International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Benefits of Decision Support Systems in Relation to Task Difficulty in Airport Security X-Ray Screening

David Huegli, Sarah Merks & Adrian Schwaninger

To cite this article: David Huegli, Sarah Merks & Adrian Schwaninger (2022): Benefits of Decision Support Systems in Relation to Task Difficulty in Airport Security X-Ray Screening, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2022.2107775

To link to this article: https://doi.org/10.1080/10447318.2022.2107775

Published online: 15 Aug 2022.

INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2022.2107775

Benefits of Decision Support Systems in Relation to Task Difficulty in Airport Security X-Ray Screening

David Huegli[a], Sarah Merks[b], and Adrian Schwaninger[a]

[a]University of Applied Sciences and Arts Northwestern Switzerland, School of Applied Psychology, Institute Humans in Complex Systems, Olten, Switzerland; [b]Center for Adaptive Security Research and Applications (CASRA), Zürich, Switzerland

ABSTRACT

Automated explosive detection systems for cabin baggage screening (EDSCB) highlight areas in X-ray images of passenger bags that could contain explosive material. Several countries have implemented EDSCB so that passengers can leave personal electronic devices in their cabin baggage. This increases checkpoint efficiency, but also task difficulty for screeners. We used this case to investigate whether the benefits of decision support systems depend on task difficulty. 100 professional screeners conducted a simulated baggage screening task. They had to detect prohibited articles built into personal electronic devices that were screened either separately (low task difficulty) or inside baggage (high task difficulty). Results showed that EDSCB increased the detection of bombs built into personal electronic devices when screened separately. When electronics were left inside the baggage, operators ignored many EDSCB alarms, and many bombs were missed. Moreover, screeners missed most unalarmed explosives because they over-relied on the EDSCB's judgment. We recommend that when EDSCB indicates that the bag might contain an explosive, baggage should always be examined further in a secondary search using explosive trace detection, manual opening of bags and other means.

1. Introduction

Automation has many functional uses including autopilots, expert systems, or alarm systems. Decision support systems are one type of automation that supports human information acquisition and analysis in order to improve decision making (Mosier & Manzey, 2020). They are used in different applied fields related to informatics (e.g., Meng et al., 2021; Qinxia et al., 2021; Xiaolong et al., 2021; Zhang et al., 2020), health care (e.g., Alberdi et al., 2008; Drew et al., 2012; Xiao et al., 2021), aviation safety (e.g., Mosier et al., 1998; Sarter et al., 2001), or airport security (e.g., Goh et al., 2005; Hättenschwiler et al., 2018; Huegli et al., 2020; Petrozziello & Jordanov, 2019; Rieger et al., 2021; Wiegmann et al., 2006). Typical examples of decision support systems are alarm systems or detection systems (Rieger et al., 2021) that provide cues regarding whether a target is present or not. In the current study, the use case is explosives detection systems for cabin baggage (EDSCB) in airport security X-ray screening (Hättenschwiler et al., 2018; Huegli et al., 2020). With modern EDSCB, airports can allow passengers to leave personal electronic devices (PEDs) such as notebooks inside their baggage. This increases passenger throughput at security checkpoints (Sterchi & Schwaninger, 2015), but also task difficulty for X-ray screeners. There are four approaches to deal with PEDs at airport security checkpoints in the presence of EDSCB: (a) Passengers must remove PEDs from their cabin baggage and screeners use EDSCB as a decision support and resolve alarms on-screen while inspecting the X-ray image. (b) Passengers can leave PEDs in their baggage, and screeners resolve EDSCB alarms on-screen while inspecting the X-ray image. (c) Passengers must remove PEDs from their baggage, and alarmed bags are sent directly to secondary search. (d) Passengers can leave PEDs in their baggage, and alarmed bags are sent directly to secondary manual search.

The contribution of the present study is to give recommendations whether PEDs can be left in cabin baggage and how EDSCB alarms should be resolved. The criteria used to evaluate these options are the detection (human–machine system hit rate) of bombs built into PEDs when placed inside bags (high task difficulty) compared to when PEDs were screened separately (low task difficulty). We also considered the human–machine system false alarm rate and response times of screeners.

1.1. Practical background

At airport security checkpoints, cabin baggage is screened using X-ray machines to prevent passengers from carrying

CONTACT David Huegli david.huegli@fhnw.ch University of Applied Sciences and Arts Northwestern Switzerland, School of Applied Psychology, Institute Humans in Complex Systems, Olten, Switzerland

prohibited articles (e.g., guns, knives, bombs) onto an airplane (Harris, 2002). The actual target prevalence of prohibited articles at these checkpoints is extremely low. However it is increased artificially to about 2–4% by using threat image projection, which is a technology that projects prerecorded X-ray images of prohibited articles onto X-ray images of real passenger bags (Hofer & Schwaninger, 2005; Meuter & Lacherez, 2016; Schwaninger et al., 2010; Skorupski & Uchroński, 2016), and covert tests with simulated prohibited articles (Schwaninger, 2009; Wetter et al., 2008).

The task of the screeners is to visually inspect X-ray images of cabin baggage for prohibited articles. This involves visual search and decision making (Koller et al., 2009; Wales et al., 2009; Wolfe & Van Wert, 2010). Bombs are the most dangerous prohibited articles in passenger baggage and are technically called improvised explosive devices (IEDs). IEDs are composed of a power source, a triggering device, a detonator, and an explosive charge usually connected by wires (Turner, 1994; Wells & Bradley, 2012). Because of these components, well-trained screeners can detect IEDs well (Halbherr et al., 2013; Koller et al., 2008, 2009; Schuster et al., 2013; Schwaninger & Hofer, 2004), but when an IED is built into a PED, its components are often concealed. Research has shown that detecting targets is more difficult in a densely packed bag than in a plastic tray (e.g., Adamo et al., 2015; Rosenholtz et al., 2007), and this is also the case when prohibited articles are built into PEDs (Mendes et al., 2013). For these reasons, at most airports, passengers must remove PEDs from their bags to have them X-rayed. However, European regulation allows passengers to leave PEDs inside their luggage when state-of-the-art multiview EDSCB technologies (EDSCB of Standard C2) which achieve automation hit rates in the range of 75–90% and automation false alarm rates of 6–20% (Hättenschwiler et al., 2018; Sterchi & Schwaninger, 2015) are in use (Commission Implementing Regulation [EU], 2015). Although leaving PEDs inside cabin baggage increases task difficulty for screeners (Bolfing et al., 2008; Mendes et al., 2013; Schwaninger et al., 2007) it also leads to faster divesting times, fewer bags per person, and thus increasing passenger throughput at security checkpoints (Sterchi & Schwaninger, 2015). As a result, many airports have bought and implemented such state-of-the-art multiview EDSCB technologies. There are two common ways in which EDSCB alarms can be resolved: First, when using EDSCB as a decision support system, EDSCB highlight areas in X-ray images that might contain explosive material (Hättenschwiler et al., 2018; Huegli et al., 2020; Singh & Singh, 2003; Wells & Bradley, 2012), and airport security screeners resolve the alarm on-screen by deciding whether the EDSCB judgment is true or not, and further examination is needed. Hence, screeners conduct on-screen alarm resolution. An alternative is to use EDSCB as an automated decision. That is, bags on which EDSCB has alarmed are sent directly to secondary search where alarm resolution takes place using explosive trace detection, manual bag search and other means (Michel et al., 2014; Sterchi & Schwaninger, 2015).

1.2. Theoretical background

When using automation as decision support, two concepts of operator behavior are particularly important: compliance and reliance (Dixon & Wickens, 2006; Meyer, 2001, 2004). Compliance is the degree to which operators respond to a decision support system when it provides an alarm. Reliance is how far operators act in line with a decision support system when it does not give an alarm. Compliance and reliance have been described as behavioral expressions of trust in automation (Hoff & Bashir, 2015; Lee & See, 2004). So far, research on how compliance and reliance relate to the subjective perception of automation's trustworthiness has been inconsistent (Avril et al., 2022; Chancey et al., 2017; Chavaillaz et al., 2016, 2019b; Dzindolet et al., 2003; Rovira et al., 2007).

In an ideal world, human operators properly assess the information provided by the decision support system and its validity. They should show compliance when it detects a target and shows reliance when it correctly advises that no target is present. However, operators sometimes use its recommendations to replace deliberate decision making and proper situation evaluation (Mosier et al., 1998; Mosier & Manzey, 2020; Parasuraman & Manzey, 2010). This can lead to less than ideal human–machine system performance, especially when target prevalence is low, which is the case at airport security checkpoints. Because most bags at airport security checkpoints do not contain IEDs, the majority of EDSCB alarms will be false alarms (on which EDSCB yields an alarm although no target is present) even when overall reliability is high (Madhavan et al., 2006; Onnasch et al., 2014; Parasuraman & Riley, 1997). Moreover, when EDSCB refrains from alarming, most of the bags do not contain IEDs. On one hand, operators experience a lot of false alarms and may ignore automation alarms when they lead to a cry-wolf effect (Breznitz, 1983). On the other hand, because the decision support system rarely misses a target, operators may rely too much on when it does not alarm. This behavior results in omission errors.

The cry-wolf effect (Breznitz, 1983) describes a lack of operator compliance resulting from exposure to many false alarms—a phenomenon that has been observed in different domains and tasks (Bliss et al., 1995; Bolton & Katok, 2018; Dixon et al., 2007; Dixon & Wickens, 2006; Huegli et al., 2020; Maltz & Shinar, 2003; Nishikawa & Bae, 2018; Parasuraman et al., 2000; Zirk et al., 2019). Low compliance results in operators ignoring some alarms (Alberdi et al., 2008; Manzey et al., 2014) and ignoring them entirely in extreme cases (Bliss et al., 1995). Ignoring a high number of alarms results in decreased human–machine system performance, even though using the highly reliable decision support system would increase target detection (Bartlett & McCarley, 2017; Boskemper et al., 2021; Dixon & Wickens, 2006; Meyer et al., 2014; Rice & McCarley, 2011).

Omission errors arise when the decision support system fails to detect a target and operators show too much reliance on automation recommendations due to inadequate monitoring (Davis et al., 2020; Mosier et al., 1998; Parasuraman & Riley, 1997). Then, operators miss the target and fail to react accordingly. Omission errors are one type of outcome of the well-investigated automation bias. The automation bias has been described as using automation as a heuristic replacement for informed decision making (Mosier et al., 1998; Parasuraman & Manzey, 2010). Whereas the cry-wolf effect results from exposure to many system false alarms, omission errors occur mainly with high system hit rates (Bailey & Scerbo, 2007; Rovira et al., 2007), because misses become rare and more likely to be overlooked.

Both the cry-wolf effect and omission errors increase when tasks are difficult. Studies on different workload conditions or task complexities (Biros et al., 2004; McBride et al., 2011) have reported an increased cry-wolf effect with high workload because operators could not evaluate the recommendations of the decision support system. Also, Huegli et al. (2020) found that operators showed lower compliance with automation (in their study EDSCB) when presented with difficult rather than easy targets. When targets were easy, operators could evaluate the alarms, and this reduced the cry-wolf effect. Turning to omission errors, previous studies have found an increase due to high task complexity (Biros et al., 2004; Lyell et al., 2018) or multitasking settings (Avril et al., 2021; Bailey & Scerbo, 2007; Parasuraman et al., 1993). Again, the reason might lie in the ability of the operators to evaluate the automated suggestion. For example, in a simulated military detection task, operators reported noticing automation errors better when task difficulty (in the form of image quality) was low rather than high (Spain, 2009). Therefore, when targets are easy for operators to detect, they should be able to evaluate whether decision support is missing a target and show less reliance when it fails to alarm a target. When targets are difficult to detect, operators show higher reliance and commit more omission errors when it fails to alarm a target.

1.3. Present study

The present study tested screeners supported by EDSCB in a simulated baggage X-ray screening task in which they had to detect prohibited articles (guns, knives, and IEDs) that were all built into PEDs. The EDSCB in this study had a hit rate of 75% and a false alarm rate of 9%. Considering our explosives target prevalence of 4.2%, the EDSCB made 90% correct decisions. This accuracy is viewed as reliable (Chavaillaz et al., 2018; Cullen et al., 2013; Ma & Kaber, 2007; Onnasch, 2015; Rice, 2009; Wiegmann et al., 2001). We used multiview X-ray imaging that is common at modern airports. We employed three realistic scenarios to simulate different concepts of operations at airport security checkpoints: In the baseline condition, PEDs were placed in plastic trays outside of bags. The trays were screened separately, and screeners were not supported by EDSCB—as is still the case at some airports today. In the second condition, PEDs were also placed in trays, but screeners were supported by EDSCB—as is the case at most airports. We refer to this test condition as PED-out; and, following Mendes et al. (2013), view it as a condition with low task difficulty. In the third condition, PEDs were left inside the cabin baggage and EDSCB supported screeners—representing a scenario at a modern airport. We refer to this test condition as PED-in, and, following Mendes et al. (2013), view it as a condition with high task difficulty. We chose one half of the PEDs to be notebooks because these are the most frequent PEDs in cabin baggage. Furthermore, we chose one half of the PEDs to be everyday electronic/electrical objects (e.g., conference speakerphones or water kettles) to investigate the practically relevant question of whether airports should differentiate between notebooks and everyday electronic/electrical objects when determining whether passengers can leave them in their baggage. Figure 1 shows four multiview X-ray images containing the same IED concealed in two different types of PED (a and b: notebook; c and d: everyday electronic/electrical object) and screened either separately (a and c: low task difficulty) or inside a bag (b and d: high task difficulty), all alarmed by EDSCB.

We anticipated that EDSCB would increase the detection of IEDs built into PEDs compared to the baseline group without EDSCB support. We further expected a lower detection performance for PED-in than for PED-out due to higher task difficulty. Also, we expected compliance with EDSCB to be lower for PED-in than for PED-out. When task difficulty is high, operators should have difficulties in identifying EDSCB hits while still experiencing many false alarms. Therefore, they will show less compliance with EDSCB, and this will lead to a more pronounced cry-wolf effect. When task difficulty is low, screeners should recognize the EDSCB hits and show more compliance. We further expected reliance to be higher for PED-in than for PED-out. This is because when EDSCB fails to detect targets with high task difficulty, screeners cannot recognize the miss and will still rely on its suggestion. As a result, screeners will make more omission errors. Because EDSCB detects only explosive targets, we did not expect EDSCB to affect the detection performance of nonexplosive targets (guns and knives built into PEDs). However, we expected a lower detection of guns and knives for PED-in than for PED-out, similar to the results reported by Mendes et al. (2013).

2. Method

2.1. Participants

Participants were 100 cabin baggage screeners from an international airport. All had been trained and certified according to national aviation security regulation standards. We excluded two participants because they were inadvertently given the same participant ID, making their data indistinguishable. The final sample consisted of 8 females and 90 males with a mean age of 28.43 years (range: 20–50 years). Fifty-one participants had professional experience of more than one year; 47, less than one year. Participants were tested during their regular working hours, informed about the study procedures and goals, and given written informed consent. The study complied with the American Psychological Association Code of Ethics and was approved by the institutional ethics review board of the School of Applied Psychology, University of Applied Sciences and Arts Northwestern Switzerland.

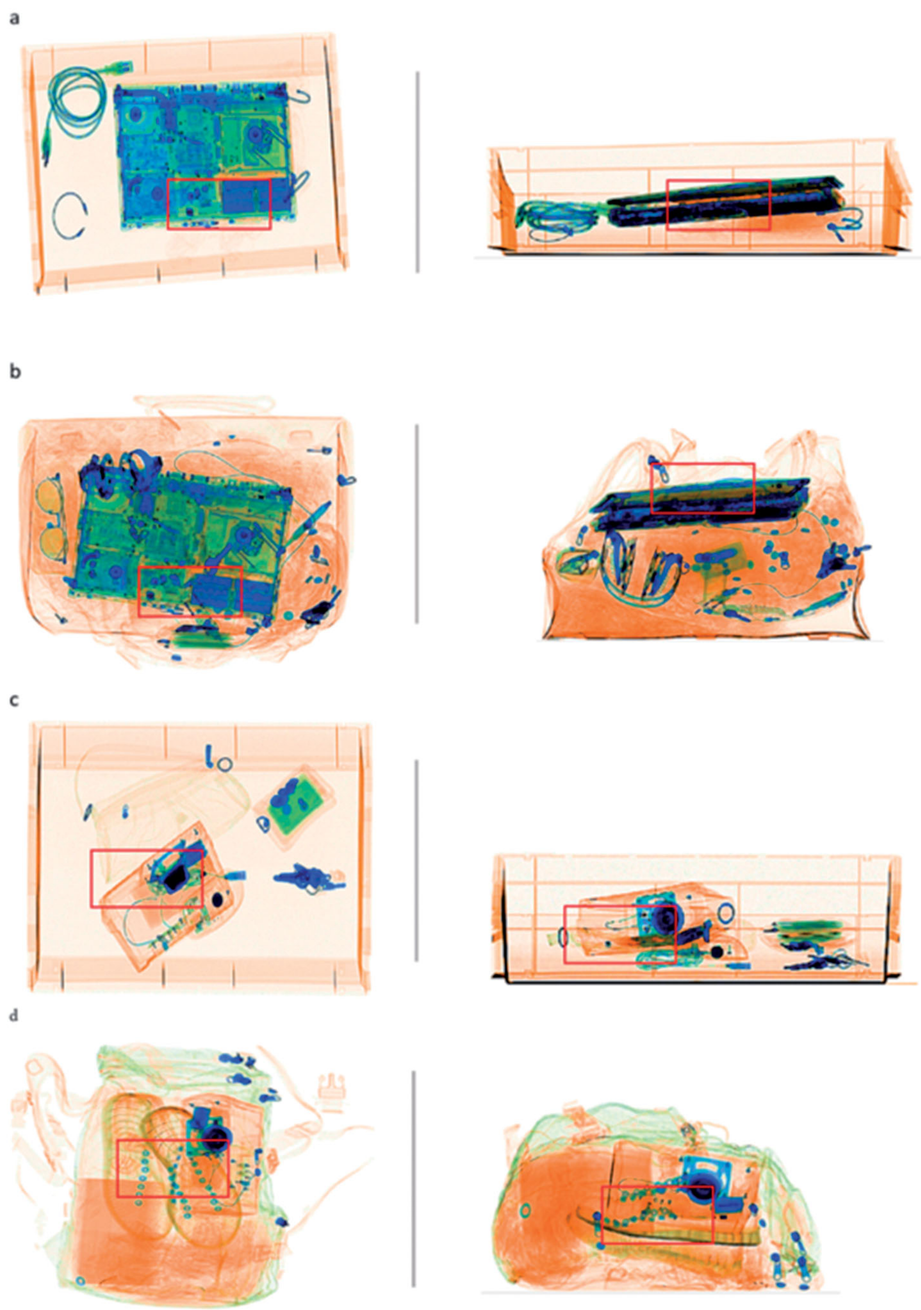

Figure 1. X-ray images from a multiview EDSCB machine used as a decision support system cueing potential targets (IEDs). The same bag is shown from two viewpoints differing by about 90°. It contains an IED concealed in two different types of PEDs (a and b: notebook; c and d: everyday electronic/electrical object) and screened either separately (a and c: low task difficulty) or inside a bag (b and d: high task difficulty), all alarmed by EDSCB.

2.2. Design

The study used a 3 × 3 × 2 mixed design with test conditions (baseline, PED-out, PED-in) as a between-subjects variable and prohibited articles category (IEDs, knives, and guns built into PEDs) and concealment type (everyday electronic/electrical object, notebook) as within-subjects variable. Participants were randomly assigned to one of the three test conditions. We chose guns built into PEDs as easy and knives built into PEDs as more difficult nonexplosive threats. This allowed us to investigate the placement effects of PEDs without the impact of EDSCB that affect human–machine system performance only for explosive threats (Huegli et al., 2020).

We examined the following dependent variables: (a) the human–machine system hit rate (proportion of target-present responses on target-present trials) and (b) human–machine system false alarm rate (proportion of target-present responses on target-absent trials) as measures of detection performance; (c) compliance (proportion of target-present responses of operators in trials in which the decision support system alarmed including both true and false alarms) and (d) reliance (proportion of target-absent responses on trials in which decision support system did not alarm including correct and incorrect refrain of an alarm) as measures of the operators' use of the decision support system; and (e) trust perception for participants who were allocated to one of the two test conditions working with EDSCB (PED-out and PED-in). Trust perception reflects the subjective trustworthiness of the system (Chavaillaz et al., 2019a). Finally, (f) we examined the response times on target-present and target-absent trials as measures of efficiency. We also assessed operator confidence for each trial. However, these confidence ratings were collected as part of a possible future investigation and were not analyzed for this study.

2.3. Materials

The simulated cabin baggage screening experiment contained 440 unique X-ray images of passenger bags and trays displayed with multiview imaging (56 practice and 384 test trials per test condition). Because most airports require passengers to remove PEDs from cabin bags, multiview images containing PEDs are rare. Therefore, three X-ray image experts recorded the images containing PEDs with an X-ray machine of the type HI-SCAN 6040 aTiX from Smiths Detection. Each target-present image had one prohibited article built into a PED: either a notebook or an everyday electronic or electrical object. Test trials included 48 target-present images, each containing one of 24 prohibited items (eight different simulated IEDs, guns, and knives). All prohibited items were presented twice: once built into a notebook and once into an everyday electronic or electrical object. For the conditions baseline and PED-out, PEDs with targets were recorded in 48 different trays. For the condition PED-in, PEDs with targets were recorded in 48 different bags.

Additionally, every bag or tray was recorded once containing the PED without the prohibited article and once without any PEDs at all. In total, we recorded 96 target-present and 192 target-absent images. More target-absent images were selected by the three X-ray image experts. They used bags from a pool of about 2,000 X-ray images recorded during regular airport security cabin baggage screening at a European airport. The recording and selection process resulted in 48 target-present and 336 target-absent images per test condition. To achieve enough target-present trials per participant and still be within a realistic range, the overall target prevalence was set at 12.5% and the target prevalence of IEDs at 4.2%. In the two automation conditions (PED-out and PED-in), red frames highlighted areas in the X-ray images that might contain IEDs (EDSCB alarms). No X-ray image had more than one red frame, and no X-ray image with a gun or knife had a red frame. The X-ray image experts set frames manually. The PED-out condition included 12 EDSCB hits (red frames on 12 of the 16 IEDs) and 34 EDSCB false alarms on trays containing PEDs. The PED-in condition had 12 EDSCB hits (red frames on 12 of the 16 IEDs) and 34 EDSCB false alarms on bags containing PEDs. This resulted in an automation hit rate of 75% and an automation false alarm rate of 9% for both automation conditions and an overall of 83% correct decisions. Appendix A presents an overview of the stimuli used for each test condition (see Tables A1–A3).

Trust perception was assessed with the 12-item questionnaire Checklist of Trust between People and Automation (Jian et al., 2000). Answers were given on a 7-point Likert scale ranging from 1 (*not at all*) to 7 (*agree totally*). An item example is "The system is reliable." Trust perception reflects the trustworthiness of the system (Chavaillaz et al., 2019a). We measured trust in automation only in participants allocated to the two test conditions working with EDSCB (PED-out and PED-in). Five participants who did not complete the questionnaire were excluded from analyses of trust perception, leaving a sample of 59 participants in PED-out and PED-in.

2.4. Procedure

Participants were tested in training facilities at the airport in groups of a maximum of six screeners in a quiet room and under supervision. They sat approximately 60 cm from HP EliteDesk 800 G1 computers with 23" EliteDisplay E232 monitors under normal lighting conditions. They responded on the screen with a mouse point and click. Participants first received oral and written instructions about the study, an explanation of the simulator's user interface, and information about the hit and false alarm rates of the EDSCB. They then performed 56 practice trials with trial feedback to familiarize themselves with the task and the reliability of EDSCB. Afterwards, participants conducted the main experiment in two test blocks, each containing 192 randomly presented trials without trial feedback. Participants had a maximum of 15 s to decide whether an X-ray image of cabin baggage contained a target or not by clicking on a "NOT OK" button (target-present) or an "OK" button (target-absent). Many airports apply such a 15-s time limit in cabin baggage screening. Participants then rated how confident they were about their decision on a scale ranging from 1 (*not confident*) to 5 (*very confident*). Then, the next trial started. European regulations (Commission Implementing Regulation [EU], 2015) mandate a break of at least 10 min after each period of a maximum of 20-min continuous X-ray image screening. Therefore, screeners took a 10-min break after each block. After the simulated X-ray screening task, participants in the EDSCB conditions (PED-out and PED-in) completed the Checklist of Trust between People and Automation (Jian et al., 2000). The complete session took approximately 60 min.

2.5. Statistical analyses

We conducted all analyses of variance (ANOVAs) and post hoc comparisons with R version 4.03 (R Core Team, 2020). Alpha was set at .05, and Holm–Bonferroni corrections were applied (Holm, 1979) for post hoc *t*-tests. We report effect sizes of ANOVAs using η_p^2 (partial eta-squared) with values of .01, .06, and .14 being interpreted as small, medium, and large effects respectively (Cohen, 1988, p. 368). The Huyn–Feldt correction was applied to p values of ANOVAs with more than two within-factor levels to address the violation of the sphericity assumption. Table 1 shows ANOVA

Table 1. *F* values, significance levels, and effect sizes for the ANOVA main effects of test condition, prohibited item category, and concealment type together with their interactions for dependent variables of human–machine system hit rate, false alarm rate, and response times.

Variable	TC			PAC			CT			TC × PAC			TC × CT			PAC × CT			TC × PAC × CT		
	F	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
Hit rate	2.47[a]	0.09	0.02	739.54[b]	<0.001	0.74	20.29[c]	<0.001	0.03	10.70[d]	<0.001	0.08	0.26[a]	0.77	0.00	32.45[b]	<0.001	0.06	2.04[d]	0.10	0.01
False alarm rate	6.87[a]	<0.01	0.08	–	–	–	6.45[b]	<0.01	0.02	–	–	–	2.26[d]	0.64	0.02	–	–	–	–	–	–
Response time																					
Target-present	2.61[a]	0.08	0.02	191.73[b]	<0.001	0.43	35.51[c]	<0.001	0.43	8.12[d]	<0.001	0.06	8.01[a]	<0.001	0.02	14.87[b]	<0.001	0.02	4.15[d]	<0.01	0.01
Target-absent	2.28[a]	0.11	0.04	–	–	–	65.19[b]	<0.001	0.09	–	–	–	27.81[d]	<0.001	.08	–	–	–	–	–	–

Note. TC: test condition; PAC: prohibited item category; CT: concealment type. [a]$df = 2, 95$; [b]$df = 2, 190$; [c]$df = 1, 95$; [d]$df = 4, 190$.

Table 2. *F* values, significance levels, and effect sizes for the ANOVA main effects of test condition and concealment type together with their interactions for operator compliance and reliance.

Variable	TC			CT			TC × CT		
	F	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
Compliance overall	4.56[a]	<0.05	0.06	0.28[a]	0.60	0.00	0.02[a]	0.89	0.00
Compliance target present	4.95[a]	<0.05	0.05	5.88[a]	<0.05	0.03	1.75[a]	0.19	0.01
Compliance target absent	2.38[a]	0.12	0.36	1.34[a]	0.25	0.03	0.00[a]	0.95	0.00
Reliance overall	15.69[b]	<0.001	0.13	0.35[b]	0.71	0.00	2.32[b]	0.10	0.02
Reliance target present	24.78[b]	<0.001	0.18	35.26[b]	<0.001	0.20	0.96[b]	0.33	0.01
Reliance target absent	14.97[b]	<0.001	0.12	1.54[b]	0.22	0.01	2.57[b]	0.08	0.02

Note. TC: test condition; CT: concealment type. [a]$df = 1, 62$; [b]$df = 2, 12$.

statistics for the human–machine system hit rate, false alarm rate, and response times. Table 2 shows the ANOVAs with overall compliance and compliance on target-present and -absent images respectively. For calculations of means, we computed basic bootstrapped 95% confidence intervals (1,000 iterations).

3. Results

3.1. Detection performance

3.1.1. Human–machine system hit rate

We computed a 3 (test condition: baseline, PED-out, PED-in) × 3 (prohibited article category: IEDs, knives, guns built into PEDs) × 2 (concealment type: everyday electronic/electrical object, notebook) mixed-design ANOVA with human–machine system hit rate as the dependent variable (Table 1). Figure 2 displays the human–machine system hit rate for the statistically significant interaction effects: (a) prohibited article categories by test condition and (b) prohibited article categories by concealment type. The horizontal green line indicates the hit rate of EDSCB by itself (automation hit rate).

The ANOVA (Table 1) and Figure 2a show that EDSCB increased the detection of IEDs when PEDs were screened separately (PED-out > baseline). However, detection of IEDs decreased again when PEDs were left inside baggage (PED-in < PED-out). Post hoc *t*-tests confirmed the significance of these differences. Moreover, the human–machine system hit rate for detecting IEDs was lower than that for EDSCB on its own, especially when PEDs were left inside baggage (PED-in), representing a strong cry-wolf effect. We shall return to the cry-wolf effect when discussing operator compliance. Moreover, screeners detected guns better than knives and IEDs, and knives better than IEDs. Knives were detected best when PEDs were placed inside luggage (PED-in). Figure 2b shows that the type of concealment had no

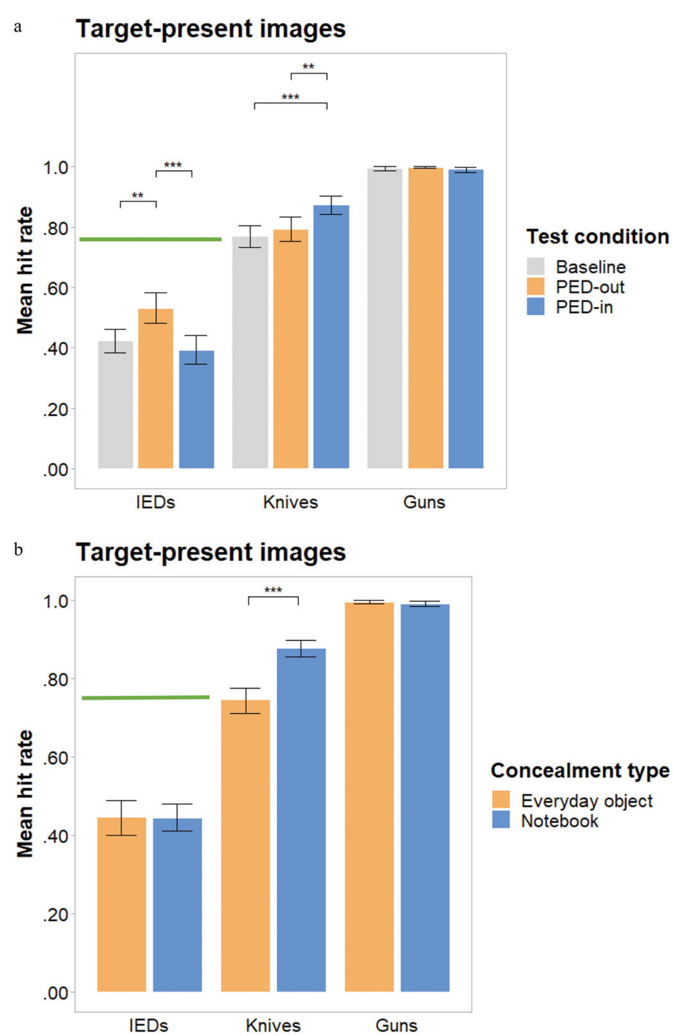

Figure 2. Mean human–machine system hit rate by test condition and prohibited article category (a) and by concealment type and prohibited article category (b). Post hoc pairwise *t* tests were computed to compare the means of test condition and concealment type within the prohibited article category. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Error bars show bootstrapped 95% confidence intervals (1,000 iterations). The horizontal green line indicates the hit rate of EDSCB by itself.

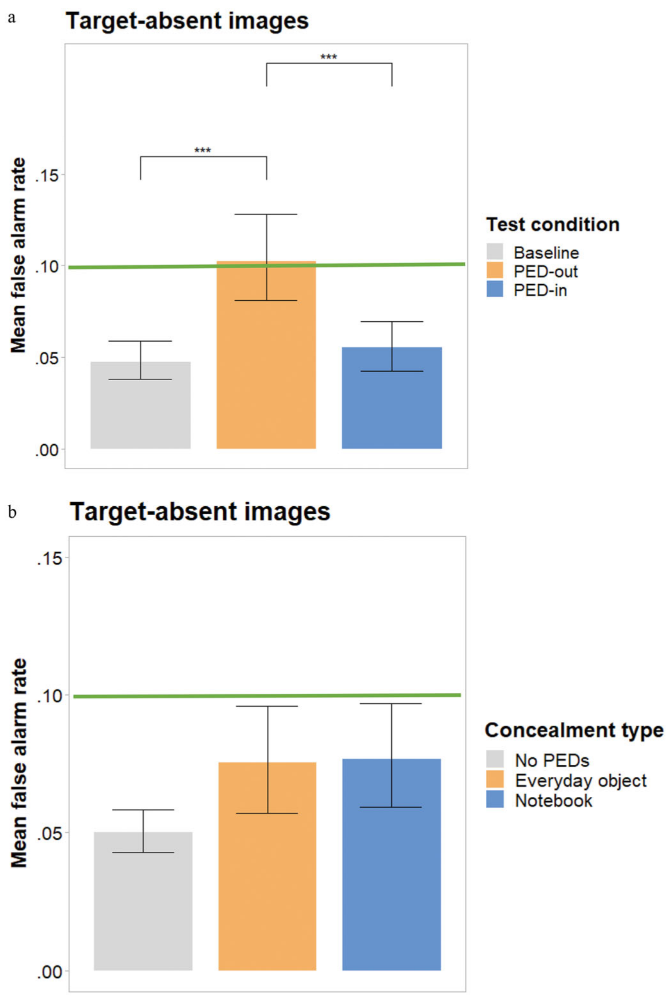

Figure 3. Mean human–machine system false alarm rate by test condition (a) and concealment type (b). Post hoc pairwise t tests were computed to compare the means of test condition and concealment type within the prohibited article category. $^{***}p < 0.001$. Error bars show bootstrapped 95% confidence intervals (1,000 iterations). The horizontal green line indicates the false alarm rate of EDSCB by itself.

impact on detection performance except for that on knives in which screeners achieved a higher detection performance when knives were concealed in notebooks compared to everyday electronic/electrical objects. Again, post hoc *t*-tests confirmed the significance of these differences. We shall explain this result in the discussion section.

3.1.2. Human–machine system false alarm rate

For target-absent trials, concealment type had three levels (no PEDs, everyday electronic/electrical object, notebook), because screeners could also report human–machine system false alarms on images that did not contain PEDs. Moreover, for target-absent trials, we dropped the variable prohibited articles category. Therefore, we computed a 3 (test condition: baseline, PED-out, PED-in) × 3 (concealment type: no PED, everyday electronic/electrical object, notebook) mixed-design ANOVA with human–machine system false alarm rate as the dependent variable (Table 1).

Figure 3 displays the human–machine system false alarm rate for the statistically significant main effects: (a) test condition and (b) concealment type. The horizontal green line indicates the false alarm rate of EDSCB by itself (automation false alarm rate). The ANOVA (Table 1) and Figure 3a show that EDSCB nearly doubled the human–machine system false alarm rate when PEDs were placed outside baggage (PED out) compared to the baseline condition. However, when EDSCB was in use and PEDs were placed inside bags (PED-in), screeners cleared more false alarms and achieved a lower false alarm rate than when PEDs were placed outside of baggage (PED-out). Post hoc *t*-tests confirmed the significance of these differences. Figure 3b and post hoc comparisons showed no significant pairwise differences between different concealment types, suggesting that the human–machine system false alarm rate does not depend on what kind of PED an image contains. Overall, the human–machine system false alarm rate did not exceed the EDSCB false alarm rate.

3.2. Behavioral trust measures and trust perception

3.2.1. Compliance

For compliance, we included only the two test conditions in which EDSCB was applied and only images in which alarms were present. We computed a 2 (test condition: PED-out, PED-in) × 2 (concealment type: everyday electronic/electrical object, notebook) mixed-design ANOVA with compliance as the dependent variable (see Table 2 for the ANOVAs with overall compliance and compliance on target-present and -absent images respectively).

Figure 4a displays overall operator compliance for the statistically significant main effect of test conditions (PED-out, PED-in). The ANOVA in Table 2 and Figure 4a shows that screeners complied more with EDSCB when PEDs were placed outside (i.e., task difficulty was low) compared to inside bags (i.e., task difficulty was high).

We compared the compliance of the two test conditions with EDSCB hits (target-present images) and EDSCB false alarms (target-absent images) to investigate whether higher compliance results from active processing of alarms due to the operators' better detection of IEDs when presented with easier images, and, hence, their use of the information provided by the EDSCB as decision support. Regarding target-present images, the ANOVA in Table 2 confirms a main effect for test condition with higher operator compliance when PEDs were placed outside ($M = 0.58$, 95% CI [0.52, 0.64]) compared to inside the bag ($M = 0.46$, 95% CI [0.41, 0.51]), $p < 0.01$. This means that screeners ignored 42% (PED-out) and 54% (PED-in) of true alarms leading to the cry-wolf effect observed in Section 3.1.1. Regarding target-absent images, the ANOVA did not reveal any significant difference for mean operator compliance between when PEDs were placed outside ($M = 0.13$, 95% CI [0.08, 0.16]) and inside the bags ($M = 0.08$, 95% CI [05, 0.10]), *ns*.

3.2.2. Reliance

We computed a 2 (test condition: PED-out, PED-in) × 2 (concealment type: everyday electronic/electrical object,

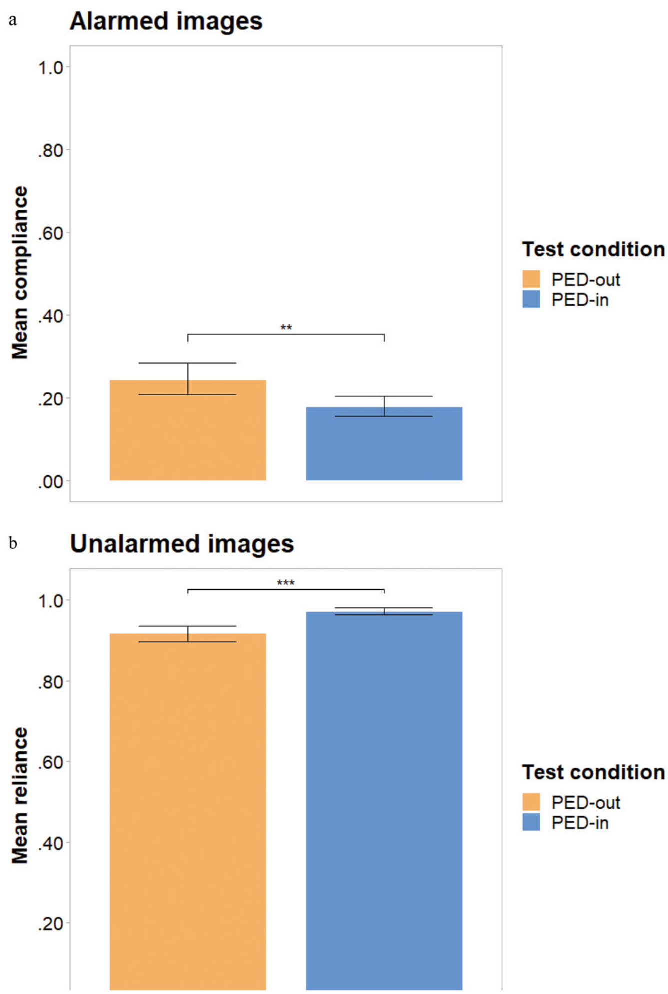

Figure 4. Operator compliance (a) and reliance (b) by test condition. Post hoc pairwise t tests were computed to compare the means of test condition and concealment type within the prohibited article category. $^{**}p < 0.01$, $^{***}p < 0.001$. Error bars show bootstrapped 95% confidence intervals (1,000 iterations).

notebook) mixed-design ANOVA with reliance as the dependent variable. We included only the two test conditions with EDSCB in use, and, in turn, only images without EDSCB alarms (see Table 2 for the ANOVAs with overall reliance and reliance on target-present and -absent images, respectively).

Figure 4b displays overall operator reliance for the significant main effect of test condition (PED-out, PED-in). The ANOVA (Table 2) and Figure 4b show that screeners relied less on EDSCB when PEDs were placed outside (i.e., when task difficulty was low) compared to inside the bag (i.e., when task difficulty was high). Post hoc t-tests confirmed the significance of these differences.

We compared the reliance of the screeners in the two test conditions in terms of EDSCB misses (target-present images) and EDSCB correct rejections (target-absent images) to investigate whether the higher reliance on more difficult images resulted in screeners committing more omission errors. Regarding target-present images, the ANOVA in Table 2 confirmed a significant main effect for test condition, with lower operator reliance when PEDs were placed outside ($M = 0.60$, 95% CI [0.54, 0.66]) compared to inside the bag ($M = 0.82$, 95% CI [0.76, 0.89]), $p < 0.001$. This means that screeners in the PED-out condition committed omission errors on 60% of the unalarmed IEDs. Screeners in the PED-in condition committed omission errors on 82% of the unalarmed IEDs. Also, reliance was higher when IEDs were built into notebooks ($M = 0.83$, 95% CI [0.77, 0.89]) rather than everyday electronic/electrical objects ($M = 0.60$, 95% CI [0.54, 0.67]), $p < 0.001$.

For target-absent images, the ANOVA in Table 2 revealed a main effect for test conditions with operators showing less reliance on EDSCB when notebooks were placed outside ($M = 0.91$, 95% CI [0.89, 0.93]) rather than inside bags ($M = 0.97$, 95% CI [0.96, 0.98]), $p < 0.001$.

3.2.3. Trust perception

The relationship between behavioral trust measures—compliance and reliance—and subjective perception of the actual trustworthiness of the automation has been found to be vague (Chavaillaz et al., 2019b). In the present study, Pearson correlations did not indicate that trust perception related significantly to compliance with EDSCB for all images, $r(57) = -0.11$, $p = 0.31$; target-present images, $r(57) = -0.22$, $p = 0.09$; or target-absent images, $r(57) = -0.03$, $p = 0.80$. Also, there were no significant correlations of trust with reliance on EDSCB for all images, $r(57) = 0.14$, $p = 0.21$; target-present images, $r(57) = 0.20$, $p = 0.09$; or target-absent images, $r(57) = 0.13$, $p = 0.24$.

3.3. Response times

3.3.1. Target-present images

Finally, we computed a 3 (test condition: baseline, PED-out, PED-in) × 3 (prohibited article category: IEDs, knives, guns built into PEDs) × 2 (concealment type: everyday electronic/electrical object, notebook) mixed-design ANOVA with target-present response time (mean of medians) as the dependent variable (Table 1).

Figure 5 displays target-present response times (mean of medians) for the statistically significant three-way interaction effect: prohibited article categories by test condition and concealment type. The ANOVA (Table 1) and Figure 5 show that EDSCB did not affect response times when PEDs were placed outside bags (baseline vs. PED-out). However, target-present response time depended on task difficulty. When EDSCB supported screeners, they detected IEDs more slowly when PEDs were placed inside rather than outside bags. However, this effect was found only for the concealment type of notebooks. Post hoc t-tests confirmed the significance of these differences. In general, screeners detected guns faster than knives and guns and knives faster than IEDs.

3.3.2. Target-absent images

Analogous to the calculations for the human–machine system false alarm rate, we further computed a 3 (test condition: baseline, PED-out, PED-in) × 3 (concealment type: no PED, everyday electronic/electrical object, notebook) mixed-

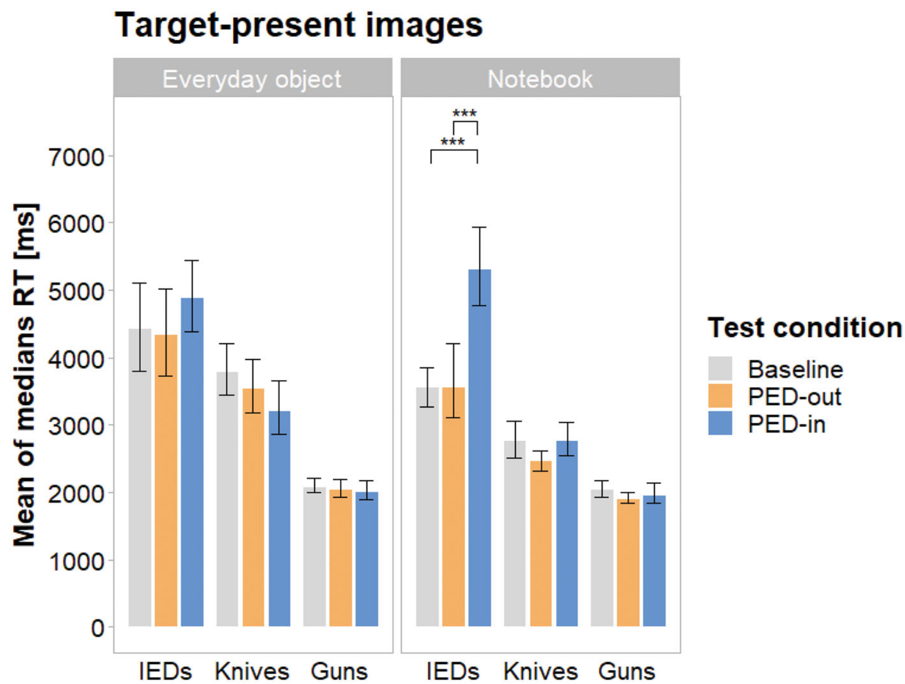

Figure 5. Target-present response times (RT) by test condition, prohibited article category, and concealment type. Post hoc pairwise *t* tests were computed to compare the means of test condition and concealment type within the prohibited article category. $^{***}p < 0.001$. Error bars show bootstrapped 95% confidence intervals (1,000 iterations).

design ANOVA with target-absent response times (mean of medians) as the dependent variable (Table 1). Figure 6 displays the target-absent response times for the statistically significant main effects: (a) test condition and (b) concealment type. The ANOVA in Table 1 and Figure 6 indicates that EDSCB did not slow down response times when PEDs were separated from the baggage. However, when EDSCB was used and PEDs were placed inside bags, images were more difficult and screeners' response times were slower. Figure 6b shows that this applied to both concealment types (everyday electronic/electrical objects and notebooks). Significant post hoc *t*-tests confirmed both differences.

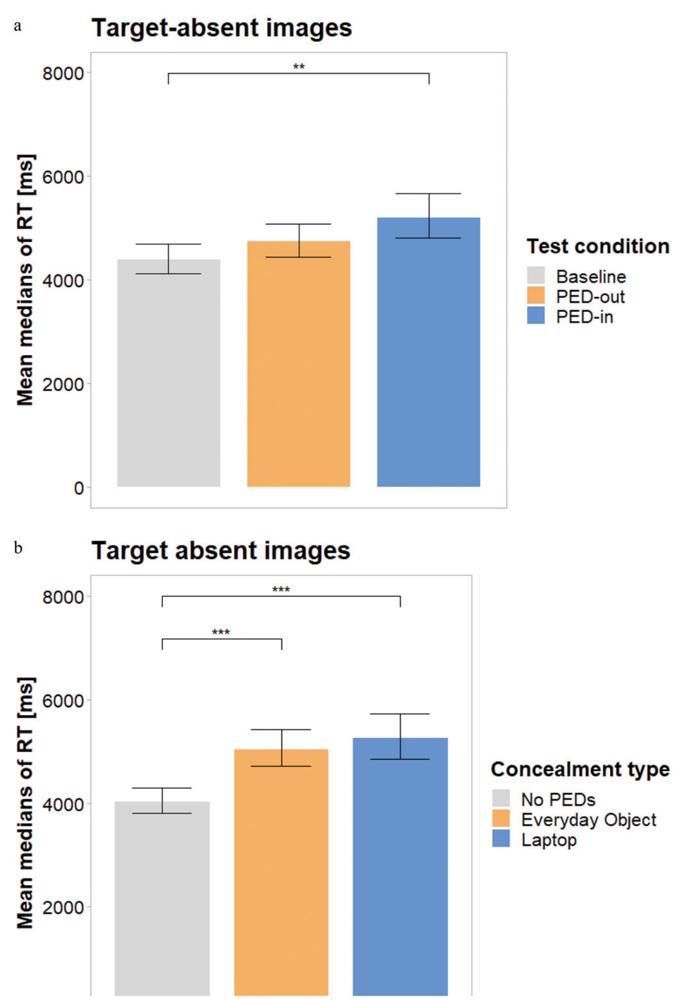

Figure 6. Target-absent response times (RT) by test condition (a) and concealment type (b). Post hoc pairwise *t* tests were computed to compare the means of test condition and concealment type within the prohibited article category. $^{**}p < 0.01$, $^{***}p < 0.001$. Error bars show bootstrapped 95% confidence intervals (1,000 iterations).

4. Discussion

The goal of the present study was to give recommendations about whether PEDs can be left in cabin baggage and how EDSCB alarms should be resolved. There are four approaches to deal with PEDs at airport security checkpoints in the presence of EDSCB: (a) Passengers have to remove PEDs from their cabin baggage and screeners use EDSCB as a decision support and resolve alarms on-screen while inspecting the X-ray image. (b) Passengers can leave PEDs in their baggage, and screeners resolve EDSCB alarms on-screen while inspecting the X-ray image. (c) Passengers have to remove PEDs from their baggage, and alarmed bags are sent directly to secondary search. (d) Passengers can leave PEDs in their baggage, and alarmed bags are sent directly to secondary manual search. We used this example to investigate the benefits of automation as a decision support under varying task difficulty. We tested whether a 90% reliable EDSCB helps professional X-ray baggage screeners to detect IEDs built into PEDs when placed inside bags (high task difficulty) compared to when PEDs were screened separately (low task difficulty). Results show that EDSCB increased the detection of IEDs built into PEDs when screened separately (i.e., low task difficulty) and screeners conducted on-screen alarm resolution. Nonetheless, detection performance was still lower than EDSCB on its own. When electronics were left inside the baggage, operators ignored many EDSCB alarms, and many IEDs were missed. Moreover, screeners missed most unalarmed explosives because they over-relied on the EDSCB's judgment. Therefore, on-screen alarm resolution, that is approaches (a) and (b), are not recommended. Instead, when the EDSCB indicates that the bag might contain an explosive, baggage should always be examined further in a secondary search using explosive trace detection, manual opening of bags and other means (approaches c and d). Our results further demonstrate that benefits of decision support systems and operator compliance with and reliance on EDSCB all depend on task difficulty. Compliance with EDSCB decreased with higher task difficulty leading to a more pronounced cry-wolf effect. Reliance on EDSCB increased with higher task difficulty leading to more omission errors. In the remaining discussion, we first look at our results on human–machine system detection performance, operator compliance, reliance, and response times in detail. After discussing the limitations of the present study and further research needs, we examine its practical implications.

4.1. Detection performance

The benefits of automation as decision support depend on the task difficulty. Our results show that EDSCB increases the detection of IEDs built into PEDs, but only when screened separately—that is, when task difficulty is low (PED-out). In both conditions in which screeners were supported by EDSCB (PED-out and PED-in), detection of IEDs by the human–machine system was lower than detection by EDSCB on its own (see the green line in Figure 2a,b). This phenomenon—referred to as the cry-wolf effect—has been investigated widely for automation and decision support systems with different reliabilities (Alberdi et al., 2004; Bartlett & McCarley, 2017, 2019; Huegli et al., 2020; Manzey et al., 2014; Meyer, 2001; Meyer et al., 2014). For PED-out, the human–machine system hit rate for IEDs was about 25% lower than the hit rate of the EDSCB. For PED-in, the human–machine system hit rate for IEDs was even 35% lower than that of the EDSCB. This result provides strong evidence that task difficulty influences not only the human–machine system hit rate but also the operator's usage of automation. Other than for knives, concealment type did not affect the human–machine system hit rate of IEDs. This result is particularly interesting from a practical point of view, as will be discussed in Section 4.4. The human–machine system hit rate for IEDs was lower than that in previous studies with screeners supported by EDSCB (Hättenschwiler et al., 2018; Huegli et al., 2020). Because IEDs were built into PEDs, the components of IEDs—which can be learned—were concealed by the PED, and IEDs became very difficult for screeners to detect, as reported in previous studies without EDSCB (Mendes et al., 2012, 2013).

Overall, guns and knives in PEDs were detected well in our study, and EDSCB did not affect their detection. This is good news, suggesting that EDSCB does not impair operators' attention when searching for nonexplosive prohibited articles. Detection rates for knives and IEDs were better than those in previous studies with well-trained screeners without automation support (Halbherr et al., 2013; Koller et al., 2008, 2009) or in a recent study that tested screeners supported with different multiview EDSCB (Huegli et al., 2020). This result might have emerged because the detection of prohibited articles in X-ray images depends on image-based factors such as the viewpoint on the prohibited article (Bolfing et al., 2008; Schwaninger et al., 2005). The options for hiding a knife in a notebook are limited to the battery pocket, restricting variation of the viewpoint on knives in X-ray images and making them easier to detect. To our surprise, detection of knives was also slightly better when PEDs appeared in bags. However, this effect was very small and did not occur for guns that were detected very well by the human–machine system in all conditions.

In terms of checkpoint efficiency, false alarm rates by airport security screeners need to be minimal, because further examination of bags at the checkpoint involves more time-consuming explosives trace detection and the manual opening of baggage (Sterchi & Schwaninger, 2015). Our results show that the increase in detection of IEDs comes at a cost. Particularly, EDSCB as a decision support system also increased the human–machine system false alarm rate compared to the baseline group, but only when PEDs were screened separately (PED-out). When PEDs were left inside bags, the human–machine system false alarm rate was reduced by half. Together with our results on the human–machine system hit rate, this suggests that screeners were more inclined to classify bags as suspicious when inspecting images with low rather than high task difficulty. This was due to the higher compliance with and lower reliance on EDSCB alarms when inspecting images with low compared to high task difficulty. We shall continue the discussion on compliance and reliance in Section 4.2. Overall, the human–machine system false alarm rate did not exceed the EDSCB false alarm rate. This result is good news because screeners could keep the human–machine system false alarm rate within a practically feasible range (Sterchi & Schwaninger, 2015).

4.2. Behavioral trust and trust perception

4.2.1. Compliance

Operator compliance with EDSCB also depended on task difficulty. Screeners complied less with EDSCB when task difficulty was high (PED-in) compared to low (PED-out), leading to a more substantial cry-wolf effect when inspecting images under high task difficulty. This finding is consistent with two previous studies using varying task complexities and workload (Biros et al., 2004; McBride et al., 2011). Both reported that the cry-wolf effect increases when operators have difficulties in evaluating the automation's recommendations because of higher task demands. According to Mosier and Manzey (2020), compliance describes the behavioral tendency to trust and follow the recommendation of the decision support system. To calibrate their behavior and show an accurate amount of compliance with the decision support system, operators must recognize true alarms (Spain, 2009). Screeners in our study did not show an accurate amount of compliance. They ignored 54% of the true EDSCB alarms when task difficulty was high (PED-in) and 42% when task difficulty was low (PED-out) on target-present images. We argue that when task difficulty is high, operators have difficulties in recognizing true EDSCB alarms while still experiencing many false alarms, and this leads to a cry-wolf effect (Breznitz, 1983). When task difficulty is low, operators make better use of the information provided by EDSCB to verify their judgment. In our study, however, both groups ignored many EDSCB alarms and showed insufficient human–machine system performance.

Screeners in both EDSCB conditions recognized most EDSCB false alarms and ignored a large number of them (87% in PED-out and 92% in PED-in). This result speaks against complacency (Parasuraman et al., 1993) that would result in a blind acceptance of alarms due to a lack of attention, poor monitoring, and vigilance issues (Alberdi et al., 2004, 2008; Meyer et al., 2014; Onnasch et al., 2014; Rice & McCarley, 2011). Although the screeners in our study ignored approximately one-half of the true EDSCB alarms, our results show that they complied more with correct EDSCB alarms than with EDSCB false alarms. Therefore, we shall not attribute our results to an extreme but rather to a moderate cry-wolf effect.

4.2.2. Reliance

We also found that operator reliance on EDSCB depended on task difficulty. Screeners showed more reliance on EDSCB when task difficulty was high (PED-in) than low (PED-out). Most importantly, when screeners were inspecting target-present images, the high reliance on EDSCB led to human–machine system misses of IEDs. In the current study, screeners showed 82% reliance on EDSCB for high task difficulty and 60% reliance for low—albeit these images contained IEDs. Especially for the high task difficulty condition, this overreliance resulted in a considerable number of omission errors; a finding that is consistent with several studies reporting an increase in omission errors with increasing task complexity (Biros et al., 2004; Lyell et al., 2018) or multitasking settings (Avril et al., 2021; Bailey & Scerbo, 2007; Parasuraman et al., 1993). Regarding target-absent images, screeners also showed less reliance on EDSCB when task difficulty was low compared to high. This behavior caused more human–machine system false alarms when task difficulty was low. Nonetheless, reliance was generally higher for target-absent than for target-present images, which also speaks for active task processing and against complacency (Alberdi et al., 2004, 2008; Meyer et al., 2014; Onnasch et al., 2014; Rice & McCarley, 2011). The high reliance on the screeners in our study seems rational, because in most images without an EDSCB alarm, there was indeed no IED in the image. Screeners were right to depend on EDSCB's recommendation as a heuristic replacement for vigilant information seeking and processing (Parasuraman & Manzey, 2010), especially when task difficulty was high, and screeners had difficulties in evaluating the EDSCB suggestion. Unfortunately, this led to omission errors when EDSCB failed—once again demonstrating the "ironies of automation" (Bainbridge, 1983).

4.2.3. Trust perception

Our analyses revealed no evidence that the two behavioral trust measures—compliance and reliance (Meyer, 2004; Meyer et al., 2014)—related to the subjective perception of trustworthiness. However, besides trust, other factors such as recognizing decision support system errors influence the effective use of a system (Spain, 2009). We argue that screeners had difficulties in processing the task and recognizing decision support system errors. Therefore, the degree of compliance and reliance may represent a more heuristic use of the decision support system (Mosier & Manzey, 2020) and not the actual degree of operators' trust in it. Examples of a heuristic use of the decision support system would be the probability matching strategy (Bartlett & McCarley, 2017, 2019; Bliss et al., 1995; Manzey et al., 2014) or the much-discussed automation bias (Lyell & Coiera, 2017; Mosier & Manzey, 2020), both of which have been found to cause insufficient human–machine system performance (Bartlett & McCarley, 2017; Boskemper et al., 2021; Huegli et al., 2020; Manzey et al., 2014; Mosier et al., 1998).

4.2.4. Response times

A previous study found that PEDs inside bags slow down screeners' response times (Mendes et al., 2013), and longer response times per visually inspected X-ray image of a passenger bag result in slower passenger throughput (Hättenschwiler et al., 2018; Sterchi & Schwaninger, 2015). However, it was unclear whether PEDs inside passenger bags also increase response times when EDSCB is in use. Our results show that screeners' target-present response times were affected by task difficulty and the concealment type when detecting IEDs. Although EDSCB did not increase response times) when task difficulty was low (PED-out), the screeners detected IEDs in PEDs more slowly compared to when task difficulty was high (PED-in). This was, however, observed only for IEDs built into notebooks. For guns and knives, task difficulty did not affect response times. Target-absent response times are more relevant to the efficiency of the human–machine system than target-present response times because most bags in baggage screening do not contain prohibited items (in both our experiment and airport X-ray screening). From a practical point of view, this is good news because, when EDSCB is in use, PEDs inside bags do not increase response times and, therefore, do not negatively affect the baggage flow. The analyses of target-absent response times show that EDSCB alone had no effect on response times and that in the conditions with EDSCB (PED-in and PED-out), task difficulty had no impact on screener response times. However, the target-absent response times analyses show that screeners' response times were slightly slower when images contained PEDs, regardless of whether they were placed in or outside the bag.

We found no evidence of a trade-off between speed and accuracy—that is, better detection at the expense of a slower response time (Heitz, 2014). On the contrary, guns in PEDs were detected better and faster than all other categories of prohibited items, probably due to the top-down guidance (Wolfe, 2021) of their familiar, specific shape and their high material density (Halbherr et al., 2013; Koller et al., 2009). Similarly, screeners detected knives in PEDs better and faster than IEDs. One possible explanation is that features of IEDs were masked by the PEDs, making IED detection more difficult (Mendes et al., 2013). At the same time, the detection of knives was facilitated by being concealed in notebooks and presented from a simple viewpoint. A speed–accuracy trade-off also failed to explain differences in human–machine system hit rates between the three test conditions because higher hit rates are not associated with slower response times.

4.3. Limitations

This study has some limitations that should be addressed in future research. First, it was conducted at one international airport with screeners trained and certified according to the national standards of that particular country. Future studies should investigate whether the effects found in this study also apply to other countries. Second, as explained in Section 1.1, in X-ray screening at airports, the extremely low target prevalence of prohibited articles is increased artificially to about 2–4% using threat image projection and covert tests (Hofer & Schwaninger, 2005; Meuter & Lacherez, 2016; Schwaninger, 2009; Skorupski & Uchroński, 2016; Wetter et al., 2008). Although target prevalence in our study was lower than in previous studies on decision support systems (Bartlett & McCarley, 2019; Boskemper et al., 2021; Chavaillaz et al., 2018, 2019b; Rice & McCarley, 2011; Wiegmann et al., 2001), it was still higher than in real-world scenarios. Third, the quality of EDSCB is still improving. For example, EDSCB based on computer tomography, which is becoming more common at airports, could soon achieve EDSCB false alarm rates below 5% (personal communication with airports and manufacturers, February 2020). This technical progress needs to be considered when discussing the practical implications of our study.

4.4. Practical implications

We have shown that EDSCB used as a decision support system increases the detection of IEDs in PEDs when they are screened separately. Further, the type of PED has almost no impact on detecting IEDs, which is particularly interesting. Nowadays, passengers often keep everyday electronic/electrical objects (e.g., speakerphones or water kettles) inside their cabin baggage. Our results question the need to screen notebooks separately. Leaving PEDs inside baggage leads to faster divesting times, fewer bags per person, and thus increased passenger throughput at security checkpoints. Unfortunately, in our study, human–machine system detection of IEDs concealed in PEDs was still insufficient and lower than the EDSCB on its own, regardless of whether PEDs were screened separately or left inside bags.

Therefore, we recommend that passengers can leave PEDs in their luggage if EDSCB of Standard C2 is used, but screeners should be given clear instructions to send all bags with EDSCB alarms to secondary search (approach d) instead of using EDSCB as a decision support system and conducting on-screen alarm resolution (approaches a and b). Sending all alarmed bags with PEDs to secondary search ensures a high detection performance of the human–machine system because EDSCB exceeds the operators' detection capabilities. Nonetheless, it is important for screeners to still inspect all X-ray images carefully for IEDs not only to prevent them from committing omission errors but also because current EDSCB still miss certain IEDs that can be detected only by visual inspection (Howell, 2017).

5. Conclusion

The present study conducted with professional airport security screeners and highly realistic stimuli investigated the benefits of EDSCB as a decision support in relation to task difficulty. Our results confirm that EDSCB increase detection of IEDs built into electronic devices when they are screened separately—that is, when task difficulty is low. Unfortunately, especially when presented with difficult images, operators ignore many EDSCB alarms due to low compliance, and this leads to a cry-wolf effect and insufficient human–machine system performance. Moreover, high task difficulty also results in operators being overly reliant on EDSCB, leading to many omission errors. We recommend that when passengers leave PEDs inside their bags and EDSCB indicates that the bag might contain an explosive, the baggage should be examined further in a secondary search.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID

David Huegli http://orcid.org/0000-0002-7176-7765

References

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2015). Targets need their own personal space: Effects of clutter on multiple-target search accuracy. *Perception*, *44*(10), 1203–1214. https://doi.org/10.1177/0301006615594921

Alberdi, E., Povykalo, A., Strigini, L., & Ayton, P. (2004). Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. *Academic Radiology*, *11*(8), 909–918. https://doi.org/10.1016/j.acra.2004.05.012

Alberdi, E., Povyakalo, A. A., Strigini, L., Ayton, P., & Given-Wilson, R. (2008). CAD in mammography: Lesion-level versus case-level analysis of the effects of prompts on human decisions. *International Journal of Computer Assisted Radiology and Surgery*, *3*(1–2), 115–122. https://doi.org/10.1007/s11548-008-0213-x

Avril, E., Cegarra, J., Wioland, L., & Navarro, J. (2022). Automation type and reliability impact on visual automation monitoring and human performance. *International Journal of Human–Computer Interaction*, *38*(1), 64–77. https://doi.org/10.1080/10447318.2021.1925435

Avril, E., Valéry, B., Navarro, J., Wioland, L., & Cegarra, J. (2021). Effect of imperfect information and action automation on attentional allocation. *International Journal of Human–Computer Interaction*, *37*(11), 1063–1073. https://doi.org/10.1080/10447318.2020.1870817

Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, *8*(4), 321–348. https://doi.org/10.1080/14639220500535301

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. https://linkinghub.elsevier.com/retrieve/pii/B9780080293486500269 https://doi.org/10.1016/0005-1098(83)90046-8

Bartlett, M. L., & McCarley, J. S. (2017). Benchmarking aided decision making in a signal detection task. *Human Factors*, *59*(6), 881–900. https://doi.org/10.1177/0018720817700258

Bartlett, M. L., & McCarley, J. S. (2019). No effect of cue format on automation dependence in an aided signal detection task. *Human Factors*, *61*(2), 169–190. https://doi.org/10.1177/0018720818802961

Biros, D. P., Daly, M., & Gunsch, G. (2004). The influence of task load and automation trust on deception detection. *Group Decision and Negotiation*, *13*(2), 173–189. https://doi.org/10.1023/B:GRUP.0000021840.85686.57

Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics*, *38*(11), 2300–2312. https://doi.org/10.1080/00140139508925269

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. In A. Holzinger (Ed.), *HCI and usability for education and work, lecture notes in computer science* (Vol. 5298, pp. 419–438). Springer.

Bolton, G. E., & Katok, E. (2018). Cry Wolf or equivocate? Credible forecast guidance in a cost-loss game. *Management Science*, *64*(3), 1440–1457. https://doi.org/10.1287/mnsc.2016.2645

Boskemper, M. M., Bartlett, M. L., & McCarley, J. S. (2021). Measuring the efficiency of automation-aided performance in a simulated baggage screening task. *Human Factors*. https://doi.org/10.1177/0018720820983632

Breznitz, S. (1983). *Cry-wolf: The psychology of false alarm*. Lawrence Erlbaum.

Chancey, E. T., Bliss, J. P., Yamani, Y., & Handley, H. A. H. (2017). Trust and the compliance-reliance paradigm: The effects of risk, error bias, and reliability on trust and dependence. *Human Factors*, *59*(3), 333–345. https://doi.org/10.1177/0018720816682648

Chavaillaz, A., Schwaninger, A., Michel, S., & Sauer, J. (2018). Automation in visual inspection tasks: X-ray luggage screening supported by a system of direct, indirect or adaptable cueing with low and high system reliability. *Ergonomics*, *61*(10), 1395–1408. https://doi.org/10.1080/00140139.2018.1481231

Chavaillaz, A., Schwaninger, A., Michel, S., & Sauer, J. (2019a). Work design for airport security officers: Effects of rest break schedules and adaptable automation. *Applied Ergonomics*, *79*, 66–75. https://doi.org/10.1016/j.apergo.2019.04.004

Chavaillaz, A., Schwaninger, A., Michel, S., & Sauer, J. (2019b). Expertise, automation and trust in X-ray screening of cabin baggage. *Frontiers in Psychology*, *10*, 256–211. https://doi.org/10.3389/fpsyg.2019.00256

Chavaillaz, A., Wastell, D., & Sauer, J. (2016). System reliability, performance and trust in adaptable automation. *Applied Ergonomics*, *52*, 333–342. https://doi.org/10.1016/j.apergo.2015.07.012

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Commission Implementing Regulation (EU). (2015). Laying down detailed measures for the implementation of the common basic standards on aviation security 2015/1998 of 5 November 2015. *Official Journal of the European Union*.

Cullen, R. H., Rogers, W. A., & Fisk, A. D. (2013). Human performance in a multiple-task environment: Effects of automation reliability on visual attention allocation. *Applied Ergonomics*, *44*(6), 962–968. https://doi.org/10.1016/j.apergo.2013.02.010

Davis, J., Atchley, A., Smitherman, H., Simon, H., & Tenhundfeld, N. (2020). *Measuring automation bias and complacency in an X-ray screening task* [Paper presentation]. 2020 Systems and Information Engineering Design Symposium (SIEDS), April, 1–5. https://doi.org/10.1109/SIEDS49339.2020.9106670

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors*, *48*(3), 474–486. https://doi.org/10.1518/001872006778606822

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*, *49*(4), 564–572. https://doi.org/10.1518/001872007X215656

Drew, T., Cunningham, C., & Wolfe, J. M. (2012). When and why might a computer-aided detection (CAD) system iinterfere with visual search? An eye-tracking study. *Academic Radiology*, *19*(10), 1260–1267. https://doi.org/10.1016/j.acra.2012.05.013

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, *58*(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

Goh, J., Wiegmann, D. A., & Madhavan, P. (2005). Effects of automation failure in a luggage screening task: A comparison between direct and indirect cueing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(3), 492–496. https://doi.org/10.1177/154193120504900359

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

Harris, D. H. (2002). How to really improve airport security. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *10*(1), 17–22. https://doi.org/10.1177/106480460201000104

Hättenschwiler, N., Sterchi, Y., Mendes, M., & Schwaninger, A. (2018). Automation in airport security X-ray screening of cabin baggage: Examining benefits and possible implementations of automated explosives detection. *Applied Ergonomics*, *72*, 58–68. https://doi.org/10.1016/j.apergo.2018.05.003

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*, 150–119. https://doi.org/10.3389/fnins.2014.00150

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

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3): 407–434. https://doi.org/10.1177/0018720814547570

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. https://doi.org/10.1142/S0218195905001683

Howell, J. (2017). The modern IED: Design and trends. Aviation Security International Magazine, 1–15. https://www.asi-mag.com/modern-ied-design-trends/

Huegli, D., Merks, S., & Schwaninger, A. (2020). Automation reliability, human–machine system performance, and operator compliance: A study with airport security screeners supported by automated

explosives detection systems for cabin baggage screening. *Applied Ergonomics*, *8*, 103094. https://doi.org/10.1016/j.apergo.2020.103094
Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04
Koller, S. M., Drury, C. G., & Schwaninger, A. (2009). Change of search time and non-search time in X-ray baggage screening due to training. *Ergonomics*, *52*(6), 644–656. https://doi.org/10.1080/00140130802526935
Koller, S. M., 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. https://doi.org/10.1007/s12198-007-0006-4
Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
Lyell, D., & Coiera, E. (2017). Automation bias and verification complexity: A systematic review. *Journal of the American Medical Informatics Association*, *24*(2), 423–431. https://doi.org/10.1093/jamia/ocw105
Lyell, D., Magrabi, F., & Coiera, E. (2018). The effect of cognitive load and task complexity on automation bias in electronic prescribing. *Human Factors*, *60*(7), 1008–1021. https://doi.org/10.1177/0018720818781224
Ma, R., & Kaber, D. B. (2007). Effects of in-vehicle navigation assistance and performance on driver trust and vehicle control. *International Journal of Industrial Ergonomics*, *37*(8), 665–673. https://doi.org/10.1016/j.ergon.2007.04.005
Madhavan, P., Wiegmann, D., & Lacson, F. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors*, *48*(2), 241–256. https://doi.org/10.1518/001872006777724408
Maltz, M., & Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition. *Human Factors*, *45*(2), 281–295. https://doi.org/10.1518/hfes.45.2.281.27239
Manzey, D., Gerard, N., & Wiczorek, R. (2014). Decision-making and response strategies in interaction with alarms: the impact of alarm reliability, availability of alarm validity information and workload. *Ergonomics*, *57*(12), 1833–1855. https://doi.org/10.1080/00140139.2014.957732
McBride, S. E., Rogers, W. A., & Fisk, A. D. (2011). Understanding the effect of workload on automation use for younger and older adults. *Human Factors*, *53*(6), 672–686. https://doi.org/10.1177/0018720811421909
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*, 654–610. https://doi.org/10.3389/fnhum.2013.00654
Mendes, M., Schwaninger, A., Strebel, N., Michel, S. (2012). *Why laptops should be screened separately when conventional x-ray screening is used* [Paper presentation]. Proceedings - International Carnahan Conference on Security Technology (pp. 267–273). https://doi.org/10.1109/CCST.2012.6393571
Meng, Y., Nazir, S., Guo, J., & Uddin, I. (2021). A decision support system for the uses of lightweight blockchain designs for P2P computing. *Peer-to-Peer Networking and Applications*, *14*(5), 2708–2718. https://doi.org/10.1007/s12083-021-01083-9
Meuter, R. F., & Lacherez, P. F. (2016). When and why threats go undetected: Impacts of event rate and shift length on threat detection accuracy during airport baggage screening. *Human Factors*, *58*(2), 218–228. https://doi.org/10.1177/0018720815616306
Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors*, *43*(4), 563–572. https://doi.org/10.1518/001872001775870395
Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors*, *46*(2), 196–204. https://doi.org/10.1518/hfes.46.2.196.37335
Meyer, J., Wiczorek, R., & Günzler, T. (2014). Measures of reliance and compliance in aided visual scanning. *Human Factors*, *56*(5), 840–849. https://doi.org/10.1177/0018720813512865
Michel, S., Hattenschwiler, N., Kuhn, M., Strebel, N., Schwaninger, A. (2014). *A multi-method approach towards identifying situational factors and their relevance for X-ray screening* [Paper presentation]. 2014 International Carnahan Conference on Security Technology (ICCST) (pp. 1–6). https://doi.org/10.1109/CCST.2014.6987001
Mosier, K. L., & Manzey, D. (2020). Humans and automated decision aids: A match made in heaven. In P. A. Hancock & M. Mouloua (Eds.), *Human performance in automated and autonomous systems. Current theory and methods* (pp. 19–41). Taylor & Francis Group.
Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *The International Journal of Aviation Psychology*, *8*(1), 47–63. https://doi.org/10.1207/s15327108ijap0801
Nishikawa, R. M., & Bae, K. T. (2018). Importance of better human-computer interaction in the era of deep learning: Mammography computer-aided diagnosis as a use case. *Journal of the American College of Radiology : JACR*, *15*(1 Pt A), 49–52. https://doi.org/10.1016/j.jacr.2017.08.027
Onnasch, L. (2015). Crossing the boundaries of automation — Function allocation and reliability. *International Journal of Human-Computer Studies*, *76*, 12–21. https://doi.org/10.1016/j.ijhcs.2014.12.004
Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors*, *56*(3), 476–488. https://doi.org/10.1177/0018720813501549
Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354
Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055
Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886
Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency". *The International Journal of Aviation Psychology*, *3*(1), 1–23. https://doi.org/10.1207/s15327108ijap0301_1
Petrozziello, A., & Jordanov, I. (2019). *Automated deep learning for threat detection in luggage from X-ray images* [Paper presentation]. International symposium on experimental algorithms (pp. 505–512). https://doi.org/10.1007/978-3-030-34029-2_32
Qinxia, H., Nazir, S., Li, M., Ullah Khan, H., Lianlian, W., & Ahmad, S. (2021). AI-enabled sensing and decision-making for IoT systems. *Complexity*, *2021*, 1–9. https://doi.org/10.1155/2021/6616279
Rice, S. (2009). Examining single- and multiple process theories of trust in automation. *The Journal of General Psychology*, *136*(3), 303–322. https://doi.org/10.3200/GENP.136.3.303-322
Rice, S., & McCarley, J. S. (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. https://doi.org/10.1037/a0024243
Rieger, T., Heilmann, L., & Manzey, D. (2021). Visual search behavior and performance in luggage screening: Effects of time pressure, automation aid, and target expectancy. *Cognitive Research: Principles and Implications*, *6*(1), 12. https://doi.org/10.1186/s41235-021-00280-7
Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 17–22. https://doi.org/10.1167/7.2.17
Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*, *49*(1), 76–87. https://doi.org/10.1518/001872007779598082
Sarter, N., Schroeder, B., & McGuirl, J. (2001). Supporting decision-making and action selection under time pressure and uncertainty:

The case of inflight icing. *39th Aerospace Sciences Meeting and Exhibit*, *43*(4), 573–583. https://doi.org/10.2514/6.2001-543
Schuster, D., Rivera, J., Sellers, B. C., Fiore, S. M., & Jentsch, F. (2013). Perceptual training for visual search. *Ergonomics*, *56*(7), 1101–1115. https://doi.org/10.1080/00140139.2013.790481
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 2004, advances in learning, commerce and security* (pp. 147–156). WIT Press. 10.13140/RG.2.1.4051.8649
Schwaninger, A. (2009). *Why do airport security screeners sometimes fail in covert tests?* [Paper presentation]. Proceedings of the 43rd IEEE International Carnahan Conference on Security Technology, October 5–8 (pp. 41–45). https://doi.org/10.1109/CCST.2009.5335568
Schwaninger, A., Michel, S., Bolfing, A. (2005). *Towards a model for estimating image difficulty in X-ray screening* [Paper presentation]. Proceedings of the 39th IEEE International Carnahan Conference on Security Technology (pp. 185–188). https://doi.org/10.1109/CCST.2005.1594875
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 screening. *GeroPsych*, *23*(3), 169–175. https://doi.org/10.1024/1662-9647/a000020
Schwaninger, A., Michel, S., & Bolfing, A. (2007). A statistical approach for image difficulty estimation in x-ray screening using image measurements. *4th Symposium on Applied Perception in Graphics and Visualization*, *1*(212), 123–130. https://doi.org/10.1145/1272582.1272606
Singh, S., & Singh, M. (2003). Explosives detection systems (EDS) for aviation security. *Signal Processing*, *83*(1), 31–55. https://doi.org/10.1016/S0165-1684(02)00391-2
Skorupski, J., & Uchroński, P. (2016). A human being as a part of the security control system at the airport. *Procedia Engineering*, *134*, 291–300. https://doi.org/10.1016/j.proeng.2016.01.010
Spain, R. D. (2009). *The effects of automation expertise, system confidence, and image quality on trust, compliance, and performance* [Doctoral dissertation]. Old Dominion Uniersity, Norfolk, VA. https://doi.org/10.25777/q87j-mr24
Sterchi, Y., Schwaninger, A. (2015). *A first simulation on optimizing EDS for cabin baggage screening regarding throughput* [Paper presentation]. Proceedings of the 49th IEEE International Carnahan Conference on Security Technology, Taipei Taiwan, September 21–24 (pp. 55–60). https://doi.org/10.1109/CCST.2015.7389657
Turner, S. (1994). *Terrorist explosive sourcebook: Countering terrorist use of improvised explosive devices*. Paladin Press.
Wales, A. W. J., Anderson, C., Jones, K. L., Schwaninger, A., & Horne, J. A. (2009). Evaluating the two-component inspection model in a simplified luggage search task. *Behavior Research Methods*, *41*(3), 937–943. https://doi.org/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. https://doi.org/10.1016/j.apradiso.2012.01.011
Wetter, O., Hardmeier, D., Hofer, F. (2008). *Covert testing at airports: Exploring methodology and results* [Paper presentation]. Proceedings of the 43rd IEEE International Carnahan Conference on Security Technology, Zurich Switzerland, October 5–8, 2009. https://doi.org/10.1109/CCST.2008.4751328
Wiegmann, D. A., McCarley, J. S., Kramer, A. F., & Wickens, C. D. (2006). Age and automation interact to influence performance of a simulated luggage screening task. *Aviation, Space, and Environmental Medicine*, *77*(8), 825–831.
Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science*, *2*(4), 352–367. https://doi.org/10.1080/14639220110110306
Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, *28*(4), 1060–1092. https://doi.org/10.3758/s13423-020-01859-9
Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*(2), 121–124. https://doi.org/10.1016/j.cub.2009.11.066
Xiao, H., Nazir, S., Li, H., Khan, H. U., & Li, C. (2021). Decision support system to risk stratification in the acute coronary syndrome using fuzzy logic. *Scientific Programming*, *2021*, 1–9. https://doi.org/10.1155/2021/6571905
Xiaolong, H., Huiqi, Z., Lunchao, Z., Nazir, S., Jun, D., & Khan, A. S. (2021). Soft computing and decision support system for software process improvement: A systematic literature review. *Scientific Programming*, *2021*, 1–14. https://doi.org/10.1155/2021/7295627
Zhang, J., Nazir, S., Huang, A., & Alharbi, A. (2020). Multicriteria decision and machine learning algorithms for component security evaluation: library-based overview. *Security and Communication Networks*, *2020*, 1–14. https://doi.org/10.1155/2020/8886877
Zirk, A., Wiczorek, R., & Manzey, D. (2019). Do we really need more stages? Comparing the effects of likelihood alarm systems and binary alarm systems. *Human Factors*, *62*(4), 540–552. https://doi.org/10.1177/0018720819852023

About the Authors

David Huegli is a research scientist and PhD student at the Institute Humans in Complex Systems (www.fhnw.ch/miks) at the School of Applied Psychology of the University of Applied Sciences and Arts Northwestern Switzerland. His research focuses on human–machine interaction at airport security checkpoints (automation and 3D imaging).

Sarah Merks is the Head of Project Management at the Center for Adaptive Security Research and Applications in Zurich. Previously, she was a Senior Research Scientist at the University of Applied Sciences and Arts Northwestern Switzerland. Her research focuses on the human factor and human–machine interaction in airport security.

Adrian Schwaninger is the head of the Institute Humans in Complex Systems at the School of Applied Psychology of the University of Applied Sciences and Arts Northwestern Switzerland. His areas of expertise are aviation security, ergonomics and human factors, and human-machine interaction.

Appendix A

Table A1. Test items for the baseline condition.

Container	Target present/absent	CT	Prohibited article	EDSCB alarms	*N*	Selection
Tray	Target present	Notebook	IED	–	8	Recorded
Tray	Target present	Notebook	Guns	–	8	Recorded
Tray	Target present	Notebook	Knives	–	8	Recorded
Tray	Target present	EO	IED	–	8	Recorded
Tray	Target present	EO	Guns	–	8	Recorded
Tray	Target present	EO	Knives	–	8	Recorded
Tray	Target absent	Notebook	–	–	24	Recorded
Tray	Target absent	EO	–	–	24	Recorded
Tray	Target absent	–	–	–	96	Selected
Bag	Target absent	–	–	–	48	Recorded
Bag	Target absent	–	–	–	144	Selected

Note. CT: concealment type; EDSCB: explosive system for cabin baggage; *N*: number; EO: everyday electronic/electrical objects; IED: improvised explosive devices.

Table A2. Test items for the PED-out condition.

Container	Target present/absent	PED	Prohibited article	EDSCB alarms	*N*	Selection
Tray	Target present	Notebook	IED	6	8	Recorded
Tray	Target present	Notebook	Guns	–	8	Recorded
Tray	Target present	Notebook	Knives	–	8	Recorded
Tray	Target present	EO	IED	6	8	Recorded
Tray	Target present	EO	Guns	–	8	Recorded
Tray	Target present	EO	Knives	–	8	Recorded
Tray	Target absent	Notebook	–	17	24	Recorded
Tray	Target absent	EO	–	17	24	Recorded
Tray	Target absent	–	–	–	96	Selected
Bag	Target absent	–	–	–	48	Recorded
Bag	Target absent	–	–	–	144	Selected

Note. EDSCB: explosive system for cabin baggage; PED: personal electronic devices; *N*: number; EO: everyday electronic/electrical objects; IED: improvised explosive devices

Table A3. Test items for the PED-in condition.

Type	Target present/absent	PED	Prohibited article	EDSCB alarms	*N*	Selection
Bag	Target present	Notebook	IED	6	8	Recorded
Bag	Target present	Notebook	Guns	–	8	Recorded
Bag	Target present	Notebook	Knives	–	8	Recorded
Bag	Target present	EO	IED	6	8	Recorded
Bag	Target present	EO	Guns	–	8	Recorded
Bag	Target present	EO	Knives	–	8	Recorded
Bag	Target absent	Notebook	–	17	24	Recorded
Bag	Target absent	EO	–	17	24	Recorded
Bag	Target absent	–	–	–	96	Selected
Tray	Target absent	–	–	–	48	Recorded
Tray	Target absent	–	–	–	144	Selected

Note. EDSCB: explosive system for cabin baggage; PED: personal electronic devices: *N*: number; EO: everyday electronic/electrical objects; IED: improvised explosive devices.