WHY LAPTOPS SHOULD BE SCREENED SEPARATELY WHEN CONVENTIONAL X-RAY SCREENING IS USED

Marcia Mendes
University of Applied Sciences and Arts
Northwestern Switzerland (FHNW), School of Applied Psychology (APS)
Institute Humans in Complex Systems (MikS)
Riggenbachstrasse 16
4600 Olten, Switzerland
marcia.mendes@fhnw.ch

Adrian Schwaninger, Prof. Dr.
University of Applied Sciences and Arts
Northwestern Switzerland (FHNW), School of Applied Psychology (APS)
Institute Humans in Complex Systems (MikS)
Riggenbachstrasse 16
4600 Olten, Switzerland
adrian.schwaninger@fhnw.ch

Nadine Strebel
University of Applied Sciences and Arts
Northwestern Switzerland (FHNW), School of Applied Psychology (APS)
Institute Humans in Complex Systems (MikS)
Riggenbachstrasse 16
4600 Olten, Switzerland
nadine.strebel@fhnw.ch

Stefan Michel, Dr.
University of Applied Sciences and Arts
Northwestern Switzerland (FHNW), School of Applied Psychology (APS)
Institute Humans in Complex Systems (MikS)
Riggenbachstrasse 16
4600 Olten, Switzerland
stefan.michel@fhnw.ch

Center for Adaptive Security Research and Applications (CASRA)
Thurgauerstrasse 39
8050 Zurich, Switzerland

Abstract – Security screening at airports is a highly relevant topic for more than a decade. Large amounts of money are still being invested to continuously improve the efficiency and effectiveness of passenger and baggage screening processes. State-of-the-art x-ray screening machines provide good quality images with high resolutions. Yet, the detection of prohibited items remains a challenging task for screening officers (screeners) and becomes even more difficult when dense electronic devices, such as laptops, are contained in baggage. Due to their compact construction, laptops can easily conceal other parts of luggage or could be used to hide threat items. International and national regulations specify that laptops need to be taken out of passenger bags at security checkpoints for x-ray screening, sometimes causing lower passenger throughput and even annoyance among passengers.

The aim of this study was to investigate how leaving laptops inside passenger bags actually affects detection performance of x-ray screening officers, when state-of-the-art technology providing single-view images is applied. The experiment was conducted with 40 certified screening officers from an international European airport. Participants were divided into two experimental groups. Both groups conducted an image interpretation test containing bags and laptops. For the first group, images of bags and laptops were displayed separately, while for the second group, laptops were left inside the passenger bags and displayed as one image. All participants had to judge whether the displayed bags and laptops contained a threat item or not. Threat items of four different categories were used: guns, knives, improvised explosive devices (IEDs) and others (e.g. electronic shock devices). Detection performance (A’) and reaction times were measured.

A highly significant difference in detection performance between both groups was observed. Throughout all analyses, detection performance scores were much higher when bags and laptops were x-rayed separately. Largest differences were found for the category IEDs. Altogether, the results imply that when no automatic threat detection and only single view images are available, the detection of prohibited items is still far more reliable when bags and laptops are screened separately. Index Terms — Aviation security, x-ray screening, single-view images, detection performance, laptops.

I. INTRODUCTION

The relevance of aviation security has increased dramatically since the 9/11 or more recent terrorist attacks and has become a highly relevant and ongoing topic all over the world. Airport security screening, in particular, has become an indispensable need for securing flight transportation.

Large amounts of money are being invested into the improvement and further development of screening technologies, continuously aiming at improving image quality, automatic explosive detection and enhancing the detection of threat items by human operators, i.e. screening officers. Especially the latter is of high importance, as the actual identification of threat items in cabin and hold baggage screening still depends on human operators, who judge the images based on the outcome of the technology applied ([1]; [2]).

[3] and [4] have shown that three image-based factors influence the difficulty for screening officers to detect threat items in x-ray images. These are view difficulty (viewpoint) of threat items, their superposition by other objects and the complexity of a bag, i.e. the number and type of objects contained in the bag. The latter comprises clutter (texture unsteadiness) and transparency/opacity (relative size of dark areas) (for more information see [5]; [6]; [7]). Depending on material density, the amount to which x-rays are able to penetrate various objects in the bag will vary.
Thus, the already challenging task of detecting threat items in x-ray images of passenger bags becomes even more difficult when certain objects, such as large electronic devices, are contained in the baggage. Due to their compact construction, electronic devices, such as laptops, are comparatively dense and hard to penetrate. Objects in front or behind such devices become difficult for screening officers to see, and hence, they could be used to conceal other parts of luggage or hide threat items (e.g., an improvised explosive device, IED).

Especially when single view x-ray systems are used or even multi-view systems, where the additional views do not provide enough meaningful information, threat items which are behind or in front of a laptop or that are hidden inside a laptop can become very difficult or even impossible for human operators to recognize [8]. At present, most machines deployed at airports provide single view images. A human operator will only be able to identify a threat item and make a correct decision, if the threat can be recognized (e.g., recognizable viewpoint) in the provided single view image [11; 3; 4].

Considering the above mentioned image-based factors and the density of electronic devices, it becomes evident why currently most international and national regulations specify that portable computers and other large electronic devices shall be removed from passenger bags and screened separately at security checkpoints (e.g., the current regulation of the European Commission [9]).

Yet, having to take items, such as laptops, out of the baggage can sometimes reduce passenger throughput at security checkpoints and provoke annoyance amongst passengers, who perhaps do not understand the necessity of this regulation. With state-of-the-art x-ray screening equipment being able to provide high quality colored images that distinguish between different materials (see [10]; [11]) and offering high image resolution, the question arises how big the advantage of separate screening actually is. Not having to unpack the baggage would certainly increase comfort for passengers and possibly increase throughput at security checkpoints.

The aim of this study was to investigate to what extent detection performance of screening officers is impaired when laptops are left inside passenger bags for security screening and when state-of-the-art technology providing single-view x-ray images without automatic detection is applied. An image interpretation test containing bags and laptops was created in two different versions. In one version x-ray images of bags and laptops were displayed separately, in the other version laptops were left inside the bags and displayed as one x-ray image. Two experimental groups with screening officers of one international European airport were formed, each group conducting one of the test versions. Detection performance scores and reaction times of both groups were compared to evaluate the effect of the two different packing conditions.

II. METHOD AND PROCEDURE

A. Participants

The study was conducted with airport security screening officers employed at an international European airport. All participants were certified screeners, meaning they were all qualified, and were trained and certified according to the standards set by the national appropriate authority (civil aviation administration). Altogether, 40 screening officers participated in this study, which were distributed into two different experimental groups (A and B, 20 per group). See Figure 1 for the experimental design. In order to control that both experimental groups were equally balanced with regard to the screeners’ x-ray image interpretation competency, test results from regular certification processes were used and compared. The test used for certification of screeners at this airport is the X-Ray Competency Assessment Test (X-Ray CAT). This computer-based test is a standardized instrument to measure x-ray image interpretation competency of airport security screening officers and has been used in several scientific studies ([2]; [12]; [13]). It is used for screener certification at several European airports. The test consists of 256 trials based on 128 different color x-ray images of passenger bags. All bag images are once displayed without (non-threat image) and once containing a threat object (threat image). For more information on the X-Ray CAT see [12]. An independent samples t-test with the X-Ray CAT results of group A and B revealed no significant difference between both groups (t[38] = 1.81, p = .078, d = 0.57).

The average age of the participants was 40.68 years (SD = 10.67), with a range between 22-56 years. 62.5% of the participants were female and 37.5% male. The average amount of job experience was M = 5.13 years (SD = 4.02, range: 1.5 – 20 years).

B. Procedure

1) Experimental Design: Both experimental groups conducted a computer-based x-ray image interpretation test. During the test, color x-ray images of passenger bags and laptops were displayed, sometimes containing threats (threat images) and sometimes without any threat items (non-threat images). All participants had to view every image and decide whether the bags and laptops could be regarded as harmless (OK) or whether they contained a threat item (NOT OK). Both test conditions differed with regard to the packing condition (laptops inside vs. outside passenger bags). Figure 1 shows the experimental design of the study with the two conditions A and B to be compared between groups.

Fig. 1: Experimental design (between-subjects design) for the comparison of both conditions (A, B).

In both test conditions the same bags were presented to the screening officers. Originally, every bag contained a laptop. In condition A the laptops were taken out of the bag and screened separately (two separate x-ray images), whereas in condition B the laptops were left inside the passenger bags (one x-ray image). See Figure 2 for an illustration of the two packing
Image Interpretation Test: The image interpretation test was based on a representative set (defined by airport security screening experts) of 96 passenger bags, all of which originally contained laptops. The test images were created and recorded in collaboration with aviation security experts from a specialized police organization and former airport security screening officers of CASRA. As explained above, in condition A the laptops were taken out of the bags and displayed separately, whereas in condition B the laptops were left inside the bags. Each bag/laptop-combination was used twice, once with a threat item included in either the bag or the laptop, and once without any threat item. The threat items contained in this test consisted of a representative sample of objects selected and built (IEDs in laptops) by screening experts from an airport police department. These threat items could be divided into four different threat categories: guns, improvised explosive devices (IEDs), knives and other threat items (e.g., electric shock devices, etc.). For all categories except “guns”, in half of the cases the threat items were placed in the bag, while in the other half of the cases the threat items were placed inside the laptop (see Figure 3). Due to their size, it would not have been realistic to place guns inside a laptop.

Each threat category contained 24 items. Therefore, the number of test images for the conditions were the following:

- **Test A (laptops and bags screened separately):**
  - 4 x 24 threat images (60 bags and 36 laptops)
  - 96 non threat BAG images
  - 96 non threat LAPTOP images
  - Total: 288 test images

- **Test B (laptops inside bags):**
  - 4 x 24 threat images (60 bags and 36 laptops)
  - 96 non threat images (combined)
  - Total: 192 test images

All participants were invited to the experimenters’ facilities to conduct the test. Four computer workstations were set up in a normally lit room using a 19” TFT monitor. X-ray images covered about 2/3 of the computer screen. Tests were conducted quietly and individually, and under supervision. The test images remained on the computer screen until the participant either pressed the "OK" or "NOT OK" button. There was no time limit set for viewing an image; yet, participants were instructed to inspect the images as quickly and accurately as possible. In order to avoid eyestrain and fatigue, and to make sure that those participants conducting the tests with more images (288 images to 192 images) would not become too tired towards the end, breaks of 10 minutes were taken in 30 minute-cycles. All participants completed the test in less than two hours, including breaks.

C. A’ as Measure of Detection Performance

Measuring detection performance in a valid way must be based on a person’s hit rate as well as the false alarm rate. The hit rate on its own would not be a valid measure, as a candidate could achieve the highest hit rate by judging all bags as NOT OK. In our analyses we applied A’ ([14]) to measure detection performance in terms of sensitivity. A’ is a non-parametric measure, meaning that its computation does not require any a priori assumptions about the underlying signal and noise distributions. A’ can be calculated by the following formula ([15]):

\[
A' = \frac{1}{\frac{1}{2} + \left(\frac{H - F}{1 + H - F}\right)\left(1 + \frac{H - F}{1 - F}\right)}
\]

H is the hit rate and F the false alarm rate. If performance is below chance, i.e. when H<F, the equation must be modified ([16]):

\[
A' = \frac{1}{\frac{1}{2} + \left(\frac{F - H}{1 + F - H}\right)\left(1 + \frac{F - H}{1 - H}\right)}
\]

Even though the computation of A’ requires no a priori assumption about underlying distributions, it does not mean that this measure is an accurate reflection of its theoretical origin (i.e., that A’ reflects the area under a reasonable ROC curve) or that it is a distribution-free measure and fully independent of response bias (see [17]). However, due to its easy computation and interpretation, A’ is often applied as a measure for detection performance in research and application. In this paper, actual performance values are not reported due to security reasons. For all relevant analyses, effect sizes are reported and interpreted based on Cohen ([18]).

---

1 The hit rate refers to how often a person has correctly judged a bag as NOT OK in proportion to all bags containing threat items.

2 The false alarm rate refers to how often a person has wrongly judged a bag as NOT OK in proportion to all bags containing no threat item.
III. RESULTS AND DISCUSSION

The main aim of this study was to investigate how leaving laptops inside passenger bags affects detection performance of x-ray screening officers, when state-of-the-art technology providing single view images is used. A test was created in two different versions. Two experimental groups with screening officers of an international European airport were formed. Each group conducted one test version (A or B), in order to investigate the effect of the two different packing conditions. If detection performance scores would be equally high for both groups (no significant difference), one could conclude that screening bags and laptops separately does actually not deliver an advantage for screening officers. Additionally to detection performance, reaction times were measured and compared.

A. Comparison of Detection Performance for Both Packing Conditions

Figure 4 displays the overall detection performance of both experimental groups (with means and standard deviations\(^3\) of detection performance), indicating that the screeners’ performance was much better when both items were screened separately. A highly significant and large difference in detection performance was found between groups A and B \((t(25.10) = 7.44, p < .001, \eta^2 = 2.35)\). Similar differences between both groups could be observed when taking a closer look at each threat category separately (Figure 5). The results of a repeated measures ANOVA with the within-participant factors threat category (guns, IEDs, knives, others) and the between-participant factor group (A, B) revealed large significant main effects for threat category, \(F(1.88, 71.26) = 69.03, p < .001, \eta^2 = .645\), and group, \(F(1, 38) = 58.92, p < .001, \eta^2 = .608\). The interaction between threat category and group also showed a large significant effect, \(F(1.88, 38) = 13.69, p < .001, \eta^2 = .265\).

As Figure 5 indicates, group A achieved higher detection performance scores for each threat category. The categories IEDs and others were most difficult to detect, in particular when the laptops were left inside the passenger bags. Largest differences between both groups were found for the category IEDs, smallest for guns (see Table 1).

These results already indicate that, in general, it is much easier to identify threat objects in laptops and passenger bags when both items are screened separately.

\(\text{Fig. 4: Mean detection performance scores (A') and standard deviations for both groups (A, B).}\)

\(\text{Fig. 5: Mean detection performance scores (A') and standard deviations for both groups (A, B) and each threat category (guns, IEDs, knives, others).}\)

\(\text{TABLE I}\)

<table>
<thead>
<tr>
<th>Threat Category</th>
<th>t</th>
<th>df</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guns</td>
<td>3.69</td>
<td>24.17</td>
<td>&lt;.01</td>
<td>1.167</td>
</tr>
<tr>
<td>IEDs</td>
<td>9.35</td>
<td>30.60</td>
<td>&lt;.001</td>
<td>2.974</td>
</tr>
<tr>
<td>Knives</td>
<td>5.61</td>
<td>25.83</td>
<td>&lt;.001</td>
<td>1.773</td>
</tr>
<tr>
<td>Others</td>
<td>6.28</td>
<td>23.00</td>
<td>&lt;.001</td>
<td>1.988</td>
</tr>
</tbody>
</table>

As described earlier, except for guns, half of the threat items were placed inside the laptops and half of them were placed inside the bags. Due to their size, all guns were only placed inside the bags and not inside laptops. Figure 6 shows how detection performance differs for each group and threat category with regard to the placement of the threat item. Separate pairwise \(t\)-tests were conducted for each threat category (except guns), to compare the differences in detection performance with regard to the placement of the threats for both groups (see Table 2). While for both groups IEDs were detected worse when these were placed inside the laptops, for the threat categories knives and others this effect was either reversed or not so evident (see Figure 6). A possible explanation for this would be that knives and other threat items often only fitted into the laptops in canonical views and thus were more exposed and easier to recognize. When comparing detection performance scores between both groups, again, for each threat category and placement condition better scores were achieved by group A. The largest difference between both groups could be found for IEDs, when these were built into the...
laptops. While detection performance was still quite high in group A, the scores achieved in group B were much lower. Thus, these analyses show that for both placement conditions detection performance is strongly impaired when laptops are not removed from bags. This applies specifically for IEDs, when these are built into laptops.

### TABLE II
RESULTS OF THE T TESTS COMPARING DETECTION PERFORMANCE WITH REGARD TO THE PLACEMENT OF A THREAT ITEM (IN BAG VS. IN LAPTOP) FOR EACH THREAT CATEGORY AND BOTH GROUPS (A, B).

<table>
<thead>
<tr>
<th>Threat Item</th>
<th>A</th>
<th>B</th>
<th>t(19)</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEDs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3.297</td>
<td>&lt;.01</td>
<td>.737</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>5.980</td>
<td>&lt;.001</td>
<td>1.337</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-12.678</td>
<td>&lt;.001</td>
<td>-2.835</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>-7.755</td>
<td>&lt;.001</td>
<td>-1.734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-14.236</td>
<td>&lt;.001</td>
<td>-3.183</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>-3.312</td>
<td>.758</td>
<td>0.070</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. Comparison of Reaction Times for Both Packing Conditions

In addition to detection performance, reaction time was measured. The reaction time refers to the amount of time it took for a screening officer to come to a decision and press a button (OK or NOT OK), after the image was fully displayed on the screen. Figure 7 shows the average reaction times needed to judge an image broken up by threat category and both participant groups (A and B). As can be seen, for all threat categories, more time was needed to analyze the images in group B, where the laptops were not removed from the bags.

A repeated-measures ANOVA with the within-participants factor threat category (guns, IEDs, knives, others) and the between-participants factor group (A, B,) revealed large significant main effects for the factors threat category, $F(1.8, 69.53) = 54.70, p < .001, \eta^2 = .590$, and group, $F(1, 38) = 14.38, p < .01, \eta^2 = .275$. The interaction between both factors showed a medium effect, $F(1.83, 69.53) = 3.49, p < .05, \eta^2 = .084$. Fastest reaction times were revealed for the category guns while longest reaction times were clearly found for the category IEDs.

In order to see which group actually took more time to complete the test all reaction times for every security screener in each group were summed and averaged across screeners. Figure 8 displays these results. As described at the beginning, for group A, where laptops and bags were displayed separately, the test contained 288 images, whereas in group B 192 images were displayed. Even though fewer images were displayed, reaction times were still significantly longer in group B.

---

4 Due to outliers or asymmetrical distributions, standard deviations from the mean reaction times can be negative.
viewed in group B compared to group A, slightly more time was needed altogether to inspect these test images (see Figure 8). It should be noted, however, that this effect was not significant, t(38) = -.29, p = .773, d = 0.09. Thus, although leaving laptops in passenger bags results in fewer x-ray images that have to be visually inspected, the x-ray screening process altogether is not necessarily speeded up. As the contents of bags with laptops become more difficult to visually search for threat items, more time is needed for such bags.

Yet, results must be handled with care. Longer reaction times might also have occurred as the screeners were not familiar with the packing condition of leaving laptops inside the baggage.

IV. SUMMARY AND CONCLUSIONS

X-ray screening of passenger bags is a challenging task for screening officers, which becomes especially difficult when objects such as large electronic devices (laptops) are contained in the baggage. Due to their compact construction, electronic devices are dense and difficult to penetrate with x-ray beams. Thus, objects behind or in front of larger electronic devices can become difficult to recognize. As most machines currently deployed at airports only provide views of one or two viewpoints, objects such as laptops could intentionally be used to conceal certain parts of luggage or hide threat items (e.g., IEDs). For this reason, most international and national regulations specify that portable computers and other large electronic devices shall be removed from passenger bags and screened separately at security checkpoints (e.g. [9]), causing lower passenger throughput or sometimes even annoyance among passengers.

The aim of this study was to investigate to what extent detection performance of screening officers is impaired when laptops are left inside passenger bags for security screening and when state-of-the-art technology providing single-view images is applied. In order to evaluate this effect, an image interpretation test was created in two different versions. Two experimental groups with screening officers of an international European airport were formed, each group conducting one test version. Detection performance scores and reaction times of both groups were measured and compared.

Results clearly showed that detection performance of the screening officers was much higher when the laptops were removed from the passenger bags and both items were x-ray screened and displayed separately. This was consistent throughout all analyses conducted. More detailed analyses focusing on the different threat categories showed that the categories IEDs and others were most difficult to detect. This could be observed in particular for group B, where laptops were left inside the passenger bags. Largest differences between both groups could be observed for the category IEDs. Further analysis considering the placement of the threat items (in bag vs. in laptop) showed that when IEDs are built into laptops and these are not removed from the bag, detection becomes quite challenging.

Additionally to detection performance, reaction time (how long it took a screener to come to a decision) was measured. Results showed that for all threat categories more time was needed in group B, where laptops were left inside the bags, to analyze the images. Longest reaction times were needed for the categories IEDs, especially in group B. Moreover, statistical analysis revealed that even though more images were viewed in test group A (through separate screening of bags and laptops), there was no significant difference in the total time needed to inspect all x-ray images.

All in all, this study clearly demonstrates that leaving laptops inside passenger bags can substantially impair detection performance of screening officers when using standard single view x-ray systems. The results imply that when no automatic threat detection is applied and only single view images are available, the detection of threat items is more reliable when bags and laptops are screened separately. Thus, the outcomes of this study underline the appropriateness and importance of current regulations specifying that portable computers have to be removed from passenger bags for x-ray screening. Further research is needed to investigate whether this regulation is still appropriate when advanced cabin baggage screening systems, using multi-view x-ray or 3D computer tomography technology, with reliable automated threat detection are available.

V. ACKNOWLEDGMENT

We are thankful to aviation security experts from the German Federal Police Technology Center for providing their valuable expertise and support for the creation and recording of x-ray images.

VI. REFERENCES


VII. VITA

Marcia Mendes works as a research scientist at the University of Applied Sciences and Arts Northwestern Switzerland (FHNW) and the Center for Adaptive Security Research and Applications (www.casra.ch) in Zurich. She is a member of the security research team led by Prof. Dr. Adrian Schwaninger and involved in several national and international projects in aviation security. Marcia finished her studies in psychology (work and organizational psychology, health psychology, clinical psychology and public health) at the University of Bremen (Germany) in 2009. Currently, she is writing her doctoral thesis in the field of aviation security human factors.

Prof. Dr. Adrian Schwaninger lectures at the University of Zurich since 1999 and at the University of Applied Sciences Northwestern Switzerland since 2008. He is the chairman of the Center for Adaptive Security Research and Applications (www.casra.ch) in Zurich and the head of the Institute Humans in complex Systems (MikS) at the School of Applied Psychology, University of Applied Sciences and Arts Northwestern Switzerland (www.fhnw.ch/miks). His areas of expertise are aviation security, human factors, scientifically based software development, applied cognitive psychology, and human-machine interaction. Adrian is a member of the ECAC Training Task Force, the ECAC Technical Task Force, and the ICAO Working Group on Training. Adrian is recognized as a leading authority on aviation security. He has more than 80 publications and more than 150 invited presentations. In 1999 he has received the Young Researcher Award in Psychology. In 2003 he has received the ASI International Award of Excellence in Aviation Security: Enhancement of Human Factors.

Nadine Strebel is a master student in Applied Psychology working as a research assistant at the University of Applied Sciences and Arts Northwestern Switzerland (FHNW) and the Center for Adaptive Security Research and Applications (www.casra.ch) in Zurich. She is involved in several national and international projects in aviation security.

Dr. Stefan Michel is currently working as a scientific researcher and project manager at the University of Applied Sciences and Arts Northwestern Switzerland (FHNW) and the Center for Adaptive Security Research and Applications (www.casra.ch) in Zurich. He manages different national and international projects in aviation security providing clients with cost effective methods to achieve security performance improvements and operational effectiveness. Stefan has a strong background in applied psychology with experience in aviation security enhancing the interaction between people, processes, and technology. Previously, he worked at the University of Zurich (Switzerland) and the Max Planck Institute for Biological Cybernetics in Tübingen (Germany) as a research scientist and received the doctoral degree in 2008.