, prohibited items are difficult to recognize (aspect of visual knowledge). In order to measure whether a screener has acquired the necessary visual knowledge, we have applied the prohibited items test (PIT). This test contains different forbidden items according to international prohibited items lists. The items are placed in X-ray images of passenger bags so that the object shapes can be seen relatively well. Since all images can be inspected for 10 seconds, failing to recognize a threat item can be mainly attributed to a lack of visual knowledge.

The object recognition test (ORT) is more related to visual processing and encoding. Three image-based factors can be distinguished that challenge different visual processing abilities. First, depending on the rotation within a bag, an object can be more or less difficult to recognize (effect of viewpoint). Second, prohibited items can be more or less superimposed by other objects, which can impair detection performance (effect of superposition). Third, the number and type of other objects in a bag can challenge visual search and processing capacity (effect of bag complexity). The ORT has been developed to measure how well screeners cope with these image-based factors. This test contains only guns and knives, placed into bags in different views with different superposition and complexity levels. Detection performance is determined by the ability of a screener to detect threat items despite rotation, superposition and bag complexity. Since the shapes of guns and knives are usually well-known even by novices, the aspect of visual threat object knowledge is of minor importance in this test.

A total of 134 aviation security screeners and 134 novices participated in this study. Detection performance was measured using A'. The three image-based factors of the ORT were validated. The effect of view, superposition, and bag complexity were highly significant. The validity of the PIT was examined by comparing the two participant groups. Large differences were found in detection performance between screeners and novices for the PIT. This result is consistent with the assumption that the PIT measures aspects related to visual knowledge. Although screeners were also better than novices in the ORT, the relative difference was much smaller. This result is consistent with the assumption that the ORT measures image-based factors that are related to visual processing abilities; whereas the PIT is more related to visual knowledge. For both tests, large inter-individual differences were found. Reliability was high for both participant groups and tests, indicating that they can be used for measuring performance on an individual basis. The application of the ORT and PIT for screener certification and competency assessment are discussed.

INTRODUCTION

The importance of aviation security has changed dramatically in the last years. As a consequence of the new threat situation large investments into technology have been made. State-of-the-art X-ray machines provide high resolution images, many image enhancement features and even automatic detection of explosive material. However, it is becoming clear since recently that the best technology is only as valuable as the humans that operate it. Indeed, reliable recognition of threat items in X-ray images of passenger bags is a demanding task. Consider the images depicted in Figure 1. Each of the two bags contains a threat item that could be used to severely harm

This research was financially supported by Zurich Airport Unique, Switzerland.

Author's Current Address:
University of Zurich, Switzerland, Psychologisches Institut, Visual Cognition Research Group (VICOREG), Klosbachstr. 107, 8032 Zurich, Switzerland.

Based on a presentation at Carnahan 2004.

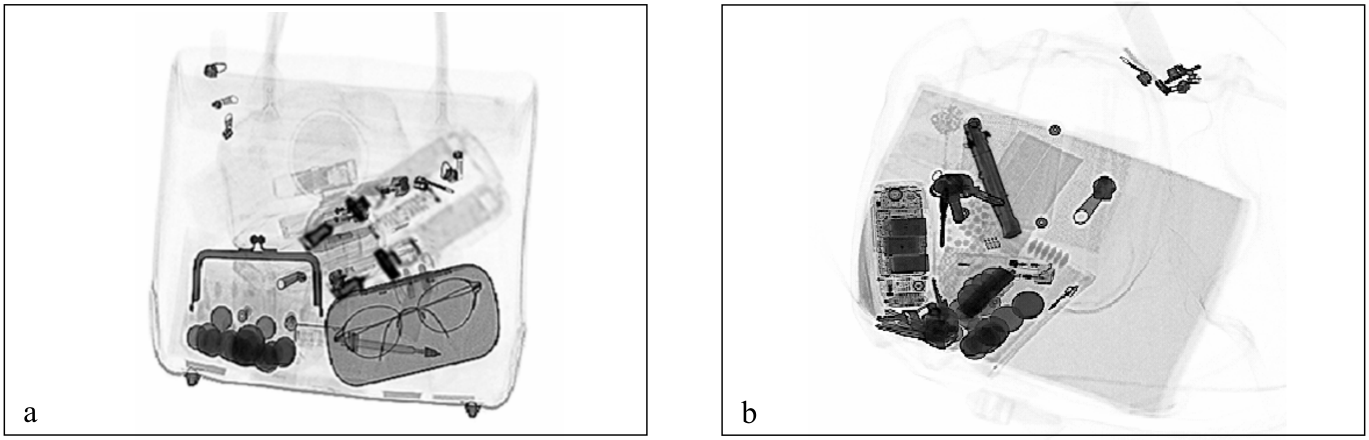

Fig. 1. Examples of prohibited items in X-ray images of passenger bags

Fig. 1A. Gas spray in the center of the baggage below the eyeglasses;
Fig. 1B. Switchblade knife slightly above the center of the baggage next to the keys

people. Even though most people would probably recognize prohibited items like the gas spray in Figure 1A when depicted in a photograph, this and other threat objects are relatively hard to recognize for novices because the shape features look quite different in an X-ray image than in reality. Other dangerous items (e.g., the switchblade knife in Figure 1B) might be missed by a novice because they look similar to harmless objects (e.g. a pen). Several other threat objects are usually not encountered in real life (e.g., improvised explosive devices, IEDs), which stresses the importance of computer-based training in order to achieve a high detection performance within a few seconds of inspection time [1].

In short, the knowledge about which items are prohibited and what they look like in an X-ray image is certainly an important determinant for detection performance. The Prohibited Items Test (PIT) has been developed to measure this knowledge-based component and it therefore contains a large number of different forbidden objects according to international prohibited items lists [2].

As pointed out by [3] several image-based effects influence how well threat items can be recognized in X-ray images (Figure 2). Viewpoint can strongly affect recognition performance, which has been shown previously in many object recognition studies (for reviews see [4 - 7]). Since objects are often superimposed on each other in X-ray images, the degree of superposition can affect detection performance substantially. Another image-based factor is bag complexity, which is determined by the type and number of objects in a bag.

The Object Recognition Test (ORT) has been developed to measure how well screeners can cope with such image-based factors [8]. In order to reduce effects of visual knowledge, only guns and knives are used in this test, i.e., object shapes that are usually well known also by novices.

The purpose of this study is to investigate the role of image-based and knowledge-based factors in X-ray screening using these two different tests. To what extent screeners know

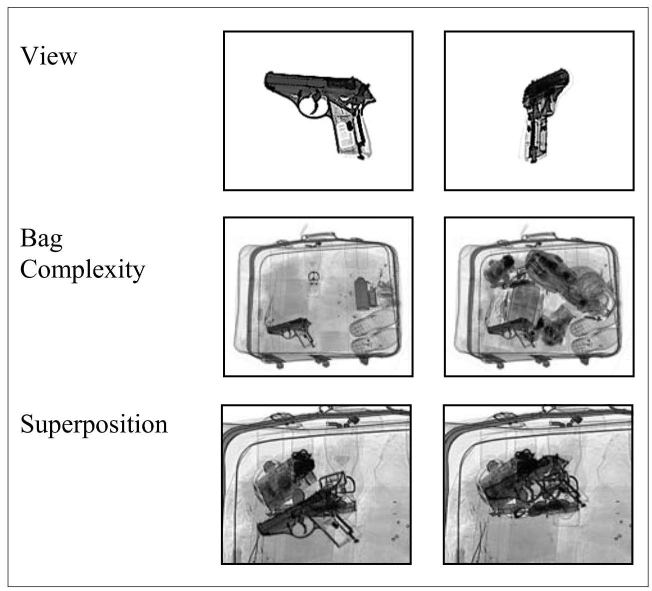

Fig. 2. Image-based factors according to [3];

Fig. 2A. Viewpoint of the threat item (canonical vs. non-canonical);
Fig. 2B. Bag complexity (low vs. high);
Fig. 2C. Superposition of the threat item (low vs. high)

which items are prohibited and what they look like in passenger bags is measured by the PIT. It includes prohibited items of different categories in X-ray images of passenger bags while keeping effects of view, superposition, and bag complexity relatively constant. The objects are displayed in an easy view with a moderate degree of superposition in bags of limited complexity during 10 seconds per image. If a participant fails to detect a threat item it is therefore rather related to a lack of visual knowledge than to an attentional failure or visual processing capacity limitations. Since many different prohibited items with shapes that are often not known from everyday experience are used in the PIT, a substantial difference in detection performance between novices and screeners could be expected. The ORT measures how well someone can cope with image-based factors such as view, superposition, and bag complexity. As mentioned above, only guns and knives are used in this test, i.e., object shapes that are well known by both screeners and novices. Therefore, smaller differences between screeners and novices might be expected for the ORT compared to the PIT. However, expertise might increase visual abilities that are necessary in order to cope with image difficulty resulting from effects of viewpoint, superposition, and bag complexity. Therefore, the effect size of the interaction between image-based effects and expertise is an important measure in this study as well.

METHOD

Participants

A total of 268 participants took part in this study. Half were aviation security screeners, the other half were novices.

All participants were tested with the ORT and then the PIT. The screener group consisted of 67 females and 67 males at the ages of 24 and 57 years (M = 41.05 years, SD = 7.84 years). All had undergone initial classroom and on the job training and they had at least two years of work experience in airport security screening of carry-on bags.

The novices group consisted of 134 males between 21 and 26 years (M = 23.24 years, SD = 1.22 years).

Materials and Procedure

Prohibited Items Test (PIT)

This test contains a wide spectrum of prohibited items which can be classified into seven categories according to international prohibited items lists [2]. The PIT version used in this study included a total of 19 guns, 27 sharp objects, 14 blunt and hunt instruments, 5 highly inflammable substances, 17 explosives, 3 chemicals, and 13 other prohibited items (e.g., buckshot, ivory). All prohibited items were depicted from an easy viewpoint and combined with a bag of medium complexity and low superposition, so their shapes could be seen relatively well and the influence of image-based factors could be minimized. X-ray images were taken from Heimann 6040i machines and displayed in color. 68 bags contained one threat item, 6 bags contained two threat items, and 6 bags contained three threat items. Each bag was shown twice resulting in a total of 160 trials. There were four blocks of 40 trials. Block order was counterbalanced across four groups of participants using a Latin Square design. Trial order was randomized within each block. Only responses to images containing one threat item were used for statistical analyses.

The PIT is fully computer-based and starts with a self-explanatory instruction, followed by a brief training session with eight examples to familiarize the participants with the procedure. Feedback is provided after each trial only in the introductory phase. Each X-ray image was displayed for a maximum of 10 seconds in the introductory and test phases. This duration is long enough to ensure that missing a threat item can be mainly attributed to a lack of visual knowledge rather than a failure of attention. For each image, participants had to decide whether the bag was OK (no threat) or NOT OK (threat) and indicate on a slider how sure they were in their decision (confidence ratings on a 50 point scale). In addition, participants had to indicate the threat category of the prohibited item(s) by clicking the corresponding buttons on the screen (for NOT OK decisions only). Pressing the space bar displayed the next image. As the test was subdivided into four blocks, participants were allowed to take a short break after a block was completed.

Object Recognition Test (ORT)

As explained in the introduction, [3] pointed out that image-based factors such as viewpoint, superposition, and bag complexity can substantially affect detection performance in X-ray images. The ORT has been designed to measure how well people can cope with such image-based factors rather than measuring knowledge-based determinants of threat detection performance [8]. To this end, guns and knives with the blade open are used in the ORT, i.e., object shapes that can be assumed to be known by most people. In addition, all guns and knives are shown for 10 seconds before the test starts, which further reduces the role of knowledge-based factors in this test.

In reality, a threat object can be depicted from a difficult viewpoint in a close-packed bag and be superimposed by other objects. The X-ray images used in the ORT vary systematically in image difficulty by varying the degree of view difficulty, bag complexity, and superposition, both independently and in combination. This makes it possible to investigate main effects as well as interactions between the image-based factors. All X-ray images of the ORT are in black-and-white, as color, per se, is mainly diagnostic for the material of objects in the bag, and thus, could be primarily helpful for experts.

Eight guns and eight knives with common shapes were used. Each gun and each knife was displayed in an easy view and a rotated view to measure the effect of viewpoint. In order to equalize image difficulty resulting from viewpoint changes, guns were more rotated than knives based on results of a pilot study. Each view was combined with two bags of low complexity: once with low superposition; and once with high superposition. These combinations were also generated using two closed-packed bags with a higher degree of bag complexity. In addition, each bag was presented once with and once without the threat item. Thus, there were a total of 256 trials: 2 weapons (guns, knives) * 8 (exemplars) * 2 (views) * 2 (bag complexities) * 2 (superpositions) * 2 (harmless vs. threat images). There were four blocks of 64 trials each. The order of blocks was counterbalanced across four groups of participants using a Latin Square. Within each block the order of trials was random.

The ORT is fully computer-based. After task instructions an introductory session followed using 2 guns and 2 knives not displayed in the test phase. Feedback was provided after each trial but only in the introductory phase. Prior to the test phase, the eight guns and eight knives used at test were presented for 10 seconds, respectively. Half of the guns and knives were shown in an easy view and half were depicted in a rotated view. At test, each object was presented in the easy and the rotated view with low and high superposition and with low and high bag complexity. Each image was displayed for 4 seconds. This duration was chosen to match the demands of high passenger flow where average X-ray image inspection time at checkpoints is in the range of 3-5 seconds. For each X-ray image, participants had to decide whether the X-ray images contained one of the guns or knives shown in the introductory phase or not (NOT OK or OK response). Confidence ratings had to be provided by changing the position of a slider (90 point scale). The next trial was started by pressing the space bar. Short breaks were possible after completing one of the four blocks.

RESULTS

It is important to take the hit rate as well as the false alarm rate into account if threat and non-threat images are used in a computer-based test requiring OK and NOT OK responses. The reason is simple: A candidate could achieve a hit rate of 100% simply by judging all bags as being NOT OK. Whether a high hit rate reflects good visual detection performance, or just a lenient response bias, can only be determined if the false alarm rate is considered, too. Psychophysics provides several methods in order to derive more valid estimates based on hit and false alarm rates. A well-known measure from signal detection theory is d' [9]. It equals z(H) – z(FA) whereas H denotes the hit rate, FA the false alarm rate and z represents the transformation into zscores (standard deviation units). An often used "non-parametric" measure is A' [10]. This measure represents an estimate of the area under an ROC curve that is specified by only one data point. More specifically, A' corresponds to the average area for the two linear ROC curves that maximize and minimize the hit rate. The term "non-parametric" is a bit misleading because it only refers to the fact that the computation of A' doesn't require an priori assumption about the underlying distributions [11, 12]. This has sometimes been regarded as an advantage over SDT measures such as d' and Δm (for a more detailed discussion of this issue. See also [13]). Although only A' data are reported in this study, it should be stressed that similar results were obtained for d' data. Moreover, correlations between A' and d' were very high for both tests and screeners groups (ORT: $r = .94$ for screeners and $r = .97$ for novices, PIT: $r = .95$ for screeners and $r = .98$ for novices, all $p < .001$).

The results section is organized as follows: first, ANOVA results of the ORT are presented. These analyses were conducted to investigate whether detection performance of aviation security screeners and novices is affected by image-based factors. In addition, the effect of expertise on the three image-based factors measured by the ORT was examined. Second, overall detection performance in the ORT is compared to overall detection performance in the PIT[1]. More specifically, the effect of expertise on image-based factors and knowledge-based factors is analyzed, comparing detection performance of aviations security screeners with that of novices in the two tests. Finally, the results of reliability analyses are presented which were conducted to evaluate whether the ORT and PIT can be used for measuring detection performance on an individual basis.

ORT and Abilities to Cope with Image-Based Factors

A' scores calculated from hit and false alarm rates of the ORT were subjected to three-way analyses of variance (ANOVA) with the three within-participants factors view, bag complexity, and superposition. Results of aviation security screeners show that there were significant main effects of view (easy vs. rotated) with an effect size of $\eta^2 = .71$, $F(1, 133) = 318.59$, $MSE = 0.003$, $p < .001$, bag complexity (low vs. high) $\eta^2 = .83$, $F(1, 133) = 652.96$, $MSE = 0.003$, $p < .001$, and superposition (low vs. high) $\eta^2 = .61$, $F(1, 133) = 203.73$, $MSE = 0.003$, $p < .001$. The following two-way interactions were significant: View * superposition, $\eta^2 = .12$, $F(1, 133) = 17.91$, $MSE = 0.002$, $p < .001$, bag complexity * superposition $\eta^2 = .12$, $F(1, 133) = 18.22$, $MSE = 0.002$, $p < .001$. Note however, that the effect sizes of these interactions are rather low when compared to the effect sizes of the main effects. All other interactions were not significant. In short, there were clear main effects of view, bag complexity, and superposition with very large effect sizes (see also conventions by [14]). Some interactions reached statistical significance, but the effect sizes were relatively small when compared to the effect sizes of the main effects.

Similar results could be observed for novices. Again, there were significant main effects of view (easy vs. rotated) $\eta^2 = .76$, $F(1, 133) = 428.33$, $MSE = 0.005$, $p < .001$, bag complexity (low vs. high) $\eta^2 = .72$, $F(1, 133) = 333.14$, $MSE = 0.005$, $p < .001$, and superposition (low vs. high) $\eta^2 = .63$, $F(1, 133) = 228.09$, $MSE = 0.004$, $p < .001$. All two-way interactions were significant: View * bag complexity $\eta^2 = .06$, $F(1, 133) = 9.07$, $MSE = 0.004$, $p < .01$, view * superposition $\eta^2 = .07$, $F(1, 133) = 10.43$, $MSE = 0.004$, $p < .01$, bag complexity * superposition $\eta^2 = .15$, $F(1, 133) = 23.15$, $MSE = 0.004$, $p < .001$. The three-way interaction between view, bag complexity and superposition also reached statistical significance, $\eta^2 = .03$, $F(1, 133) = 4.14$, $MSE = 0.004$, $p < .05$. As for screeners, very large effect sizes were found for main effects whereas the interactions showed much smaller effect sizes.

Figure 3 shows the main effects of each of the three image-based factors, averaged across the other two factors. A comparison of Figure 3A (aviation security screeners) and Figure 3B (novices) reveals that screeners were slightly better than novices while both screener groups are substantially affected by the image-based factors view, bag complexity, and superposition. In order to examine whether expertise has a

[1] A' scores for the PIT were calculated using the responses to images of the following categories: guns, sharp objects, hunt and blunt instruments.

differential effect on these image-based factors, a four-way analysis of variance (ANOVA) with the within-participants factors view, bag complexity, and superposition and the between-participant factor expertise was computed. There were again significant main effects of view (easy vs. rotated) $\eta^2 = .74$, $F(1, 266) = 744.57$, $MSE = 0.004$, $p < .001$, bag complexity (low vs. high) $\eta^2 = .77$, $F(1, 266) = 884.75$, $MSE = 0.004$, $p < .001$, and superposition (low vs. high) $\eta^2 = .62$, $F(1, 266) = 428.20$, $MSE = 0.004$, $p < .001$. Two-way interactions between view and bag complexity $\eta^2 = .04$, $F(1, 266) = 10.23$, $MSE = 0.003$, $p < .01$, view and superposition $\eta^2 = .09$, $F(1, 266) = 26.17$, $MSE = 0.003$, $p < .001$, view and expertise $\eta^2 = .10$, $F(1, 266) = 30.52$, $p < .001$, and superposition and expertise $\eta^2 = .03$, $F(1, 266) = 9.39$, $p < .01$ were significant, as well as the tree-way interactions between view, bag complexity, and superposition $\eta^2 = .02$, $F(1, 266) = 5.47$, $MSE = 0.003$, $p < .05$, and bag complexity, superposition and expertise $\eta^2 = .13$, $F(1, 266) = 41.13$, $p < .001$. Although these interactions were significant, all have relatively low effect sizes when compared to the main effects. All other interactions were not significant.

In short, these results indicate that the effects of image-based factors are apparent for novices as well as for aviation security screeners and expertise does only slightly reduce these effects of view, bag complexity, and superposition.

PIT, Visual Knowledge and Expertise

In contrast to the ORT, the PIT has been developed to measure whether screeners know which items are prohibited and how they look in X-ray images of passenger bags [2]. Whereas in the ORT only guns and knives are used – object shapes that are also familiar to novices – the PIT contains all kinds of forbidden objects based on international prohibited items lists. In this test, all target objects are shown in an easy viewpoint with a moderate degree of superposition in bags of moderate bag complexity. As mentioned above, each image is shown for 10 seconds and therefore missing a threat item in the PIT can rather be attributed to a lack of visual knowledge than to an attentional failure or visual processing capacity limitations. If detection performance in the PIT is indeed mainly determined by visual experience and training with X-ray images, large differences between novices and aviation security screeners should be observed in this test. As reported in the previous section, only moderate differences between novices and screeners were found for the ORT.

In order to compare relative difference between experts and novices for the PIT and ORT, overall hit and false alarm rates were used to compute relative detection performance difference separately for the ORT and PIT using the following formula:

$$(A'_{experts} - A'_{novices}) / A'_{novices}$$

Relative detection performance difference between experts and novices was indeed much higher for the PIT than for the ORT (15.89% vs. 6.05%). This is consistent with the view that the PIT measures visual knowledge dependent on training and expertise, whereas the ORT measures more stable visual abilities used to cope with image-based factors such as effects of view, bag complexity, and superposition.

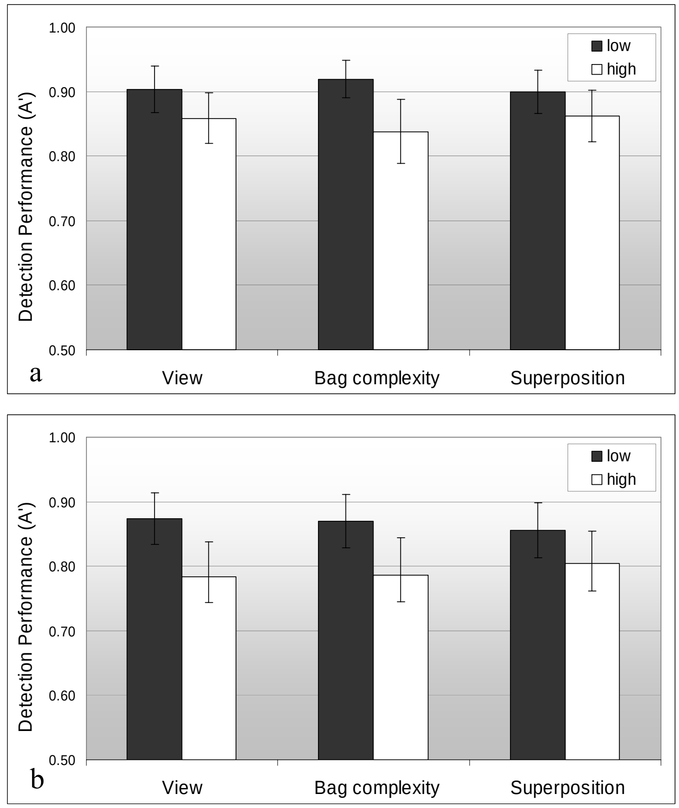

Fig. 3. Detection performance (A') in the ORT with standard deviations;

Fig. 3A. For aviations security screeners; Fig. 3B. For novices

This main finding was further analyzed using a two-way analysis of variance (ANOVA) with the within-participant factor test type (ORT, PIT) and the between-participant factor expertise using overall A' scores from each test. There was a significant effect of test type (ORT vs. PIT) $\eta^2 = .80$, $F(1, 266) = 1075.10$, $MSE = 0.002$, $p < .001$, a significant effect of expertise (experts vs. novices) $\eta^2 = .44$, $F(1, 266) = 206.11$, $MSE = 0.004$, $p < .001$, and a significant interaction of test type and expertise $\eta^2 = .20$, $F(1, 266) = 65.30$, $p < .001$. The interaction between test type and expertise is consistent with the hypothesis that the ORT measures rather image-based factors whereas the PIT measures rather knowledge-based factors.

It must also be noted however, that correlation analyses showed that the two tests are far from being orthogonal. Overall detection performance A' of the two tests correlates with $r = .51$, $p < .001$ for experts, and $r = .42$, $p < .001$ for novices. This could at least indicate that detection performance in PIT is not only determined by visual knowledge but also by visual abilities used to cope with image-based factors as measured by the ORT.

Table 1. Reliability Analyses

		Reliability Coefficients	PC SN	PC N	CR SN	CR N
PIT	Screeeners	Cronbach Alpha	.840	.878	.887	.924
		Split-half	.811	.915	.859	.948
	Novices	Cronbach Alpha	.871	.877	.885	.914
		Split-half	.882	.862	.883	.890
ORT	Screeners	Cronbach Alpha	.862	.934	.902	.962
		Split-half	.733	.813	.792	.887
	Novices	Cronbach Alpha	.899	.910	.916	.959
		Split-half	.778	.810	.759	.907

NOTE: Cronbach Alpha values and split-half reliabilities (Guttman) for both tests in each group (experts and novices separately) calculated for percentage correct (PC) and confidence ratings (CR) separately for signal plus noise (SN) and noise trials (N).

One potential argument against the analyses of this section could be that the expert group consisted of males and females, whereas the novices group consisted only of males. However, it is unlikely that gender effects can explain the differences found between experts and novices since no significant differences were found between male and female screeners, neither for the ORT (p = .70) nor for the PIT (p = .78).

Reliability Analyses

Internal reliability was analyzed using Cronbach's Alpha and Guttman split-half coefficients separately for both participant groups (aviation security screeners and novices). Analyses were computed for signal plus noise trials (bags including a threat item) and noise trials (harmless bags), respectively. Reliability coefficients were computed on the basis of the percentage correct measures (i.e., hit and correct rejections), as well as on the basis of the screeners' confidence ratings (CR) for hits and correct rejections. As can be seen in Table 1 high reliability coefficients were found for both tests and participant groups.

The results section has clearly shown that item difficulty in the ORT depends on the main effects and interactions between view, bag complexity, and superposition. Therefore, the high internal consistency also found for the ORT is a nice example for the fact that a test can be homogenous and multifactorial (see [15]).

DISCUSSION

The objective of this study was to examine the role of image-based and knowledge-based factors for detecting threat items in passenger bags. As pointed out by [3], image-based factors such as effects of viewpoint, bag complexity, and superposition can substantially affect detection performance. The ORT has been developed to measure how well a participant can cope with these image-based factors [8]. This test contains guns and knives depicted in an easy and difficult view shown in bags with low and high bag complexity while being strongly or little superimposed by other objects. Main effects with large effect sizes were found for aviation security screeners as well as novices. While screeners achieved a moderately better detection performance in the ORT, they were still significantly affected when threat items were rotated, superimposed by other objects, or shown in complex bags. This result is consistent with the view that the ORT does measure visual abilities necessary to cope with image difficulty resulting from effects of viewpoint, bag complexity and superposition. Large inter-individual differences were found both for novices as well as experts. Internal reliability was very high for both groups. Therefore, this test could be a useful tool both for competency assessment of screeners as well as for pre-employment assessment purposes.

The PIT has been developed to measure whether a screener knows which items are prohibited and what they look like in X-ray images of passenger bags [2]. In this test, all objects are depicted in an easy view. Bag complexity and superposition are moderate so that the threat item shapes are visible. Images are shown for 10 seconds, i.e., missing a threat item can be attributed to a lack of visual knowledge rather than to an attentional failure or a visual processing capacity limitation. If the PIT is indeed related to visual knowledge based on expertise and training, large differences between novices and experts should be observed. Indeed, relative detection performance difference between novices and experts was about three times higher for the PIT than for the ORT. This result is consistent with the view that the PIT measures knowledge-based factors whereas the ORT measures visual abilities used for coping with image-based factors. As for the ORT, excellent reliability coefficients were found for the PIT. This test could therefore provide a useful tool for certification, competency, and risk assessment as well as for quality control in general.

In summary, the results of this study confirm that X-ray detection performance relies on visual abilities necessary for coping with image-based effects such as view, bag complexity, and superposition. Visual experience and training are necessary to know which items are prohibited and what they look like in X-ray images of passenger bags. Both aspects are prerequisites for a good screener and can be evaluated using the ORT and PIT.

ACKNOWLEDGMENT

This research was financially supported by Zurich Airport Unique, Switzerland. We are thankful to Zurich State Police, Airport Division for their help in creating the stimuli and the good collaboration for conducting the study.

REFERENCES

[1] A. Schwaninger and F. Hofer, Evaluation of CBT for increasing threat detection performance in X-ray screening, in *The Internet Society 2004, Advances in Learning, Commerce and Security*, K. Morgan and M. J. Spector, Eds. Wessex: WIT Press 2004, pp. 147-156.

[2] A. Schwaninger, *Prohibited items test (PIT): Test manual and user's guide.* Zurich: APSS.

[3] A. Schwaninger, Detection systems: Screener evaluation and selection, *AIRPORT*, Vol. 2, pp. 14-15, 2003.

[4] M.J. Tarr and H.H. Bülthoff, Is human object recognition better described by geon structural descriptions or by multiple views?, *Journal of Experimental Psychology, Human Perception and Performance*, Vol. 21(6), pp. 1494-1505, 1995.

[5] M.J. Tarr and H.H. Bülthoff, *Object recognition in man, monkey and machine*. Cambridge, Massachusetts: MIT Press, 1999.

[6] A. Schwaninger, *Object recognition and signal detection*, in Praxisfelder der Wahrnehmungspsychologie, B. Kersten and M.T. Groner, Eds., Bern: Huber, in press.

[7] M. Graf, A. Schwaninger, C. Wallraven and H.H. Bülthoff, Psychophysical results from experiments on recognition & categorisation, Information Society Technologies (IST) programme, Cognitive Vision Systems – CogVis; (IST-2000-29375), 2002.

[8] A. Schwaninger, Object recognition test (ORT): *Test manual and user's guide*. Zurich: APSS.

[9] D.M. Green and J.A. Swets, Signal detection theory and psychophysics, New York: Wiley, 1966.

[10] I. Pollack and D.A. Norman, A non-parametric analysis of recognition experiments, *Psychonomic Science*, Vol. 1, pp. 125-126, 1964.

[11] R.E. Pastore, E.J. Crawley, M.S. Berens and M.A. Skelly, Nonparametric" A' and other modern misconceptions about signal detection theory, *Psychonomic Bulletin & Review*, Vol. 10, pp. 556-569, 2003.

[12] N.A. MacMillan and C.D. Creelman, *Detection theory: A user's guide*. Cambridge: University Press, 1991.

[13] F. Hofer and A. Schwaninger, Reliable and valid measures of threat detection performance in X-ray screening, *IEEE ICCST Proceedings*, Vol. 38, pp. 303-308, 2004.

[14] J. Cohen, *Statistical power analysis for the behavioral sciences*. New York: Erlbaum, Hillsdale, 1988.

[15] P. Kline, *The handbook of psychological testing*. London: Routledge, 2000.