Screening baggage in airports now and then – cognitive and operational challenges for training and competence development

Signe Maria WAECHTER, Stephanie WALTER

Zurich State Police Airport Division, Research and Development
Postfach, CH-8058 Zurich

Abstract. Within seconds, a screener must decide whether a bag is clear and can be released or if the item needs further investigation. The content and structure of passenger baggage is influenced by various factors, e.g., increasing security restrictions or different airline fares). The goal of this study was to investigate the challenges of cabin baggage now and then for the screeners. Certified screeners (*N* = 68) saw a series of baggage items from 2008 and 2016 and judged whether the item was clear or not while their reaction time was measured. Further, they rated the complexity and difficulty of the bag images. The reaction time for newer images containing a forbidden item was longer. Further, error rate was higher for newer images. On the other hand, older bag images were perceived as more complex and difficult. The results underscore the cognitive demands that might implicitly impede screeners in their work.

Keywords. screener, security control, cabin baggage, cognition

1. Introduction

The proximity and the frequency of terror attacks is increasing. Only in the last two years there was a series of terrorist attacks in Europe (Brussels, March 2016; Istanbul, June 2016). In the past, aviation has often been chosen as targets and remains an attractive target, because the possible fatalities are high and bringing down an airplane has an enormous symbolic impact. Moreover, an attack on an airport can also damage the economy and the reputation of a country. Under these considerations, the importance of security screening is crucial. Simultaneously, economic pressure on airports is also rising. Screeners conducting security controls at airports are challenged to work quickly and efficiently without compromising security levels. This job requires specific cognitive capabilities that should be present in new recruits, and these capabilities should also be instituted in screeners' ongoing training (Hardmeier et al. 2005).

Screeners have to decide within seconds whether a cabin baggage does not contain any forbidden item and can be released to the passenger or whether it needs a more precise investigation. This means that screeners have to be able to detect forbidden and dangerous goods. To work as a screener, the person has to pass a certification test and to remain the certificate, screeners have to train on a regular basis with a computer-based training program. However, the certification and training program match the reality only partially as the content of cabin baggage might change quicker and can be influenced by other factors, such as airline fares. For example, the so-called light fare that does not include a hold baggage could lead to denser cabin baggage that might be more challenging to screen. Furthermore, the

increasing amount of cell phones and other home electronics (e.g., tablets, laptops) can be hindering factors. Moreover, also new security regulations can influence the screening process, such as the liquid regulations that regulates the transportation of liquids on airplanes since 2008. Finally, human decision-making is not always purely rational and certain pieces of information or an affect provoked by a stimulus can bias a decision (Gigerenzer & Gaissmaier 2011; Tversky & Kahneman 1974). People often rely on heuristics to avoid cognitive effort. Therefore, also the decision-making process when confronted with an item can be influenced by the use of certain heuristics, that is, cognitive shortcuts to reach a fast and frugal decision. Keeping all these factors in mind, permanently increasing the competence of screeners and understanding the psychological factors of screening processes is essential to ensure the detection of forbidden and dangerous goods in baggage (Schwaninger et al. 2007).

This study investigated whether the screening of current cabin baggage is more difficult and takes more time compared to some years ago. The authors of the paper hypothesized that due to the beforehand mentioned factors current cabin baggage is packed denser and therefore it can be hypothesized that screeners need more time for their decisions, especially when the baggage contains a forbidden item. An experimental study was conducted to investigate these research questions. In order to gain an objective measurement, participants' reaction times where assessed in a first part of the experiment. In a second part of the study, participants' perceived difficulty and density of different bag images were assessed. Finally, this paper discusses reasons and implications of the results for training and competence development of screeners

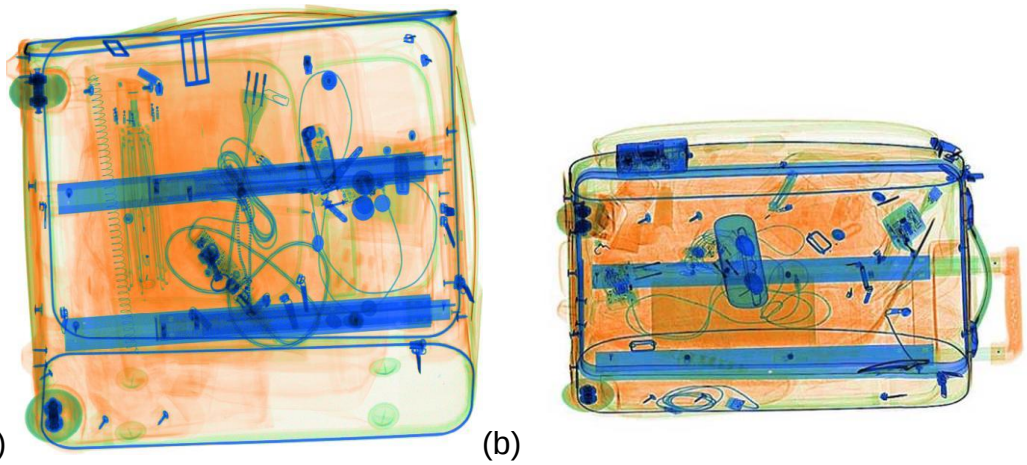
2. Methods and Materials

The experiment consisted of two parts. The goal of the first part was to analyse participant's reaction times when evaluating items from 2008 and from 2016. It was assumed that the reaction times for baggage from 2016 would be longer compared to 2008 due to increased density. The goal of the second part was to analyse participants' subjective perception of the difficulty and the density of the bag items from 2008 and 2016. Again, it was assumed that participants' would judge items from 2016 as more difficult and denser than items from 2008.

An experimental within-subjects design was chosen to investigate the research questions. Real pictures of screened baggage from 2016 and from 2008 were chosen. It was taken care to select different types of luggage (e.g., trolley, back bag) to be as close to the reality as possible. The selected pictures were verified by an independent and well-experienced screener. The experiment was programmed online with the provider Sosci Survey (https://www.soscisurvey.de/). 68 certified screener currently working at a big international airport participated in the experiment. The mean age was 43 years (SD = 11). It was arranged that the screeners could participate in a short break during their regular working shift. They did not receive any incentive for their participation.

For the first task, 30 pictures from 2008 and from 2016 were selected. Each picture equalled one baggage or one tray, respectively. Out of the 30 items, 7 items contained a forbidden good (e.g., liquids). Participants read a short introduction about the study informing them that they would see a series of pictures. They were asked to evaluate the items as they would do at the security checkpoint. They were not

informed that they were presented items stemming from two different years. After agreeing to participate in the study, participants were presented one picture and where asked whether the baggage is clear or needs further investigation (i.e., contains forbidden items, items partially covered). After selecting *yes* or *no* the next picture was presented. The order of the pictures changed randomly. Two exemplary items are depicted in Figure 1. In this part, the reaction time was assessed to gain an objective measurement of the difficulty to judge the item.



(a) (b)

Figure 1. *Examples of items presented to participants in the first task without any forbidden goods. Item (a) is from 2016 and item (b) from 2008.*

In the second task, participants were again informed that they would see a series of pictures and would need to answer some questions. In total, they were presented with five pictures of baggage from 2016 and five from 2008 that they did not have seen before and that did not contain any forbidden goods. After each picture, a new page appeared. The order of the pictures changed randomly. Participants were asked to judge how difficult the item is to evaluate and how dense the luggage was. An example of the second task is depicted in Figure 2. Finally, participants were asked to answer some sociodemographic questions and were then thanked for their contribution.

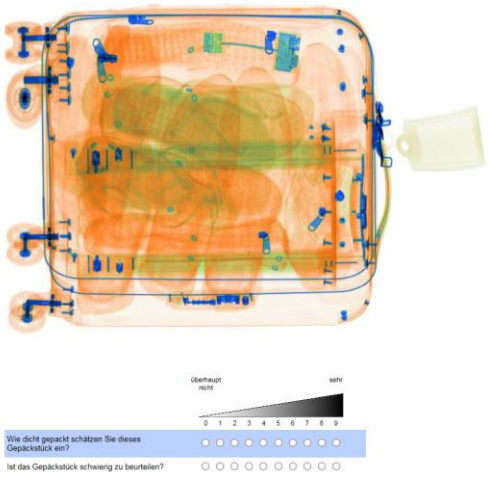


Figure 2*. Example of an item in the second task with a luggage from 2016.*

3. Results

The analysis of the results was conducted with RStudio (RStudio Team 2015). The mean reaction time for all pictures from 2008 and all pictures from 2016 was calculated as well as the mean reaction time for the items containing a forbidden item (i.e., DG items). To ensure normal distribution, the values were log-transformed for further analysis. A paired t-test was applied revealing a non-significant result ($t\ [67]$ = .411, p = .68) for the difference in reaction time for all pictures from 2008 and all pictures from 2016. Furthermore, there was a significant difference in reaction time for DG items ($t\ [67]$ = -5.60, p < .001) indicating that the detection of forbidden goods in items from 2016 takes more time compared to 2008 (see Figure 3.).

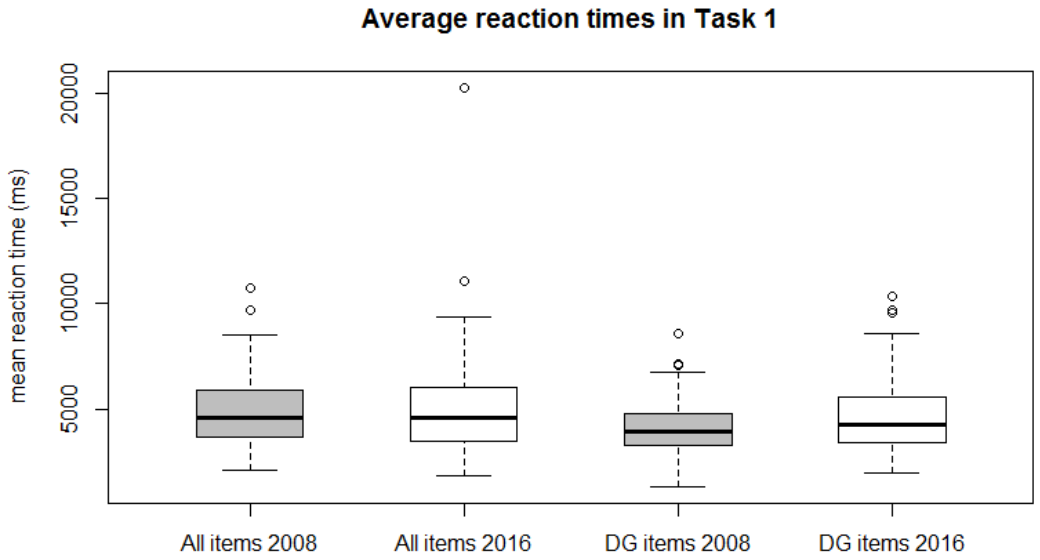


Figure 3. *Mean reaction times for 2016 and 2008. Values before transformation are shown.*

As for luggage without any DG item, there might be some individual variance whether a screener would send the luggage to a re-screening or not. However, when confronted with a DG Item, all participants should send the item to re-screening (i.e., tick *yes* in the question). Therefore, the percentage of correct answers for DG items for 2008 and for 2016 was calculated and compared using a paired t-test. The test revealed a significant difference ($t\ [67]$ = 4.83, p < .001) indicating that detecting forbidden goods in items from 2008 (M = 72.69, SD = 16.87) was easier compared to items from 2016 (M = 61.34, SD = 20.51).

For the second task, the mean rating for difficulty and density for 2008 and 2016 was calculated. Using a paired t-test, results showed that the mean rating for 2008 was lower in both difficulty ($t\ [67]$ = 10.86, p < .001) and density ($t\ [67]$ = 10.38, p < .001) compared to 2016 (see Figure 4.).

4. Discussion

This paper aimed to show that detecting forbidden items in more recent items is more difficult (i.e., longer reaction times) and that these items are perceived as more difficult to evaluate and as higher in density. Results have shown that on average

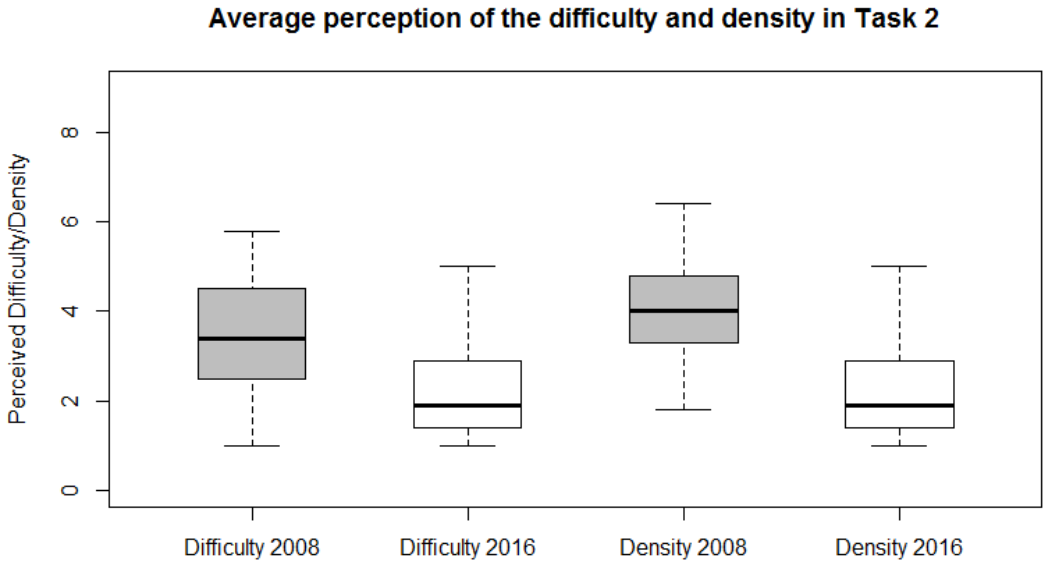


Figure 4*. The graph shows the mean ratings for difficulty and density for items from 2008 and from 2016. Participants were asked how difficult the item is to evaluate and how dense the item is. Lower values indicate that the items were easier and less dense, respectively.*

evaluating luggage items from the years 2008 and 2016 takes about the sameamount of time. Nevertheless, the detection of forbidden items takes about 15% more time for items from 2016 than from 2008. Further, they were less often correctly identified. For the prevention of a terrorist attack on airplanes, this means that screeners might be more challenged by the current luggage. One reason for the increasing difficulty can be that hand luggage might be denser packed due to different airline fares (e.g., light fare). However, the actual density of the used bag images was not determined.

Nevertheless, a second reason is provided by the results of the second task of the study. It was shown that more recent items (i.e., items from 2016) are perceived as less dense and less difficult to evaluate by screeners. This result might indicate that recent luggage is (unconsciously) more familiar to the screeners due to the better-known content compared to the luggage from eight years ago. Due to this perceived familiarity with the items, participants might have rated the items as less difficult and less dense. This type of behaviour can be traced back to research on decision making under uncertainty suggesting that people often do not make purely rational decisions, but rely on certain heuristics, that is, cognitive shortcuts (Tversky & Kahneman 1974). For example, people tend to develop a stronger positive attitude towards stimuli they are more acquainted and more frequently confronted with (Harrison 1977; Slovic et al. 2007, Zajonc 1968) compared to more ambiguous stimuli. According to this reasoning, heuristic judgment and decision-making can therefore be biased due to the affect provoked by a stimuli. As a consequence, screeners might unconsciously judge items appearing more familiar as less challenging and – potentially – scan these items less thoroughly. In other words, screeners could apply a cognitive shortcut and might unwittingly pay less attention to some luggage. The reliance on such a decision heuristic, could jeopardize the

security control. Moreover, the influence of such a familiarity bias could especially be concerning as this study has found that the detection and correct identification of forbidden goods in recent items was more challenging. It cannot be ruled out that the effect was influenced by the items used and would be different with other items. Moreover, familiarity or affect towards different items was not measured. This was an exploratory study and more research is therefore required to fully understand the results.

To conclude, the tendency to get accustomed to items due to their familiarity thereby neglecting forbidden goods (i.e., familiarity bias) could constitute a potential threat risk. The development and impact of such a familiarity bias on screeners working performance needs to be further investigated. Based on these results, it can only be speculated what constitutes familiarity of luggage. For competence development, it seems imperative that screeners are confronted with a variety of pictures of luggage. This could for example be achieved by assigning screeners to different lines (e.g., staff, business, economy) and different baggage types (i.e., cabin and hold baggage). It could also have some implications for the use of threat-image protection (TIP) that is currently used in many airports (Schwaninger 2006). Keeping screeners aware and curious seems crucial to ensure the correct identification of forbidden goods.

5. Literatur

Hardmeier, D., Hofer, F., & Schwaninger, A. (2005) The x-ray object recognition test (x-ray ort) – a reliable and valid instrument for measuring visual abilities needed in x-ray screening. *IEEE ICCST Proceedings, 39,* 189-192. doi:10.1109/CCST.2005.1594876

Harrison, A. A. Mere exposure. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 10). New York: Academic Press, 1977.

Gigerenzer G, Gaissmaier W (2011) Heuristic decision making. Annu Rev Psychol. 62(1):451–82. doi: 10.1146/annurev-psych-120709-145346 PMID: 21126183.

RStudio Team (2015) RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

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

Schwaninger, A. (2006). Threat image projection: enhancing performance? *Aviation Security International, December 2006,* 36-41.

Slovic, P., Finucane, M.L., Peters, E., MacGregor, D.G. (2007) The affect heuristic. Eur. J.Oper. Res. 177, 1333e1352. http://dx.doi.org/10.1016/j.ejor.2005.04.006.

Tversky, A., & Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131.

Zajonc, R. B. (1968) Attitudinal effects of mere exposure. Journal of Personality and Social Psychology Monographs, 9(2, Pt.

Apparies RJ, Riniolo TC, Porges SW (1998) A psychophysiological investigation of the effects of driver longer-combination vehicles. Ergonomics 41:581-592.

Acknowledgment: The authors would like to thank Zurich Airport AG for their kind support.

Gesellschaft für
Arbeitswissenschaft e.V.

Soziotechnische Gestaltung des digitalen Wandels – kreativ, innovativ, sinnhaft

63. Kongress der
Gesellschaft für Arbeitswissenschaft

FHNW Brugg-Windisch, Schweiz

15. – 17. Februar 2017

GfA Press

Bericht zum 63. Arbeitswissenschaftlichen Kongress vom 15. – 17. Februar 2017

FHNW Brugg-Windisch, Schweiz

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Dortmund: GfA-Press, 2017
ISBN 978-3-936804-22-5

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle erhältlich.

Schriftleitung: Matthias Jäger
im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

USB-Print: Dr. Philipp Baumann, Olten

Screen design und Umsetzung
© 2017 fröse multimedia, Frank Fröse
office@internetkundenservice.de · www.internetkundenservice.de