Applied Ergonomics 72 (2018) 58–68

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Automation in airport security X-ray screening of cabin baggage: Examining benefits and possible implementations of automated explosives detection

Nicole Hättenschwiler*, Yanik Sterchi, Marcia Mendes, Adrian Schwaninger

School of Applied Psychology, University of Applied Sciences and Arts Northwestern, Switzerland

ARTICLE INFO

Keywords:
Airport security X-ray screening
Explosives detection
Automation

ABSTRACT

Bomb attacks on civil aviation make detecting improvised explosive devices and explosive material in passenger baggage a major concern. In the last few years, explosive detection systems for cabin baggage screening (EDSCB) have become available. Although used by a number of airports, most countries have not yet implemented these systems on a wide scale. We investigated the benefits of EDSCB with two different levels of automation currently being discussed by regulators and airport operators: automation as a diagnostic aid with an on-screen alarm resolution by the airport security officer (screener) or EDSCB with an automated decision by the machine. The two experiments reported here tested and compared both scenarios and a condition without automation as baseline. Participants were screeners at two international airports who differed in both years of work experience and familiarity with automation aids. Results showed that experienced screeners were good at detecting improvised explosive devices even without EDSCB. EDSCB increased only their detection of bare explosives. In contrast, screeners with less experience (tenure $<$ 1 year) benefitted substantially from EDSCB in detecting both improvised explosive devices and bare explosives. A comparison of all three conditions showed that automated decision provided better human–machine detection performance than on-screen alarm resolution and no automation. This came at the cost of slightly higher false alarm rates on the human–machine system level, which would still be acceptable from an operational point of view. Results indicate that a wide-scale implementation of EDSCB would increase the detection of explosives in passenger bags and automated decision instead of automation as diagnostic aid with on screen alarm resolution should be considered.

1. Introduction

Secure air transportation is vital for both the economy and society (Abadie and Gardeazabal, 2008). For several decades now, airplanes have been interesting targets for terrorists (Baum, 2016). Looking at the history of attacks against airplanes (both successful and near misses), one of the biggest concerns is bombs – that is, improvised explosive devices (IEDs; Novakoff, 1993; Singh and Singh, 2003; Baum, 2016). The Global Terrorism Database (2017) lists 893 attacks on airports or aircrafts with explosives, 247 of which occurred after 2001. Quite recently, on the 29th of July 2017, a terrorist plot was prevented at Sydney airport when an IED was found concealed inside a bag (Westbrook and Barrett, 2017). In response to heightened risk, especially since 9/11, airports and governments have increased their investments in aviation security (Gillen and Morrison, 2015). In the last few years, explosive detection systems for cabin baggage screening (EDSCB) have also become available (Sterchi and Schwaninger, 2015). Whereas a few countries such as the United States are using these systems (Neffenger, 2015), they have not been implemented widely in European countries and on other continents (Pochet, 2016). We investigated the benefits of EDSCB with two different levels of automation that are both being discussed currently by regulators and airport operators. We were able to recruit airport security officers (screeners) from two different European airports to work on two experiments using a simulated cabin baggage screening task. In this introduction, we first summarize previous research on visual inspection and conventional cabin baggage screening before going on to discuss automation and EDSCB.

1.1. Visual inspection and conventional cabin baggage screening

To prevent terrorist attacks and other acts of unlawful interference, passengers and their belongings have to be screened before they are allowed to enter the secure areas of airports and board airplanes (Thomas, 2009). Screeners visually inspect X-ray images of cabin baggage for prohibited items such as guns, knives, and improvised

---

* Corresponding author. University of Applied Sciences and Arts Northwestern Switzerland, School of Applied Psychology, Institute Humans in Complex Systems, Riggenbachstrasse 16, CH-4600 Olten, Switzerland.

E-mail address: nicole.haettenschwiler@fhnw.ch (N. Hättenschwiler).

https://doi.org/10.1016/j.apergo.2018.05.003
Received 15 December 2016; Received in revised form 4 May 2018; Accepted 5 May 2018

explosive devices (IEDs) as well as other items such as self-defence gas sprays or Tasers (Schwaninger, 2005). This inspection involves visual search and decision making (Koller et al., 2009; Wales et al., 2009; Wolfe and Van Wert, 2010). The challenges when performing visual search in X-ray baggage screening include a low target prevalence, the variation in target visibility, the search for an unknown target set, and the possible presence of multiple targets (for recent reviews, see Biggs and Mitroff, 2014; Mitroff et al., 2015). When deciding whether or not a bag contains a prohibited item, screeners need to know which items are prohibited and what they look like as X-ray images (Schwaninger, 2005, 2006). Whereas even novices can recognize certain object shapes such as guns and knives in X-ray images (Schwaninger et al., 2005), other prohibited items such as IEDs are difficult to recognize without training (Schwaninger and Hofer, 2004; Koller et al., 2008, 2009; Halbherr et al., 2013). An IED is composed of a triggering device, a power source, a detonator, and explosive that are usually all connected by wires (Turner, 1994; Wells and Bradley, 2012). Through computer-based training, screeners can learn to recognize these components, and they can achieve and maintain a high detection performance for IEDs (Schwaninger and Hofer, 2004; Koller et al., 2008, 2009; Halbherr et al., 2013; Schuster et al., 2013). In cabin baggage screening, bare explosives also pose a threat, because these could be combined with other IED components after passing an airport security checkpoint. Detecting bare explosives can be a challenge even for well-trained screeners, because they often look like a harmless organic mass (Jones, 2003). So far, no study has investigated how well screeners can detect bare explosives and whether automation and EDSCB can increase human–machine system performance in response to such threats. Before discussing automation and EDSCB as a specific application, it is worth considering important findings and concepts on automation and human–machine system performance in general.

1.2. Automation and human–machine system performance

Automation refers to functions performed by machines (usually computers) that assist or replace tasks performed by humans (for reviews, see Parasuraman and Wickens, 2008; Sheridan, 2011; Vagia et al., 2016). One form of automation assisting humans is the diagnostic aid (Wickens and Dixon, 2007). This provides support in the form of alerts or alarms and influences attention allocation (Cullen et al., 2013). Examples include collision warning systems for driving and air traffic control (Lehto et al., 2000; Abe and Richardson, 2006; Liu and Jhuang, 2012; Biondi et al., 2017) or aids assisting radiologists in making diagnostic decisions from mammograms (e.g. Vyborny et al., 2000; Fenton et al., 2007). Other examples are systems that indicate potentially threatening objects in X-ray images of passenger baggage. These systems have been investigated in laboratory studies with student participants (Wiegmann et al., 2006; Rice and McCarley, 2011). Common to this type of automation is that it categorizes events into target or non-target states (Wickens and Dixon, 2007). Signal detection theory (Green and Swets, 1966, 1972) provides a useful framework with which to describe the performance (reliability) of such diagnostic automation (Wickens and Dixon, 2007; Parasuraman and Wickens, 2008; Rice and McCarley, 2011). In signal detection theory, high performance (reliability) in terms of d' is achieved when targets are detected well (high hit rate) and the false alarm rate is low. The criterion (or response bias) is a threshold that can be changed while d' remains constant (Macmillan and Creelman, 2005). The criterion can be changed by adjusting thresholds for alerts, resulting in a trade-off between two types of automation errors: misses and false alarms (Parasuraman, 1987; Parasuraman and Riley, 1997; Wickens and Colcombe, 2007). Designers often set low thresholds, because the consequences of automation misses are considered to be more costly than false alarms (Parasuraman and Wickens, 2008). However, if the base rate of dangerous events to be detected is low, the result will be many false alarms and only few hits (Parasuraman and Riley, 1997). This can produce a 'cry wolf' effect with operators ignoring system warnings (Breznitz, 1983; Bliss, 2003). Such an effect can drastically reduce or even eliminate the benefits of automation when it is implemented as a diagnostic aid.

Alongside automation as a diagnostic aid, other levels of automation are possible. Sheridan and Verplank (1978) proposed a taxonomy with 10 levels of automation ranging from fully manual to fully computer automated. Parasuraman et al. (2000) proposed a taxonomy with four processing stages: 1) sensory processing, 2) perception/working memory, 3) decision making, and 4) response/action. Several other taxonomies for different levels of automation have been proposed (for a review, see Vagia et al., 2016). Kaber and Endsley (2003) have pointed out that specifying the 'best' level of automation is not as straightforward as one might think. Moreover, familiarity with automation can affect how people interact with it (Parasuraman and Manzey, 2010; Sauer et al., 2016; Strauch, 2016; Sauer and Chavaillaz, 2017). Indeed, deciding how best to organize human–machine function allocation and the level of automation remains a difficult task that can also depend on the specific application (Sheridan, 2011). Parasuraman et al. (2000) have suggested that appropriate criteria for selecting the level of automation for a particular application are human performance, automation reliability, and the cost associated with outcomes.

1.3. Automation and EDSCB

For X-ray screening of cabin baggage, regulators and airport operators are currently discussing two EDSCB implementation scenarios differing in their level of automation and human–machine function allocation: on-screen alarm resolution (OSAR) and automated decision (Sterchi and Schwaninger, 2015). In the OSAR scenario, automation is implemented as a diagnostic aid. Screeners visually inspect *every* piece of cabin baggage. During this inspection, EDSCB indicates potential explosive material by either marking an area on the X-ray image of a passenger bag with a coloured rectangle or highlighting it in a special colour (Nabiev and Palkina, 2017). Screeners then have to resolve this; that is, they have to visually inspect the X-ray image and decide whether the area indicated by the machine is harmless (EDSCB false alarm) or whether it actually could be explosive material, making it necessary to subject the baggage to a secondary inspection. This is also conducted at the airport security checkpoint and involves explosive trace detection, opening the bag, and manually searching it (Sterchi and Schwaninger, 2015). EDSCB systems with high hit rates (close to 90%) have false alarm rates in the range of 15–20% (personal communication with EDSCB experts, summer 2016). As mentioned above, system reliability can be described by d' from signal detection theory (Green and Swets, 1966, 1972). For example, an EDSCB with a hit rate of 88% and a false alarm rate of 17% would have a system reliability of $d' = 2.1$. In operation, most of the EDSCB alarms are cleared by screeners, leaving only a small percentage of bags on which EDSCB has raised an alarm that then requires a secondary inspection. Although OSAR is the scenario currently employed at airports that have already introduced EDSCB, its effectiveness can be questioned, because screeners might not be able to distinguish explosive material from benign material (as pointed out already by Jones, 2003). Moreover, EDSCB false alarm rates of 15–20% could result in a cry wolf effect leading screeners to potentially ignore system warnings (Breznitz, 1983; Bliss, 2003). Screeners might therefore be prone to mistakenly clearing bags that contain explosives. This would drastically reduce the effectiveness of EDSCB in the OSAR scenario. In other words, the probability of detecting explosives on the human–machine system level equals about 90% (EDSCB) *minus* the erroneously cleared alarms by screeners. This could result in a much lower detection rate.

The automated decision scenario uses a higher level of automation with different human–machine function allocation. Bags on which the EDSCB raises an alarm are sent *automatically* to secondary inspection using manual search and/or explosive trace detection (Sterchi and