A First Exploratory Study on the Relevance of Everyday Object Knowledge and Training for Increasing Efficiency in Airport Security X-ray Screening

CONFERENCE PAPER · SEPTEMBER 2015

READS
10

5 AUTHORS, INCLUDING:

Nicole Hättenschwiler
University of Applied Sciences and Arts Nor...
5 PUBLICATIONS 1 CITATION
SEE PROFILE

Stefan Michel
University of Applied Sciences and Arts Nor...
26 PUBLICATIONS 124 CITATIONS
SEE PROFILE

Adrian Schwaninger
University of Applied Sciences and Arts Nor...
128 PUBLICATIONS 1,285 CITATIONS
SEE PROFILE

Available from: Nicole Hättenschwiler
Retrieved on: 23 September 2015
A First Exploratory Study on the Relevance of Everyday Object Knowledge and Training for Increasing Efficiency in Airport Security X-ray Screening

Nicole Hättenschwiler, Stefan Michel, Milena Kuhn, Sandrina Ritzmann and Adrian Schwaninger

School of Applied Psychology
University of Applied Sciences and Arts Northwestern Switzerland (FHNW)
Olten, Switzerland
and
Center for Adaptive Security Research and Applications (CASRA)
Zurich, Switzerland

Abstract — Secure air transportation is vital for economy and society and it relies heavily on airport security screening. Passenger bags and other belongings are screened using X-ray machines to ensure that they do not contain prohibited items. Human operators (X-ray screeners) visually inspect X-ray images to decide whether they are harmless or whether they might contain a prohibited item and therefore require secondary search (typically using manual search and/or explosive trace detection technology). Several previous studies have shown that learning which items are prohibited and what they look like in X-ray images is important to achieve good detection performance. As passenger bags contain a large variety of harmless everyday objects, it could be assumed that knowing what such objects look like in X-ray images could help X-ray screeners to work more efficiently by reducing false alarms (i.e. sending a passenger bag to secondary search even though it does not contain a prohibited item). In the first experiment, the relationship between knowledge of harmless everyday objects and false alarm rate was investigated with 15 certified X-ray screeners of one large European airport. Statistical analyses revealed a good knowledge of harmless everyday objects on average with some variation between X-ray screeners and a negative correlation with false alarm rate. In the second experiment, the effectiveness of an e-learning course for acquiring knowledge of everyday objects in X-ray images was evaluated. Thirty novices conducted a test-retest experiment where half of the participants conducted an e-learning course about harmless everyday objects in X-ray images between the two tests. The results revealed that e-learning can be an effective and efficient method for increasing the knowledge of everyday objects in X-ray images. Based on the results of both studies, the relevance to learn everyday objects as part of initial and recurrent training of X-ray screeners is discussed.

Keywords — aviation security; X-ray screening; detection performance; everyday object recognition; everyday object training

I. INTRODUCTION

Reviews of attacks against civil aviation since 11th September 2001 underline the importance of aviation security measures (e.g. [1]). At airport security checkpoints, passengers and their belongings are screened to ensure that prohibited items (guns, knives, improvised explosive devices (IED) and other threat items) are not carried. State-of-the-art X-ray screening equipment offers good image quality with high resolution, automated detection of explosives, several image enhancement functions and other features [2]; [3]; [4]. The main task of an X-ray screener is to visually inspect X-ray images of passenger bags and to decide whether a bag is harmless or whether it might contain a prohibited item and therefore needs further inspection by secondary search (typically by using manual search and/or explosive trace detection technology). As pointed out by [5]; [6]; [7] prohibited items are difficult to recognize without training because 1) objects often look very different in X-ray images than in reality, 2) certain prohibited items are not known from everyday experience (e.g. IEDs), 3) some prohibited items look similar to harmless objects (e.g. a switchblade knife can resemble a pen), and 4) when objects are depicted from unusual viewpoints, they become difficult to recognize. During initial classroom, computer-based and on the job training, X-ray screeners learn how to interpret X-ray images in order to recognize everyday objects and prohibited items. In the last decade, several studies have shown that computer-based training (CBT1) is important, efficient and effective to achieve good detection performance in X-ray image interpretation ([8]; [9]; [10]; [11]; [12]; [13]; [14]; [15]). While these studies have provided converging on the importance to learn which items are prohibited and what they

1 In the literature, the term CBT is used in different ways. The definition used here refers to all forms of self-paced distance training and learning activities using computers [16] and therefore includes e-learning courses.
look like in X-ray images, the role of everyday object knowledge has not yet been addressed specifically. This is an interesting topic especially from an operational point of view. In particular, one could assume that the knowledge on what everyday objects look like in X-ray images could result in fewer cases where an everyday object is confused with a prohibited item (e.g. pen can resemble a switchblade knife). This would result in fewer false alarms, i.e. wrongly judging a bag to contain a prohibited item. False alarms have to be resolved by secondary search which typically involves manual search and/or alarm resolution using explosive trace detection technology [17]. Due to the additional time needed for secondary search, high false alarm rates can have a strong negative impact on throughput [17] and they could also result in lower passenger satisfaction [18]. Therefore, it is worth investigating the role of everyday object knowledge and training because it could be very relevant for more efficient X-ray screening.

In this first exploratory study, two experiments were conducted. For both experiments, a new test was created to measure how well novices and X-ray screeners can categorize and name everyday objects in X-ray images. In Experiment 1 bivariate and partial correlation analyses were performed to examine whether there is a statistically significant and meaningful relationship between everyday object knowledge and false alarm rate in a simulated X-ray baggage screening task. In Experiment 2 it was investigated whether e-learning can be used to learn effectively and efficiently everyday object knowledge which could be relevant for more efficient X-ray screening.

II. EXPERIMENT 1

The main aim of Experiment 1 was to investigate the relationship between everyday object knowledge and false alarm rates. To this end, X-ray screeners were tested using an everyday object categorization and naming test and a simulated X-ray baggage screening task. Two X-ray screeners with experience in test development and teaching X-ray image interpretation assisted for creating the stimuli. They did not participate as subjects in the experiments and are referred to as X-ray screening experts from now on.

III. METHOD AND PROCEDURE

Fifteen X-ray screeners (7 female), with a mean age of 39.8 years (SD = 10.13) and a mean work experience of 6.03 years (SD = 4.44) in cabin baggage screening at a large international European airport were tested. The X-ray screeners first conducted an X-ray object categorization and naming test (X-Ray OCNT) and four weeks later a simulated X-ray baggage screening task (XBST).

A. X-Ray Object Categorization and Naming Test (X-Ray OCNT)

For the purpose of this study, 32 X-ray images of passenger bags were selected from a pool of 2800 images by the X-ray screening experts to be representative.

B. Simulated X-ray baggage screening task (XBST)

160 color X-ray images of passenger bags were selected by the two X-ray screening experts to be representative. In half of the X-ray images, one prohibited item was added using the validated X-ray image blending software mentioned above. Four categories of prohibited items were used (guns, knives, IEDs and other prohibited items). For each category there were ten exemplars. Each exemplar was displayed once in easy view (as defined by the two X-ray screening experts and the authors) and difficult view (rotated around the horizontal or vertical axis by 85 deg.). For each category, half of the prohibited items were part of the CBT system used at this airport. Since it cannot be assumed that terrorists would use a prohibited item that is contained in the CBT, half of the prohibited items were newly recorded and visual comparison was used to make sure that they are different from the prohibited items contained in the CBT. The X-ray images were displayed on the screen without time limit and X-ray screeners had to decide whether it contained a prohibited item or not by pressing a key. There was no feedback on the correctness of responses and the participants took about 30 minutes to complete the test.

IV. RESULTS AND DISCUSSION

On average, 96.19% of the objects were correctly categorized (SD = 1.67%) in the X-ray OCNT. Harmless
everyday objects were correctly named on average in 82.07% of the cases ($SD = 11.19\%$). In the XBST, when X-ray screeners wrongly judged a bag to contain a prohibited item, this counted as a false alarm. False alarm rate was calculated by number of false alarms divided by number of X-ray images of bags not containing a prohibited item. On average, screeners had a mean false alarm rate of $M = .037$ ($SD = .031$). As can be seen in Fig. 2 there is a negative linear relationship between % correctly named harmless everyday objects and false alarm rate. Even though a relatively small number of X-ray screeners participated in Experiment 1 ($n=15$), a one-tailed Pearson correlation was significant, $r(13) = -.471$, $p=.038$, showing an effect size which is modest according to [19].

Two observations are worth mentioning when looking at Fig. 2. First, there was substantial variation between X-ray screeners regarding % correctly named harmless everyday objects ranging from 54% to 96%. Two X-ray screeners had quite low values of 54%. Since correlations are very sensitive to outliers, it was examined whether the values of these two screeners on % correct naming are below the often used criterion of three standard deviations ($11.19\% \times 3 = 33.57\%$) from the mean (82.07%). Since this was not the case, the found correlation can be regarded as valid subject to verification with a larger sample size (for a review and discussion on different methods for outlier analysis, see [20]). Second, although all X-ray screeners had false alarm rates measured with the XBST that were below 11%, there was still enough variation between screeners so that a correlation with % correct naming could be revealed.

A key concept of signal detection theory [21] is that variation in false alarm rate can also result from differences in response bias, i.e. a tendency to report more or less often that a signal (in this case a prohibited item) is present. Response bias depends on a variety of factors including individual ones, stimuli used, test design and other situational factors ([21]; for a more recent detailed overview and discussion on detection theories, see [22]). An often used measure of response bias is the criterion which can be calculated as follows:

$$c = -0.5[z(H) + z(FA)]$$

(1)

$H$ refers to hit rate which in the XBST corresponds to number of correct decisions for X-ray images containing a prohibited item divided by number of X-ray images of passenger bags containing a prohibited item.

FA refers to false alarm rate, which is calculated as defined above.

$z$ refers to the $z$ transformation which is used to convert hit and false-alarm rates to $z$ scores (i.e. standard deviation units).

Fig. 3 shows the relationship between response bias (criterion) and false alarm rate. The high negative bivariate Pearson correlation, $r(13) = -.870$, $p<.001$, with a high effect size according to [19] is consistent with signal detection theory according to which response bias strongly influences the false alarm rate.

Although it seems not plausible to assume that X-ray screeners with good knowledge of harmless everyday objects would have systematically different response biases, this possibility should be ruled out by showing that the correlation between % correctly named harmless everyday objects in the X-ray OCNT and the false alarm rate from the XBST is not due to the background variable response bias. Therefore, a partial correlation (Pearson) between % correctly named harmless everyday objects in the X-ray OCNT and the false alarm rate from the XBST controlling for response bias (criterion $c$) was performed. The result $r(13) = -.492$, $p=.037$ (one-tailed) shows that variation in response bias cannot explain the correlation between everyday object knowledge and false alarm rate. In contrary, when controlling for response bias, the correlation between everyday object knowledge and false alarm rate becomes even slightly higher.

---

2 The results for naming prohibited items in the X-ray OCNT are not reported because this information could be regarded as security sensitive and because these results are not relevant for testing the relationship between knowledge of harmless everyday objects and false alarm rates from the XBST.

3 Due to recording failure of the software a few trial responses were missing. More specifically, for six of the 15 X-ray screeners, one response was missing. This resulted in a total of 0.25% missing responses which for the calculation of the false alarm rate and the criterion resulted in negligible deviations compared to when all responses would have been recorded.

4 Since the hypothesis stated in the introduction would predict that with increasing everyday object knowledge, false alarm rates would decrease, one-tailed significance levels are reported.
These results are highly consistent with the hypothesis stated in the introduction, namely that everyday object knowledge helps reducing false alarm rates in X-ray screening tasks which is important for increasing efficiency.

It should be noted that causation cannot be inferred from correlational analyses even if partial correlation is used. However, the results of Experiment 1 are at least encouraging with regard to further studies in which the role of everyday object knowledge could be examined using an experimental design in which the effect of systematic variation of everyday object knowledge on X-ray screening performance is investigated, ideally including a control group.

V. EXPERIMENT 2

While several studies exist on the detection of prohibited items by X-ray screeners before and after several months of CBT ([8]; [10]; [11]; [12]; [13]; [14]; [15]; [23]), no study has been conducted yet to investigate how well novices can recognize harmless everyday objects in X-ray images and whether this can be trained efficiently and effectively. To address this topic, a test-training-test design was used with an experimental and a control group.

VI. METHOD AND PROCEDURE

The experimental design is illustrated in Fig.4. Thirty novices (15 female) with a mean age of 31.1 years (SD = 11.87) and no work experience in X-ray screening were first tested on their visual abilities using the X-Ray Object Recognition Test (X-Ray ORT; [24]) as described below. Afterwards, from these thirty novices two equal groups (an experimental and a control group) were built based on the results from the X-Ray ORT so that the mean and standard deviation of $d'$ (a measure of sensitivity, Green & Swets, 1966) did not differ significantly between both groups, $t(28) = 1.176$, $p = .25$. Afterwards, all participants conducted an adapted version of the X-Ray OCNT (X-Ray ONT) used in Experiment 1 (prohibited items were excluded) to assess their knowledge of harmless everyday objects. Afterwards, the experimental group conducted an e-learning course on everyday objects, which is described below. The control group did not receive any training. Both groups were tested again one week later using the X-Ray ONT.

A. X-Ray Objects Recognition Test (ORT)

The X-Ray ORT is an X-ray image interpretation test which was developed to measure the ability to cope with image-based factors in X-ray image interpretation (i.e. effects of viewpoint, superposition by other objects, and bag complexity). It consists of 256 grey-scale X-ray images of passenger bags. Half of them contain either a gun or a knife; the other 128 X-ray images are harmless bags. Each bag is displayed for 4 seconds on the screen and participants have to decide for each image whether the bag is OK (i.e. there is no prohibited item in the bag) or whether it is NOT OK (i.e. the bag contains a gun or a knife) by clicking on a button on the screen. The X-Ray ORT takes about 30 minutes to complete. Information on test construction, its reliability and validity measures can be found in [24]; [25].

B. X-Ray Introduction of Everyday Objects (XRI)

X-Ray Introduction (XRI) is an e-learning course for beginners teaching the look of harmless everyday objects in cabin baggage and air cargo, which was provided by CASRA. The XRI is recommended for people who have not worked with X-ray images before. In a short introduction, X-ray technology and the meaning of colors in X-ray images are explained. In the following nine sessions, 90 harmless everyday objects are introduced in three steps. First, harmless everyday objects are shown as X-ray images together with a photograph. As objects sometimes look very different in X-ray images compared to reality, this is an important step to learn what everyday objects look like in X-ray images. Second, X-ray images of a passenger bags containing these everyday objects are displayed one after the other, giving the trainees the opportunity to view the introduced objects within a realistic context. Then, trainees have to find each learned everyday object in X-ray images of passenger bags by clicking on the object indicated in the instruction. Participants can also monitor their learning progress in form of a self-evaluation exercise.

The XRI was conducted without interruption and took about 2.5h to complete.
C. Everyday Objects Naming Test (X-Ray ONT)

The X-Ray ONT from Experiment 1 was used with two adaptations: 1) the red frames which marked the prohibited items were deleted which resulted in 19 X-ray images containing three harmless everyday objects, nine X-ray images containing two harmless objects and four X-ray images containing one harmless object. 2) Objects had only to be named and not categorized\(^5\). Different X-ray images were used for the X-Ray ONT and the XRI. The participants completed the test in about 45 - 60 minutes.

VII. RESULTS AND DISCUSSION EXPERIMENT 2

Fig.6 shows the results of the X-Ray ONT for the experimental and the control group. When conducting the X-ray ONT the first time, the groups did not differ significantly, \(t(28) = -0.377, p = .709\). The control group achieved 52.72% correct naming (SD = 12.48), and the experimental group 54.55% (SD = 14.11). This result indicates that novices could recognize about half of the everyday objects in X-ray images used in the X-Ray ONT before taking any training. While this result is encouraging, it also means that many everyday objects cannot be recognized in X-ray images without training. When conducting the X-Ray ONT the second time, the control group achieved 58.28% correct naming (SD = 10.82), whereas the experimental group reached 66.05% (SD = 10.29). Bonferroni-corrected post-hoc analyses revealed a significant difference between the X-ray ONT performance before and after training for the experimental group \(t(28) = 2.549, p = .008\), but not for the control group \(t(28) = 1.305, p = .101\). To examine main effects and interaction, a mixed 2x2 univariate ANOVA was conducted with the dependent variable % correct naming (X-Ray ONT), the within-subjects factor time (before and after taking the XRI) and the between-subjects factor groups (experimental vs. control group). There was no main effect of group, \(F(1, 28) = 1.26, p = .272\), a main effect of time, \(F(1, 28) = 77.89, p < .001\), and a significant interaction between time and group, \(F(1, 28) = 9.42, p = .005\).

These results can be summarized as follows: Novices could recognize about half of the everyday objects in the X-Ray ONT without training. Using a short e-learning module (XRI) during 2.5 hours did already increase % correct naming in the X-ray ONT by about 12%. This result is very encouraging regarding the question whether knowledge of harmless everyday objects in X-ray images can be trained effectively and efficiently.

VIII. SUMMARY, CONCLUSIONS AND LIMITATIONS

To get first insights into the relevance of knowledge of everyday objects in X-ray security screening, two experiments were conducted. Firstly, to explore the relationship between the knowledge about harmless everyday objects and the false alarm rate measured in a simulated X-ray baggage screening task, and secondly, to investigate the effectiveness of an e-learning course on harmless everyday objects in X-ray images.

From an operational perspective, a low false alarm rate is desirable to guarantee the efficiency of the security screening process. We assumed a negative relationship between the percentage of correct answers for the naming of harmless everyday objects in an object categorization and naming task and the false alarm rate in a simulated X-ray baggage screening task. This hypothesis could be confirmed while controlling for variation in response bias using partial correlation. As mentioned earlier, it should be noted that causation cannot be inferred from correlational analyses even if partial correlation is used. Moreover, Experiment 1 was conducted with a rather small sample size (n=15). Further studies with a bigger sample size using an experimental design with systematic variation of everyday object knowledge and a control group would lead to stronger causal conclusions. Also, the performance of X-ray screeners with CBT on prohibited item detection who either do or do not receive additional everyday object training could be compared.

Experiment 2 showed promising results on e-learning as an efficient and effective tool for building knowledge of harmless everyday objects in X-ray images. This is certainly important for initial training so that X-ray screeners can work efficiently as early as possible. Although on average, certified X-ray screeners reached high values in everyday object naming in Experiment 1, there was still substantial inter-individual variation. Therefore, it could makes sense to run additional studies in which cost benefit analyses are conducted to see whether everyday object recognition training should be implemented not only for initial but also for recurrent training.

To conclude, this report shed light on the role of knowledge of harmless everyday objects for X-ray screening. Such knowledge, in combination with sound visual knowledge of prohibited items, could have positive effects on the effectiveness and the efficiency of airport security screening.

REFERENCES


