Detecting Bombs in X-Ray Images of Hold Baggage: 2D Versus 3D Imaging

Nicole Hättenschwiler, **Marcia Mendes**, and **Adrian Schwaninger**, University of Applied Sciences and Arts Northwestern Switzerland, Olten

Objective: This study compared the visual inspection performance of airport security officers (screeners) when screening hold baggage with state-of-the-art 3D versus older 2D imaging.

Background: 3D imaging based on computer tomography features better automated detection of explosives and higher baggage throughput than older 2D X-ray imaging technology. Nonetheless, some countries and airports hesitate to implement 3D systems due to their lower image quality and the concern that screeners will need extensive and specific training before they can be allowed to work with 3D imaging.

Method: Screeners working with 2D imaging (2D screeners) and screeners working with 3D imaging (3D screeners) conducted a simulated hold baggage screening task with both types of imaging. Differences in image quality of the imaging systems were assessed with the standard procedure for 2D imaging.

Results: Despite lower image quality, screeners' detection performance with 3D imaging was similar to that with 2D imaging. 3D screeners revealed higher detection performance with both types of imaging than 2D screeners.

Conclusion: Features of 3D imaging systems (3D image rotation and slicing) seem to compensate for lower image quality. Visual inspection competency acquired with one type of imaging seems to transfer to visual inspection with the other type of imaging.

Application: Replacing older 2D with newer 3D imaging systems can be recommended. 2D screeners do not need extensive and specific training to achieve comparable detection performance with 3D imaging. Current image quality standards for 2D imaging need revision before they can be applied to 3D imaging.

Keywords: human–automation interaction, visual search, graphical user interfaces (GUI), experience, transfer of training

---

Address correspondence to Nicole Hättenschwiler, School of Applied Psychology, Institute Humans in Complex Systems, University of Applied Sciences and Arts Northwestern Switzerland, Riggenbachstrasse 16, CH-4600 Olten, Switzerland; e-mail: nicole.haettenschwiler@fhnw.ch.

HUMAN FACTORS
Vol. 61, No. 2, March 2019, pp. 305–321
DOI: 10.1177/0018720818799215

On December 21, 1988, Pan Am Flight 103 exploded over Lockerbie, Scotland, due to a bomb in a passenger bag transported in the hold of the aircraft (Strantz, 1990). Since then, many terrorist attacks have targeted airplanes (Baum, 2016; Singh & Singh, 2003). The most recent involving a bomb in hold baggage occurred on October 31, 2015, when Metrojet Flight 9268 was blown up during flight killing all 224 passengers (Baum, 2016). In response to such bomb threats, explosive detection systems (EDS) based on 2D imaging for hold baggage screening (HBS) were developed and introduced about 15 years ago (Caygill, Davis, & Higson, 2012; Harding, 2004; Singh & Singh, 2003). Such EDS-HBS assist airport security officers (screeners) who visually inspect X-ray images of passenger bags before they are loaded into the hold of an aircraft (Wells & Bradley, 2012). Newer 3D imaging technology uses computer tomography (CT). Technically, this has better automated explosive detection, higher baggage throughput, and 3D-rotatable images. Nonetheless, it also has lower image resolution and, therefore, poorer image quality than older 2D imaging technology (Flitton, Breckon, & Megherbi, 2013; Mouton & Breckon, 2015; Oftring, 2015; Wells & Bradley, 2012).

Human-machine system performance depends on technology and human factors. For instance, if lower image quality with 3D imaging would make it harder for screeners to decide whether a bag contains an improvised explosive device (IED), then 3D screening could in fact be inferior to 2D screening despite having better automated explosive detection. On the other hand, and this is a very important point to consider, if screeners would achieve at least similar detection performance with 3D imaging compared with 2D imaging, then the human-machine system as a whole would perform better with 3D imaging because this technology has better automated explosive detection and higher baggage throughput. Investigating this issue is of major practical relevance: Although some countries introduced 3D imaging several years ago, other countries do not accept such technology due to their lower image quality compared with older 2D imaging EDS-HBS technology, even though 3D imaging has better automated explosive detection capability and higher baggage throughput (Flitton et al., 2013; Oftring, 2015). Moreover, there is a current debate on the international regulatory level regarding whether screeners working with 2D imaging need extensive and specific training before they can be allowed to work with 3D imaging technology. Our study addressed both issues by testing 2D and 3D screeners with 2D and 3D imaging in a simulated hold baggage screening task with the following research questions: (a) Can screeners achieve at least similar detection performance using 3D imaging compared with 2D imaging despite lower image resolution? (b) Does visual inspection competency acquired with one type of imaging transfer to the other type of imaging? These research questions are also interesting from a theoretical perspective—in particular, with regard to human-machine interaction, visual information processing and transfer of learning. Before discussing the relevant literature, it is important to clarify important terms and processes regarding the airport security screening of cabin and hold baggage.

Passengers store their carry-on bags in the cabin of airplanes. Because such cabin baggage can be accessed during flight, guns, knives, IEDs, and other items that could pose a threat (e.g., electric shock devices) are prohibited (Hancock & Hart, 2002; Harris, 2002; Schwaninger, 2005). As required by law (e.g., European Commission, 2015), screeners visually inspect every piece of cabin baggage at airport security checkpoints using X-ray machines. Larger baggage, in contrast, is stored in the hold of an aircraft and processed differently (Shanks & Bradley, 2004). Passengers have to register such hold baggage at check-in stations before going through airport security checkpoints. Hold baggage is then processed by a baggage handling system containing X-ray machines that have EDS-HBS (Level 1 of hold baggage screening) that highlights areas on the X-ray image that might contain explosive with colored rectangles (2D imaging systems) or by coloring the suspect area (3D imaging systems; see Figure 1 for illustrations). Whereas there are multiple target types (guns, knives, IEDs, explosives, other prohibited items) in cabin baggage screening, this is not the case in hold baggage screening. Because passengers cannot access items stored in the hold of an aircraft, guns or knives do not pose a threat, and hold baggage screening targets only fully functioning IEDs (Bretz, 2002). Only X-ray images of hold baggage on which an EDS-HBS has raised an alarm are sent to remote screening locations for on-screen alarm resolution by screeners (Level 2 of hold baggage screening). They visually inspect the X-ray images and decide whether the bag is harmless or contains a fully functioning IED with the following components: a triggering device, a power source, an explosive, and a detonator that need to be connected to each other by, for example, wires (Turner, 1994; Wells & Bradley, 2012). If screeners decide that an X-ray image is suspicious, more time-consuming investigations follow including rescreening with other X-ray technology, trace detection, explosive detection dogs, passenger reconciliation, and the opening of bags (Shanks & Bradley, 2004; Singh & Singh, 2003).

Since the terrorist attacks on September 11, 2001, there have been many studies on the visual inspection of X-ray images of cabin baggage, which consists of visual search and decision making (Koller, Drury, & Schwaninger, 2009; McCarley, Kramer, & Wickens, 2004; Wales, Anderson, Jones, Schwaninger, & Horne, 2009; Wolfe & Van Wert, 2010). Visual search challenges include low target prevalence, variations in target visibility, and the possible presence of multiple targets (Biggs & Mittroff, 2014; Clark, Cain, Adamo, & Mitroff, 2012; Godwin et al., 2010; Godwin, Menneer, Cave, Thaibsyah, & Donnelly, 2015; Mitroff, Biggs, & Cain, 2015). When it comes to decision making on whether a bag contains a prohibited item, screeners need to know which items are prohibited and what they look like in X-ray images (Schwaninger, 2005). Several studies have shown the importance of computer-based training in helping screeners to

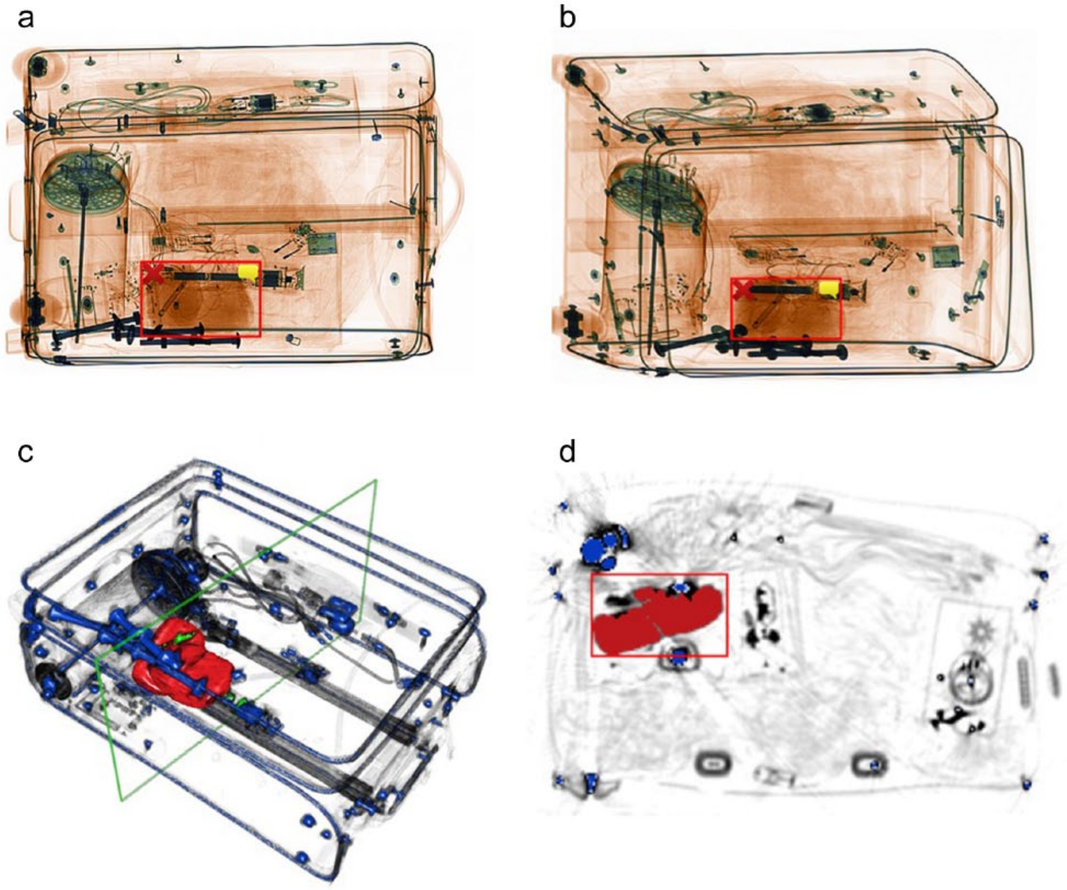

Figure 1. Target-present bag containing an IED recorded with a 2D multiview X-ray and a 3D CT imaging system currently used at airports: (a) 2D default image, (b) second 2D image with 30 degrees difference in perspective, (c) 3D-rotatable image, and (d) 3D-sliceable image. Explosive material is highlighted by the 2D imaging system with red rectangles (Figure 1a and 1b) and by the 3D imaging system with red coloring (Figure 1c and 1d). With 3D imaging, the detonator is visible in green (Figure 1c) and in blue (Figure 1d).

achieve and maintain high visual inspection performance (Fiore, Scielzo, Jentsch, & Howard, 2006; Halbherr, Schwaninger, Budgell, & Wales, 2013; Koller et al., 2009; Koller, Hardmeier, Michel, & Schwaninger, 2008; Schuster, Rivera, Sellers, Fiore, & Jentsch, 2013; Schwaninger & Hofer, 2004; Schwaninger, Hofer, & Wetter, 2007). International regulations take this into account by mandating initial and recurrent training of screeners. For example, European regulations mandate at least 6 hr of image recognition training and testing in every 6-month period for cabin- and hold-baggage screeners (European Commission, 2015).

Target prevalence in real-world baggage screening is about 2% because airports use threat image projection, a technology that projects X-ray images containing targets into the flow of images that are visually inspected by screeners (Hofer & Schwaninger, 2005; Schwaninger, 2006; Schwaninger et al., 2007; Schwaninger, Hardmeier, Riegelnig, & Martin, 2010). The challenge of low target prevalence in visual search refers to the finding that rare targets are frequently missed (Godwin et al., 2010; Wolfe, Brunelli, Rubinstein, & Horowitz, 2013; Wolfe, Horowitz, & Kenner, 2005). This is consistent with signal detection theory (SDT, Green &

Swets, 1966) according to which the probability of signal occurrence (target prevalence) influences the probability of responding that a signal (target) is present. Using the SDT framework, the target prevalence effect can be explained as a shift in response bias (Fleck & Mitroff, 2007; Godwin et al., 2010; Lau & Huang, 2010; Wolfe et al., 2007; Wolfe & van Wert, 2010). SDT provides a measure of detection performance (d′) that is independent of response bias (and therefore also of target prevalence). This has been confirmed for different domains and tasks (Green & Swets, 1966; MacMillan & Creelman, 2005; Swets, 1996) including X-ray image inspection and visual search (Meneer, Donnelly, Godwin, & Cave, 2010; Verghese, 2001; Wolfe & Reynolds, 2008; Wolfe & Van Wert, 2010). Moreover, Schwaninger, Hofer, and Wetter (2007) found very similar detection performance (d′) in screeners when performing a computer-based test with a target prevalence of 50% compared with detection performance (d′) measured on the job using threat image projection data with a target prevalence of 2%.

Regarding target visibility, studies have shown how image-based factors impact on visual inspection performance (e.g., Bolfing, Halbherr, & Schwaninger, 2008; Schwaninger, Hardmeier, & Hofer, 2005; Schwaninger, Michel, & Bolfing, 2005, 2007). For example, objects depicted from unusual viewpoints are more difficult to recognize (effect of viewpoint). Moreover, in X-ray images, objects appear with overlay, and detecting prohibited items depends on how much they are superimposed by other objects (effect of superposition). Finally, prohibited items are more difficult to recognize in complex bags containing many other items and clutter (effect of bag complexity). These challenges can be reduced with 2D imaging that displays a passenger bag as two X-ray images from different perspectives (dual-view imaging). However, previous studies on cabin baggage screening have shown that although dual-view imaging leads to higher detection performance than single-view X-ray imaging, it also increases response time (von Bastian, Schwaninger, & Michel, 2008; Franzel, Schmidt, & Roth, 2012). Similar results have been found for motion imaging in which bags are displayed as an animated sequence of X-ray images depicting a bag from different viewpoints (Mendes, Schwaninger, & Michel, 2013).

Several years ago, advanced CT technology, which has been implemented highly successfully in medical imaging (Barrat, 2000), became available for hold baggage screening (Mouton & Breckon, 2015; Wetter, 2013). Compared with the older 2D imaging technology used in HBS, state-of-the art CT scanners feature better automated explosive detection, slicing, and 3D-rotatable images (Flitton et al., 2013; Mouton & Breckon, 2015; Oftring, 2015; Wells & Bradley, 2012). Slicing refers to the production of cross-sectional images or “slices” of a bag. From a series of image slices, a bag can be reconstructed as a 3D CT volume image and the bag can be displayed as a 3D-rotatable and 3D-sliceable image (Flitton, Breckon, & Megherbi, 2010, 2013). This could result in better detection performance among screeners for two reasons: First, it might be easier to recognize the different components of an IED that, in certain 2D views, would be displayed from a difficult viewpoint and/or superimposed by other items in a complex bag (Bolfing et al. 2008; Schwaninger, Michel, & Bolfing, 2005, 2007). Second, object recognition research has shown that exposure to 3D images results in richer visual object representations (Tarr & Vuong, 2002; Vuong & Tarr, 2004). This could improve screeners’ detection performance not only in 3D but also in 2D images. On the other hand, CT systems have lower image resolution and therefore lower image quality compared with EDS-HBS 2D imaging (Flitton et al., 2010, 2013; Mouton & Breckon, 2015), and this could impair screeners’ detection performance with 3D imaging. Regarding response times (RT), screeners might take more time to visually inspect 3D images because rotating X-ray images and slicing both require additional time.

This study extends previous research on cabin baggage screening by addressing questions of high practical and theoretical relevance for hold baggage screening. We wanted to know (a) whether screeners using 3D imaging can achieve at least similar detection performance to that when using 2D imaging despite lower image

TABLE 1: Description of Screeners Participating in the Study

Participants	*n*	% female	Age	Work experience with 2D imaging (months)	Work experience with 3D imaging (months)
2D Screeners	42	61%	*M* = 44.90 *SD* = 10.36	*M* = 138.31 *SD* = 78.35	
3D Screeners	42	35%	*M* = 36.76 *SD* = 9.22	*M* = 86.02 *SD* = 64.42	*M* = 19.12 *SD* = 5.07

resolution, and (b) whether the visual inspection competency acquired with one type of imaging transfers to the other type of imaging. We addressed these research questions by asking two screener groups that differed in their experience in working with the two imaging technologies to perform a simulated hold baggage screening task with both 2D and 3D imaging. In order to achieve high external validity, we used X-ray images that were recorded with 2D and 3D imaging systems that are currently operational at airports. It is important to note that the reason for comparing 2D and 3D imaging differing in image quality is that the two types of imaging tested in this study are from real-world systems; there is therefore a need to know whether 3D screening results in better human-machine system performance despite lower image quality.

Our main dependent variable was detection performance (d′), which has high external validity for real-world baggage screening because it is independent of target prevalence. Due to the fact that airports use threat image projection with a target prevalence of about 2% (Hofer & Schwaninger, 2005; Schwaninger, Hofer, & Wetter, 2007), target-absent RT were also important, because they account for about 98% of X-ray images in real-world hold baggage screening. Based on our results, we shall discuss whether replacing older 2D with newer 3D imaging technology improves the human-machine system performance in terms of efficiency and effectiveness of the hold baggage screening process as a whole. In addition, our results have important implications in light of current international discussions on whether extensive and specific training should be mandated for 2D screeners before allowing them to work with 3D imaging technology.

METHOD

Participants

Participants were professional hold baggage screeners from two international airports (see Table 1 for details). All screeners had been selected, qualified, trained, and certified according to the standards set by the appropriate national authority (civil aviation administration) in compliance with the relevant EU regulation (European Commission, 2015). Eighty-eight screeners consented to participate in the study (43 2D screeners and 45 3D screeners). Three screeners (one 2D and two 3D) who could not attend the main test due to illness were excluded. One further 3D screener had to be excluded due to a malfunction of the simulator. This left a total of 84 screeners (42 2D screeners [21 tested with 2D imaging and 21 tested with 3D imaging] and 42 3D screeners [23 tested with 2D imaging and 19 tested with 3D imaging]). The current research complied with the American Psychological Association Code of Ethics and was approved by the institutional review board of the University of Applied Sciences and Arts Northwestern Switzerland. Informed consent was obtained from all participants.

Design

All participants attended the airport test facilities twice. First, they completed a pretest to familiarize themselves with the 2D and 3D simulators and the testing procedure. For the main test 2 weeks later, screeners were randomly assigned to be tested with either 2D or 3D imaging. The experiment (main test) used a between-subjects design with X-ray imaging technology (2D vs. 3D imaging) and screener group (3D vs. 2D screeners) as independent variables and visual inspection performance measures as dependent variables (detection performance [d′], target-absent RT, and target-present RT).

Materials

Aviation security experts from a specialized police organization running one of the test centers responsible for airport security equipment testing and certification in Europe created 64 different IEDs (32 for the pretest and 32 for the main test, IEDs were randomly assigned to be used in the pretest or the main test). X-ray images of hold baggage were recorded at this test center by five aviation security experts and the first and second author using 2D multiview X-ray and 3D CT imaging systems that are currently being used at airports (see Figure 1 for examples of images and further information).

Thirty-two different bags were used repeatedly by repacking them to create unique stimuli for the pretest and the main test. All bags were of medium complexity as defined by the aviation security experts. Target-present images contained one IED. Target-absent images contained EDS-HBS false alarms (e.g., cheese, certain liquids, etc.). To ensure that the 3D imaging condition had the same system reliability (e.g., Rice & McCarley, 2011) as the 2D imaging condition, we used EDS-HBS alarms from 3D imaging as a reference when setting red frames manually around the same objects of interest in 2D imaging stimuli.

The pretest consisted of 64 2D X-ray images and 64 3D CT images of different bags. Target prevalence was 50%. Each IED was used twice in different bags: once recorded from a more frontal perspective displaying more surface area, and once from a horizontally or vertically rotated perspective using medium superposition. The main test consisted of 256 bags that were recorded with 2D and 3D imaging. Target prevalence was 50%. Each of the 32 IEDs was used four times in four different bags by varying viewpoint and superposition.

As described in the introduction, 3D imaging systems have lower image quality than 2D imaging systems. To assess such differences, we used the standard test piece (STP) and protocol, which is currently the most widely used international standard for the assessment of image quality of 2D imaging systems (see the Appendix for details).

Procedure

Tests were conducted without giving performance feedback using simulators provided by the manufacturer of the 2D and 3D imaging systems. Six computer workstations with 19” TFT monitors were set up in a normally lit room. Each screener sat approximately 50 cm away from the monitor. The X-ray images covered about two thirds of the screen. Four to six participants performed the test in each session while working individually, quietly, and under supervision. This is a typical scenario in hold baggage screening (Kuhn, 2017). Screeners received instructions before the start of each test informing them about the imaging systems, the number of images, and that the target items were IEDs. To prevent a criterion shift (change of response bias) during the experiment, we informed the screeners beforehand about the target prevalence in the experiment (see also McCarley, 2009; Rich et al., 2008).

Screeners were instructed to visually inspect each X-ray image as if they were working at the airport and to decide as accurately and quickly as possible whether or not the image contained a target by clicking on a target-present or a target-absent button on the simulator interface (a yes–no task in signal detection theory; see MacMillan & Creelman, 2005). After receiving their instructions, all participants started the experiment with 10 practice trials (5 target-absent and 5 target-present images in random order). A time limit of 90 s was set for viewing an X-ray image; afterwards, the image disappeared, but the screeners still had to make a decision.

European regulations mandate that screeners have to take a break of at least 10 min after 20 min of continuous visual inspection of X-ray images (European Commission, 2015). Therefore, tests were divided into four blocks, and screeners were asked to take breaks of 10 to 15 min after completing each block. Block order was counterbalanced across participants. Images appeared in random order within a block. All participants completed the pretest in less than 40 min and the main test in less than 1.5 hr including breaks.

TABLE 2: Definition of Hit, False Alarm, Miss, and Correct Rejection According to SDT (Green & Swets, 1966)

Stimulus	Target-present response	Target-absent response
Target-present stimulus	Hit	Miss
Target-absent stimulus	False alarm	Correct rejection

Note. SDT = signal detection theory (Green & Swets, 1966).

Analyses

We computed analyses of covariance (ANCOVA) with detection performance (d′), target-absent RT, and target-present RT as dependent variables and age and 2D work experience as covariates (using SPSS version 22 and an alpha level of .05). Age was used as covariate because 3D screeners were, on average, younger than 2D screeners (see Table 1) and because previous research showed a negative correlation between age and the visual inspection performance of screeners (Ghylin, Drury & Schwaninger, 2006; Schwaninger et al., 2010). 2D work experience was used as covariate because 2D screeners had on average more 2D work experience than 3D screeners (see Table 1). We conducted post hoc comparisons with R version 3.22 (R Core Team, 2015) and applied Holm–Bonferroni corrections (Holm, 1979). We report ANCOVA effect sizes with η_p^2; effect sizes of *t* tests, with Cohen's *d*.

According to SDT (Green & Swets, 1966), there are four possible outcomes depending on stimuli and participant responses (Table 2). Detection performance (d′) was calculated using the following SDT formulae, whereby *z* refers to the inverse of the cumulative distribution function of the standard normal distribution (Green & Swets, 1966; MacMillan & Creelman, 2005):

$$Hit\ Rate\,(HR) = Hits\,/\,(Hits + Misses) \quad (1)$$

$$False\ Alarm\ rate\,(FAR) = False\ Alarms\,/\,(False\ Alarms + Correct\ Rejections) \quad (2)$$

$$d' = z(HR) - z(FAR) \quad (3)$$

RESULTS

Image Quality

Detailed results on image quality assessment with six tests of the STP are reported in the Appendix. In summary, results confirmed that the 2D imaging system passed all image quality tests. The 3D imaging system did not pass two of the six tests: The spatial resolution and useful penetration tests could not be solved using either the 3D-rotatable or the 3D-sliceable image. Nonetheless, taking all test results into account, it should be possible to recognize main IED components (triggering devices, power sources, explosives, and detonators) to a similar degree with 2D and 3D imaging. However, recognizing thin wires when they are hidden behind aluminum of a thickness of 7.9 mm or more was not possible with 3D imaging.

Visual Inspection Performance

We first present the results on detection performance (d′) because this is the main dependent variable for addressing our research questions. We then present RT, whereby target-absent RTs are more important due to the fact that they account for about 98% of all X-ray images in real-world hold baggage screening when using threat image projection (Hofer & Schwaninger, 2005; Schwaninger, 2006; Schwaninger, Hofer, & Wetter, 2007; Schwaninger et al., 2010). Figure 2 shows detection performance d′ depending on X-ray imaging technology (2D vs. 3D imaging) and screener group (2D vs. 3D screeners).

A 2 (2D vs. 3D imaging) × 2 (2D vs. 3D screeners) ANCOVA with d′ as dependent variable while controlling for age and 2D work experience revealed a trend toward better detection performance (d′) with 3D imaging (mean values of main effect: 2D imaging d′ = 1.80; 3D

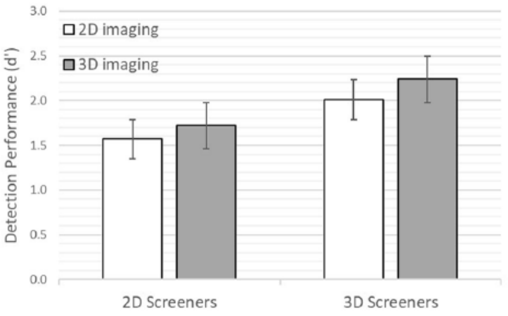

Figure 2. Detection performance (d′) by X-ray imaging technology (2D vs. 3D imaging) and screener group (2D vs. 3D screeners). Error bars are ± one standard error.

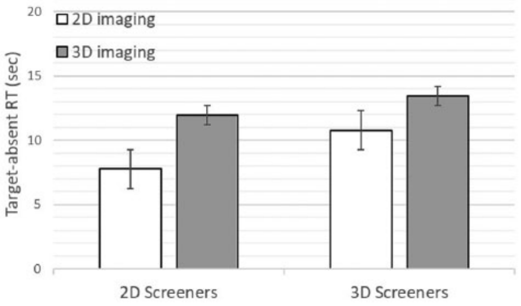

Figure 3. Target-absent RT by X-ray imaging technology (2D vs. 3D imaging) and screener group (3D vs. 2D screeners). Error bars are ± one standard error.

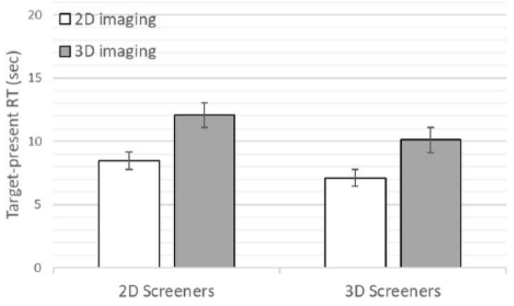

Figure 4. Target-present RT by X-ray imaging technology (2D vs. 3D imaging) and screener group (2D vs. 3D screeners). Error bars are ± one standard error.

imaging d′ = 1.97). However, this effect did not attain statistical significance, $F(1, 78) = 3.56$, $p = .065$, $\eta_p^2 = .04$. There was a significant effect of screener group with 3D screeners performing better with both types of imaging than 2D screeners (mean values of main effect: 2D screeners d′ = 1.72; 3D screeners d′ = 2.05), $F(1, 78) = 10.18$, $p = .002$, $\eta_p^2 = .12$. The interaction between imaging and screener group was not significant. There was a significant effect of the covariate age, $F(1, 79) = 2.86$, $p < .001$, $\eta_p^2 = .16$ but not of the covariate 2D work experience.

Figure 3 shows target-absent RT by X-ray imaging technology and screener group.

We calculated a 2 (2D vs. 3D imaging) × 2 (2D vs. 3D screeners) ANCOVA for target-absent trials with RT as dependent variable while controlling for age and 2D work experience. We found a main effect of imaging $F(1, 78) = 12.12$, $p < .001$, $\eta_p^2 = .13$, and screener group, $F(1, 78) = 11.22$, $p < .001$, $\eta_p^2 = .13$, but no significant effect for their interaction. Further, there was a significant effect of the covariate age, $F(1, 78) = 7.75$, $p = .007$, $\eta_p^2 = .09$, but not of the covariate 2D work experience. To examine whether speed–accuracy trade-offs can explain why 3D screeners had higher detection performance (d′) than 2D screeners with both imaging systems, we used two-tailed independent samples *t* tests to examine accuracy in target-absent trials (percent correct rejections, PCR). For 2D imaging, 2D screeners had significantly higher PCR than 3D screeners, $t(42) = -3.88$, $p < .001$. For 3D imaging, we did not find a difference between the screener groups for PCR, $t(38) = -.00$, $p = .997$. This means that we found no evidence that the better detection performance (d′) of 3D screeners compared with 2D screeners could be explained by a speed–accuracy trade-off in target-absent trials.

Figure 4 shows target-present RT dependent on X-ray imaging technology and screener group.

We calculated a 2 (2D vs. 3D imaging) × 2 (2D vs. 3D screeners) ANCOVA for target-present trials with RT as dependent variable

TABLE 3: Correlations Between Speed (Target Absent and Target Present RT) and Detection Performance (d′) Controlling for Age and Work Experience

	3D Screener Group		2D Screener Group	
Trial type	2D imaging	3D imaging	2D imaging	3D imaging
Target-present trials	$r = -.01$ $p = .97$	$r = -.33$ $p = .10$	$r = .03$ $p = .92$	$r = .22$ $p = .36$
Target-absent trails	$r = .28$ $p = .20$	$r = .25$ $p = .31$	$r = .16$ $p = .49$	$r = .19$ $p = .40$

while controlling for age and 2D work experience. We found a main effect of imaging, $F(1, 78) = 20.32, p < .001, \eta_p^2 = .21$, and a significant effect of the covariate age, $F(1, 78) = 25.52, p < .001, \eta_p^2 = .25$ but not of the covariate 2D work experience. Neither the main effect of screener group nor the interaction was significant, making a speed–accuracy trade-off an implausible explanation for the better detection performance (d′) of 3D screeners compared with 2D screeners.

To examine speed-accuracy trade-offs within the screener groups, we also calculated two-tailed partial correlations between response times and detection performance (d′) while controlling for age and work experience (Table 3). A speed-accuracy trade-off would have been supported if at least one significant positive correlation (longer reaction times and higher detection performance [d′]) would have been found. This was not the case, which makes speed-accuracy trade-offs very unlikely.

DISCUSSION

This study addressed two questions of high practical and theoretical relevance for the airport security screening of hold baggage: (a) Can screeners achieve at least similar detection performance using 3D imaging compared with 2D imaging despite the lower image quality of 3D imaging? (b) Does visual inspection competency acquired with one type of imaging transfer to the other type of imaging? We addressed these questions by asking 2D screeners and 3D screeners to perform a simulated hold baggage screening task with both types of imaging. We first discuss the results on detection performance (d′), the main dependent variable for our research questions. We then discuss the results on response times (RT) whereby target-absent RTs are more meaningful for real-world baggage screening. We conclude by discussing implications of our results for the efficiency and effectiveness of hold baggage screening using 2D versus 3D imaging systems.

Despite lower image quality (see the Appendix for these results and their discussion), 3D imaging resulted in a similar detection performance (d′) of screeners compared with that for 2D imaging. Benefits of 3D imaging allowing three-dimensional rotation and slicing seem to compensate for the potentially negative effects of lower image quality. This is consistent with earlier research on cabin baggage screening that showed better detection performance for motion imaging compared with static 2D imaging (Mendes et al., 2013). 2D screeners achieved a similar detection performance (d′) with 3D imaging to that with 2D imaging. This indicates a very large transfer effect and has important practical implications in light of the current international discussions on whether specific training should be mandated for 2D screeners before allowing them to work with 3D imaging systems. Our results suggest that 2D screeners do not need extensive and specific training to achieve similar detection performance with 3D imaging compared with that attained with 2D imaging.

3D screeners also achieved similar detection performance (d′) with both imaging systems, but they performed better than 2D screeners with both types of imaging. As explained in the introduction, object recognition research has shown that exposure to 3D images results in richer visual representations that could therefore also increase detection performance in 2D images

(Tarr & Vuong, 2002; Vuong & Tarr, 2004). This is a plausible explanation for our finding that 3D screeners performed better than 2D screeners not only with 3D imaging but also with 2D imaging. Alternative explanations might be based on group differences in age, cognitive abilities, training, or work experience along with speed–accuracy trade-offs. Because we used age as covariate, age differences are an unlikely explanation for performance differences between 2D and 3D screeners. Visual-cognitive abilities have also been shown to impact on screener performance (Hardmeier & Schwaninger, 2008; Rusconi, Ferri, Viding, & Mitchener-Nissen, 2015; Rusconi, McCrory, & Viding, 2012; Schwaninger, Hardmeier, & Hofer, 2005). However, it is also unlikely that differences in these abilities can explain the detection performance differences between 3D and 2D screeners in our study. The organization providing the 2D screeners had implemented a very selective pre-employment screening procedure including a visual-cognitive test battery and an X-ray object recognition test (Hardmeier, Hofer, & Schwaninger, 2006; Schwaninger, Hardmeier, & Hofer, 2005). Moreover, it is difficult to explain differences between 2D and 3D screeners by amount of training because both screener groups were qualified, trained, and certified according to the same European standards including a 6-hr mandatory image recognition training and testing every 6 months (European Commission, 2015). Finally, differences in 2D work experience cannot explain why 3D screeners were better with 2D imaging than 2D screeners, because the latter had more work experience with 2D imaging, and 2D work experience was used as covariate. Thus, the most plausible explanation based on results from object-recognition research (Tarr & Vuong, 2002; Vuong & Tarr, 2004) would seem to be that extensive exposure to 3D imaging during work and training resulted in richer visual representations and therefore better performance of 3D screeners than 2D screeners for both types of imaging.

The target-absent RT of 2D screeners when using 2D imaging was 8 s. Threat image projection data from experienced 2D screeners working with a similar 2D imaging system revealed target-absent RTs of about 7 s (Schwaninger, Hofer, & Wetter, 2007). This suggests that the target-absent RT found in our study would generalize quite well to real-world conditions (at least for 2D screeners when using 2D imaging) despite large differences in target prevalence. Both screener groups needed more time (about 2 s [3D screeners] and 4 s [2D screeners]) when using 3D imaging compared with 2D imaging. This result was anticipated, because rotating and slicing 3D images takes longer to process than the visual inspection of static 2D X-ray images. When no target was present, 3D screeners took longer for visual inspection than 2D screeners. For 3D imaging, the difference was small (about 1 s). For 2D imaging, 3D screeners took 3 s longer than 2D screeners. Although speculative, one possible explanation could be that 3D screeners were used to rotating and slicing images but were unable to do this when using 2D imaging. This may have resulted in longer target-absent RT. However, the important result is that the higher detection performance (d′) of 3D screeners with both imaging systems compared with 2D screeners could not be explained by a speed–accuracy trade-off.

The target-present RT of 2D screeners when using 2D imaging was 8 s. This was similar to the real-world target-present RT of 9 s for experienced 2D screeners when using 2D imaging for hold baggage screening (Schwaninger, Hofer, & Wetter, 2007). This provides further support for the view that the RT found in our study would generalize to real-world conditions despite large differences in target prevalence. As for target-absent RT, both screener groups needed more time: 3 s (3D screeners) and 4 s (2D screeners) when using 3D imaging. Differences between screener groups were not significant for target-present RT, making a speed–accuracy trade-off an extremely implausible explanation for the better detection performance (d′) of 3D screeners compared with 2D screeners with both imaging systems.

Whereas 2D work experience did not have an impact, age had an influence on all dependent variables: Older screeners had lower detection performance (d′) and longer response times. This result is consistent with previous research showing a negative correlation between age and visual inspection performance of screeners

TABLE 4: Estimation of Efficiency Increase (Throughput) When Using 3D Imaging Compared With 2D Imaging Based on Target-Absent RT Results

Scenario	Bags per hour	EDS-HBS FAR	Approval capacity	Efficiency increase Level 1	Bags sent to visual inspection	Target absent RT [sec]	Visual inspection time [hr]	Efficiency increase Level 2
2D screeners / 2D imaging	1,500	35%	975		525	8	1.2	
2D screeners / 3D imaging	1,500	15%	1,275	31%	225	12	0.8	36%
3D screeners / 2D imaging	1,500	35%	975		525	11	1.6	
3D screeners / 3D imaging	1,500	15%	1,275	31%	225	13	0.8	49%

Note. EDS = explosive detection systems; HBS = hold baggage screening; FAR = false alarm rate; RT = response time.

(Ghylin et al., 2006; Schwaninger et al., 2010). Because we used 2D work experience and age as covariates, the observed screener group differences in detection performance (d′) and response times cannot be explained by preexisting differences in the covariates.

To summarize, the results on detection performance (d′) answered our two research questions: (a) Screeners achieved a similar detection performance (d′) using 3D imaging compared with 2D imaging despite lower image resolution of 3D imaging. (b) Visual inspection competency acquired with one type of imaging transferred to visual inspection with the other type of imaging. However, both screener groups needed more time (2–4 s) when using 3D imaging compared with 2D imaging.

What do our results on screeners' visual inspection performance mean for the efficiency (throughput) of 2D versus 3D hold baggage screening at airports? According to Oftring (2015), 2D and 3D imaging systems can process about 1,500 bags per hour, but 2D imaging systems have false alarm rates of at least 35%, whereas 3D imaging systems achieve much lower false alarm rates (15%). The installation of 3D imaging (Level 1 in hold baggage screening) should therefore already result in a 31% increase in efficiency. Based on the amount of bags sent to visual inspection and the target-absent RTs found in our study, an efficiency increase from 36% to 49% on Level 2 of hold baggage screening (alarm resolution of screeners) can be achieved (see Table 4 for the calculation). As explained in the introduction, if screeners decide that an X-ray image is suspicious, more time-consuming investigations follow including rescreening with other X-ray technology, trace detection, explosive detection, dogs, passenger reconciliation, and the opening of bags (Shanks & Bradley, 2004; Singh & Singh, 2003). Therefore, efficiency gains will be even higher in practice because 3D imaging results in less hold baggage being sent to Level 2.

Estimating the increase in effectiveness (detection of IEDs) is more difficult, because the detection rates of 2D and 3D imaging systems are not publicly available for security reasons. However, it is clear that 3D imaging systems achieve substantially higher detection of explosives than 2D imaging systems (e.g., Oftring, 2015; Singh & Singh, 2003; Wells & Bradley, 2012). Moreover, in Europe, EDS-HBS have to meet European detection standards and be approved by test centers of the European Civil Aviation Conference (ECAC). So far, only 3D imaging systems have met ECAC Standard 3, whereas 2D imaging systems achieve only Standard 2 (European Civil Aviation Conference, 2018). ECAC Standard 3 requires higher hit rates and lower false alarm rates and therefore higher detection performance (d′) of EDS-HBS.

Taking together the results of our study on screeners' visual inspection performance with the performance advantages of 3D imaging technology, it is reasonable to infer that the whole human-machine system performance when using 3D imaging technology is superior to 2D imaging not only in terms of efficiency (throughput) but also in terms of effectiveness (detection of IEDs) of the HBS process as a whole. The results of our study further suggest that extensive and specific training is not needed for 2D screeners before allowing them to work with 3D imaging systems. Nonetheless, some limitations do call for further research: Screener performance was tested with only one 2D and 3D imaging system. It would be interesting to see whether different results would be obtained with other 2D systems using a larger angular difference between the two views of a bag (e.g., 60–90 deg), with 3D systems that have higher image resolution, and with hybrid systems that show four views (3D-rotatable, 3D sliceable, and two different STP-compliant 2D views). Although it is not possible to conclude from our study that higher image resolution of 3D imaging systems would result in better visual inspection performance among screeners, it would be worth investigating this in future studies. Second, it would be interesting to see whether the results of our study can be replicated with screeners from other airports using a within-subjects design to investigate transfer effects from 2D to 3D imaging and vice versa over several months (although this might be rather difficult to achieve in practice). Conducting such a study with student participants is not an option for reasons of external validity as well as the security-sensitive nature of the image material and on-screen alarm resolution protocols.

Despite these limitations, we believe that our study is robust enough to make a significant contribution to the theory, practice, and knowledge base of human factors and ergonomics—particularly with regard to its practical relevance. First, we can recommend a wide-scale implementation of 3D imaging systems with an image quality equal to or higher than that of the 3D imaging system tested in this study, because it can be expected to result in better human–machine system performance in terms of efficiency and effectiveness of the hold baggage screening process as a whole. Second, due to large transfer effects, 2D screeners do not require extensive and specific training to achieve similar detection performance with state-of-the-art 3D imaging. Third, image quality standards and procedures need revision before they can be applied to 3D imaging systems.

APPENDIX

Assessment of Image Quality

The most widely used international standard for assessing the image quality of X-ray imaging systems is the standard test piece (STP) and a procedure developed about 30 years ago for 2D imaging systems (WG Standard Test Piece, n.d.). Whereas 2D imaging systems comply with the STP standard, this is not the case for many 3D imaging systems. This is not surprising given the fact that the STP was developed for 2D imaging and that there is no specific image quality assessment procedure available yet for 3D imaging. Some countries have implemented 3D imaging systems at many of their airports (Oftring, 2015) because they have better automated explosive detection capability and higher baggage throughput (Flitton et al., 2013; Mouton & Breckon, 2015; Wells & Bradley, 2012). Other countries hesitate to change from 2D imaging to 3D imaging because it is unclear whether screeners achieve a similar detection performance using 3D imaging compared with 2D imaging due to the lower image quality of 3D imaging.

Our study evaluated screener performance using 2D and 3D imaging while also evaluating image quality differences between the two imaging systems using the STP. We used the STP to provide a quantitative measure of the differences of the 2D and 3D imaging systems tested in this study. We first explained the STP and the image quality assessment procedure and then presented the results for the 2D and 3D imaging systems used in our study. Based on the results, we discussed whether current image quality standards for 2D imaging need to be revised before they can be applied to 3D imaging systems.

The STP contains samples of materials of varying density and needs to be X-rayed with the tested machine. Based on the X-ray image of the STP, six tests are carried out to assess single wire resolution, useful penetration, spatial resolution, simple penetration, and material

TABLE A1: Results of Image Quality Tests for the 2D and 3D Imaging Systems Used in This Study

		3D imaging system		
Test	Requirement	3D-rotatable image	3D-sliceable image	2D imaging system
Test 1 STP: Single wire resolution	Ability to display a single thin wire (30 American Wire Gauge = 0.254 mm) when not covered by the aluminum step wedge.	Yes	Yes	Yes
Test 2 STP: Useful penetration	Wire (24 American Wire Gauge = 0.5105 mm) needs to be visible behind different thickness of aluminum (4.8 mm, 7.9 mm, and 11.1 mm).	4.8mm: No 7.9mm: No 11.1mm: No	4.8mm: Yes 7.9mm: No 11.1mm: No	4.8 mm: Yes 7.9 mm: Yes 11.1 mm: Yes
Test 3 STP: Spatial resolution	Ability to distinguish and display objects that are close together; gaps between the relevant vertical and horizontal gratings can be seen (2.0 mm slots on a 4.0 mm pitch).	No	No	Yes
Tests 4 & 5 STP: Simple penetration	Thin materials: The relevant steel plate (0.10 mm thick) can be seen.	No	Yes	Yes
	Thick materials: The lead bar (1.5 mm thick) can be seen behind 14 mm of steel.	No	Yes	Yes
Test 6 STP: Material discrimination	Different colors are allocated to the sample of organic and inorganic substances (sugar and salt discrimination).	Yes	Yes	Yes

Note. "Yes" means the requirement of the STP is fulfilled. For a machine to be STP compliant, it must pass all tests. STP = standard test piece.

discrimination. For each measure, certain requirements need to be fulfilled for the machine to pass the test. Test 1: Single Wire Resolution. This defines the ability to display a single thin wire. Test 2: Useful Penetration. This determines what level of detail should be seen behind a thickness of known material. Test 3: Spatial Resolution. This defines the ability to distinguish and display objects that are close together. Tests 4 and 5: Simple Penetration. These test the X-ray machine's ability to image thin and thick material as well as the thickness of steel the X-ray machine should be able to penetrate. Test 6: Material Discrimination. This ensures that different colors are allocated to organic and inorganic substances.

In our study, we used one 2D multiview X-ray and one 3D CT imaging system; both are operational at airports and representative of their category (the names of the systems cannot be revealed for this publication, but we can state that the 3D imaging system belongs to the most widely used in the world). A certified European test center conducted the image quality assessment using the STP. The results are shown in Table A1.

As expected, the 2D imaging system was STP compliant; that is, it passed all tests. The 3D imaging system did not pass two of the six tests: The spatial resolution and useful penetration tests could not be solved using either the 3D-rotatable or the 3D-sliceable image. However, when comparing the STP results of both systems, it can be assumed that it should be possible to recognize main IED components (triggering devices, power sources, explosives, and detonators) to a similar degree with 2D and 3D imaging. Nonetheless, recognizing thin wires when they are hidden behind aluminum of a thickness of 7.9 mm or more is not possible with 3D imaging.

To summarize, 3D imaging systems have lower image quality than 2D imaging systems according to tests using the STP and protocol, which is currently the most widely used international standard for assessing the image quality of 2D imaging systems. Despite this, our study could show that 2D and 3D screeners attained the same detection performance with 2D and 3D imaging. Based on the fact that newer 3D imaging technology has better automated explosive detection and therefore higher baggage throughput (Flitton et al., 2013; Mouton & Breckon, 2015; Oftring, 2015; Wells & Bradley, 2012), we argue that 3D imaging is superior to 2D imaging despite its lower image quality. Whether 3D screening is superior or inferior to 2D screening also depends on the visual inspection performance of X-ray screeners working with such systems. Therefore, when installing new technology at airports, it is important to consider not only technical features but also human factors.

Regarding image quality testing, no specific image quality assessment procedure is yet available for 3D imaging. Based on the results of our study, we argue that image quality standards and procedures need revision before they can be applied to 3D imaging systems. Regulatory bodies should not only evaluate technical aspects when testing image quality but also take human factors into account using experiments with highly realistic images, simulators, and screeners as participants as we did in our study.

ACKNOWLEDGMENTS

We thank the German Federal Police Technology Center for the valuable expertise and support for creating the stimulus material.

KEY POINTS

- This study compared the performance of airport security officers (screeners) using state-of-the-art 3D imaging and older 2D imaging for airport security screening of hold baggage.
- Despite lower image quality, screeners achieved a similar detection performance with 3D imaging to that for 2D imaging.
- 3D screeners revealed higher detection performance with both types of imaging than 2D screeners.
- Features of 3D imaging systems (3D rotation and slicing) seem to compensate for the lower image quality.
- Visual inspection competency acquired with one type of imaging seems to transfer to the other type of imaging.
- 2D and 3D screeners required more time for visual inspection of 3D versus 2D images. However, baggage throughput would still be substantially higher with 3D imaging systems for hold baggage screening due to lower EDS alarm rates than those for older 2D imaging systems.
- Replacing older 2D with newer 3D imaging systems for hold baggage screening can be recommended to increase the efficiency and effectiveness of hold baggage screening.
- Extensive and specific training of 2D screeners before allowing them to work with 3D imaging is not needed to achieve a similar performance to that with 2D imaging.
- Current image quality standards for 2D imaging need to be revised before they can be applied to 3D imaging systems for hold baggage screening.

REFERENCES

Barrat, H. H. (2000). *Handbook of medical imaging*. Bellingham, WA: SPIE Press.

Baum, P. (2016). *Violence in the skies: A history of aircraft hijacking and bombing*. Chichester, England: Summersdale Publishers.

Biggs, A. T., & Mitroff, S. R. (2014). Improving the efficacy of security screening tasks: A review of visual search challenges and ways to mitigate their adverse effects. *Applied Cognitive Psychology*, *29*(1), 142–148. doi:10.1002/acp.3083

Bolfing, A., Halbherr, T., & Schwaninger, A. (2008). How image based factors and human factors contribute to threat detection performance in X-ray aviation security screening. *HCI and Usability for Education and Work, Lecture Notes in Computer Science*, *5298*, 419–438. doi:10.1007/978-3-540-89350-9_30

Bretz, E. A. (2002). Slow takeoff. *IEEE Spectrum*, *39*(9), 37–39.

Caygill, J. S., Davis, F., & Higson, S. P. (2012). Current trends in explosive detection techniques. *Talanta*, *88*, 14–29. doi:10.1016/j.talanta.2011.11.043

Clark, K., Cain, M. S., Adamo, S. H., & Mitroff, S. R. (2012). Overcoming hurdles in translating visual search between the lab and the field. *Nebraska Symposium on Motivation*, 59, 147–181.

European Commission (2015, November 5). Commission implementing regulation (EU) 2015/1998 of 5 November

2015 laying down detailed measures for the implementation of the common basic standards on aviation security. *Official Journal of the European Union*. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015R1998&from=EN

European Civil Aviation Conference (ECAC). (2018, April). Retrieved from https://www.ecac-ceac.org/cep

Fiore, S. M., Scielzo, S., Jentsch, F., & Howard, M. L. (2006). Understanding performance and cognitive efficiency when training for X-ray security screening. *In Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 2610–2614). Santa Monica, CA: Human Factors and Ergonomics Society.

Fleck, M. S., & Mitroff, S. R. (2007). Rare targets are rarely missed in correctable search. *Psychological Science*, *18*, 943–947.

Flitton, G., Breckon, T., & Megherbi, N. (2010). Object recognition using 3D SIFT in complex CT volumes. In *British Machine Vision Conference* (pp. 11.1–11.12). Aberystwyth, Wales: BMVA Press. doi:10.5244/C.24.11

Flitton, G., Breckon, T., & Megherbi, N. (2013). A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery. *Pattern Recognition*, *46*(9), 2420–2436. doi:10.1016/j.patcog.2013.02.008

Franzel, T., Schmidt, U., & Roth, S. (2012). Object detection in multi-view X-ray images. In A. Pinz, T. Pock, H. Bischof, & F. Leberl (Eds.), *Pattern recognition: Joint 34th DAGM and 36th OAGM Symposium, Graz, Austria* (pp. 144–154). Berlin, Germany: Springer.

Ghylin, K. M., Drury, C. G., & Schwaninger, A. (2006). *Two-component model of security inspection: Application and findings*. 16th World Congress of Ergonomics, IEA 2006, Maastricht, The Netherlands, July, 10–14, 2006. doi:10.13140/RG.2.1.2216.8567

Godwin, H. J., Menneer, T., Cave, K. R., Helman, S., Way, R. L., & Donnelly, N. (2010). The impact of relative prevalence on dual-target search for threat items from airport X-ray screening. *Acta Psychologica*, *134*(1), 79–84. doi:10.1016/j.actpsy.2009.12.009

Godwin, H. J., Menneer, T., Cave, K. R., Thaibsyah, M., & Donnelly, N. (2015). The effects of increasing target prevalence on information processing during visual search. *Psychonomic Bulletin & Review*, *22*(2), 469–475. doi:10.3758/s13423-014-0686-2

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Halbherr, T., Schwaninger, A., Budgell, G. R., & Wales, A. W. J. (2013). Airport security screener competency: A cross-sectional and longitudinal analysis. *International Journal of Aviation Psychology*, *23*(2), 113–129. doi:10.1080/10508414.2011.582455

Hancock, P. A., & Hart, S. G. (2002). Defeating terrorism: What can human factors/ergonomics offer? *Ergonomics in Design*, *10*(1), 6–16.

Harding, G. (2004). X-ray scatter tomography for explosives detection. *Radiation Physics and Chemistry*, *71*, 869–881. doi:10.1016/j.radphyschem.2004.04.111.

Hardmeier, D., Hofer, F., & Schwaninger, A. (2006, June). Increased detection performance in airport security screening using the X-Ray ORT as pre-employment assessment tool. *Proceedings of the 2nd International Conference on Research in Air Transportation, ICRAT 2006, Belgrade, Serbia and Montenegro* (pp. 393–397). doi:10.5167/uzh-97986

Hardmeier, D., & Schwaninger, A. (2008, June). Visual cognition abilities in X-ray screening. *Proceedings of the 3rd International Conference on Research in Air Transportation, ICRAT 2008* (pp. 311–316). Fairfax, VA. doi:10.13140/RG.2.1.4335.7924

Harris, D. H. (2002). How to really improve airport security. *Ergonomics in Design*, *10*(1), 17–22.

Hofer, F., & Schwaninger, A. (2005). Using threat image projection data for assessing individual screener performance. *WIT Transactions on the Built Environment*, *82*, 417–426. doi:10.2495/SAFE050411

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Koller, S., Drury, C., & Schwaninger, A. (2009). Change of search time and non-search time in X-ray baggage screening due to training. *Ergonomics*, *52*(6), 644–656. doi:10.1080/00140130802526935

Koller, S., Hardmeier, D., Michel, S., & Schwaninger, A. (2008). Investigating training, transfer and viewpoint effects resulting from recurrent CBT of X-ray image interpretation. *Journal of Transportation Security*, *1*(2), 81–106. doi:10.1007/s12198-007-0006-4

Kuhn, M. (2017). Centralised image processing: The impact on security checkpoints. *Aviation Security International*, *23*(5), 28–30.

Lau, J. S., & Huang, L. (2010). The prevalence effect is determined by past experience, not future prospects. *Vision Research*, *50*(15), 1469–1474. doi:10.1016/j.visres.2010.04.020.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

McCarley, J. S. (2009). Response criterion placement modulates the benefits of graded alerting systems in a simulated baggage screening task. *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (pp. 1106–1110). Santa Monica, CA: Human Factors and Ergonomics Society.

McCarley, S., Kramer, A. F., & Wickens, C. D. (2004). Visual skills in airport-security screening. *Psychological Science*, *15*(5), 302–306. doi:10.1111/j.0956-7976.2004 .00673.x

Mendes, M., Schwaninger, A., & Michel, S. (2013). Can laptops be left inside passenger bags if motion imaging is used in X-ray security screening? *Frontiers in Human Neuroscience*, *7*, 1–10. doi:10.3389/fnhum.2013.00654

Menneer, T., Donnelly, N., Godwin, H. J., & Cave, K. R. (2010). High or low target prevalence increases the dual-target cost in visual search. *Journal of Experimental Psychology. Applied*, *16*(2), 133–144. doi:10.1037/a0019569

Mitroff, S. R., Biggs, A. T., & Cain, M. S. (2015). Multiple-target visual search errors: Overview and implications for airport security. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 121–128. doi:10.1177/237273221560

Mouton, A., & Breckon, T. P. (2015). A review of automated image understanding within 3D baggage computed tomography security screening. *Journal of X-ray Science and Technology*, *23*(5), 531–555. doi:10.3233/XST-150508

Oftring, C. (2015). Assessing the impact of ECAC3 on baggage handling systems – considerations for upgrading existing ECAC2 systems. Retrieved from https://www.copybook.com/media/airport/profiles/beumer/documents/1464176454_ECAC%20Standard%203.pdf

R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rice, S., & McCarley, J. (2011). Effects of response bias and judgment framing on operator use of an automated aid in a target detection task. *Journal of Experimental Psychology: Applied*, *17*(4), 320–331.

Rich, A. N., Kunar, M. A., Van Wert, M. J., Hidalgo-Sotelo, B., Horowitz, T. S., & Wolfe, J. M. (2008). Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *Journal of Vision*, *8*(15), 1–17. doi:10.1167/8.15.15

Rusconi, E., Ferri, F., Viding, E., & Mitchener-Nissen, T. (2015). XRIndex: A brief screening tool for individual differences in security threat detection in X-ray images. *Frontiers in Human Neuroscience*, *9*, 439. doi:10.3389/fnhum.2015.00439

Rusconi, E., McCrory, E., & Viding, E. (2012). Self-rated attention to detail predicts threat detection performance in security X-ray screening. *Security Journal*, *25*, 356–371. doi:10.1057/sj.2011.27

Schuster, D., Rivera, J., Sellers, B. C., Fiore, S. M., & Jentsch, F. (2013). Perceptual training for visual search. *Ergonomics*, *56*(7), 1101–1115. doi:10.1080/00140139.2013.790481.

Schwaninger, A. (2005). Increasing efficiency in airport security screening. *WIT Transactions on the Built Environment*, *82*, 407–416. doi:10.2495/SAFE050401

Schwaninger, A. (2006). Threat image projection: Enhancing performance? *Aviation Security International*, 36–41.

Schwaninger, A., Hardmeier, D., & Hofer, F. (2005). Aviation security screeners' visual abilities and visual knowledge measurement. *IEEE Aerospace and Electronic Systems*, *20*(6), 29–35.

Schwaninger, A., Hardmeier, D., Riegelnig, J., & Martin, M. (2010). Use it and still lose it? The influence of age and job experience on detection performance in X-ray. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry*, *23*(3), 169–175. doi:10.1024 /1662-9647/a000020

Schwaninger, A., & Hofer, F. (2004). Evaluation of CBT for increasing threat detection performance in X-ray screening. In K. Morgan & M. J. Spector (Eds.), *The Internet society: Advances in learning, commerce and security* (pp. 147–156). Southampton, England: WIT Press. doi:10.13140/RG.2.1.4051.8649

Schwaninger, A., Hofer, F., & Wetter, O. E. (2007). Adaptive computer-based training increases on the job performance of x-ray screeners. *Proceedings of the 41st Carnahan Conference on Security Technology*, *Ottawa, October 8–11, 2007*. doi:10.1109/CCST.2007.4373478

Schwaninger, A., Michel, S., & Bolfing, A. (2005). Towards a model for estimating image difficulty in X-ray screening. *Proceedings of the 39th Carnahan Conference on Security Technology*, 39, 185–188. doi:10.1109/CCST.2005.1594875

Schwaninger, A., Michel, S., & Bolfing, A. (2007). A statistical approach for image difficulty estimation in X-ray screening using image measurements. *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization* (pp. 123–130). New York, NY: ACM Press. doi:10.1145/1272582.1272606

Shanks, N. E. L., & Bradley, A. L. W. (2004). *Handbook of checked baggage screening: Advanced airport security operation*. London, England: Professional Engineering Publishing.

Singh, S., & Singh, M. (2003). Explosives detection systems (EDS) for aviation security. *Signal Processing*, 83(1), 31–55. doi:10.1016/S0165-1684(02)00391-2

Strantz, N. J. (1990) Aviation security and Pan Am Flight 103: What have we learned. *Journal of Air Law and Commerce*, *56*, 413.

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics*. Mahwah, NJ: Erlbaum.

Tarr, M. J., & Vuong, Q. C. (2002). Visual object recognition. In H. Pashler (Series Ed.) & S. Santis (Ed.), *Stevens' handbook of experimental psychology: Vol. 1. Sensation and perception* (3rd ed., Vol. 1, pp. 287–314). New York, NY: Wiley. doi:10.1002/0471214426.pas0107

Turner, S. (1994). *Terrorist explosive sourcebook countering terrorist use of improvised explosive devices*. Boulder, CO: Paladin Press.

Verghese, P. (2001). Visual search and attention: A signal detection approach. *Neuron*. *31*, 523–535. doi:10.1016/S0896-6273(01)00392-0

Vuong, Q. C., & Tarr, J. T. (2004). Rotation direction affects object recognition. *Vision Research*, *44*, 1717–1730. doi:10.1016/j.visres.2004.02.002

von Bastian, C. C., Schwaninger, A., & Michel, S. (2008). Do multi-view X-ray systems improve X-ray image interpretation in airport security screening? *Zeitschrift für Arbeitswissenschaft*, *3*, 166–173. doi:10.3239/9783640684991

Wales, A., Anderson, C., Jones, K., Schwaninger, A., & Horne, J. (2009). Evaluating the two-component inspection model in a simplified luggage search task. *Behavior Research Methods*, *41*(3), 937–943. doi:10.3758/BRM.41.3.937

Wells, K., & Bradley, D. A. (2012) A review of X-ray explosives detection techniques for checked baggage. *Applied Radiation and Isotopes*, *70*(8), 1729–1746. doi:10.1016/j.apradiso.2012.01.011

Wetter, O. E. (2013). Imaging in airport security: Past, present, future, and the link to forensic and clinical radiology. *Journal of Forensic Radiology and Imaging*, 1(4), 152–160. doi:10.1016/j.jofri.2013.07.002

WG Standard Test Piece (n.d.). Retrieved April 4, 2018 from https://www.wi-ltd.com/wp-content/uploads/2016/03/WG-X-Ray-Machine-Standard-Test-Piece-STP.pdf

Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of Vision*, *13*(3), 33. doi:10.1167/13.3.33

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, *435*, 439–440. doi:10.1038/435439a

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, *136*(4), 623–638. doi:10.1037/0096-3445.136.4.623

Wolfe. J. M., & Reynolds, J. H. (2008). Visual search. In A. I. Basbaum, A. Kaneko, G. M. Shepherd, & G. Westheimer (Eds.), *The senses: A comprehensive reference* (Vol. 2, pp. 275–280). San Diego, CA: Academic Press.

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two, dissociable decision criteria in visual search. *Current Biology*, *20*: 121–124. doi:10.1016/j.cub.2009 .11.066.

Nicole Hättenschwiler is a PhD student working at the University of Applied Sciences and Arts, Northwestern Switzerland, in the field of human factors in aviation security. She obtained her Master of Science in Psychology from the University of Bern in 2014.

Marcia Mendes earned her PhD in the field of human factors in aviation security from the University of Basel in 2016.

Adrian Schwaninger received his PhD in psychology from the University of Zurich in 2003. Since 2008, he has been Professor of Psychology at the Institute Humans in Complex Systems of the School of Applied Psychology of the University of Applied Sciences and Arts Northwestern Switzerland. Since 2009, he has been the Head of this Institute.

Date received: September 20, 2017
Date accepted: August 15, 2018