International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Evaluating Artifact Effects in Threat Image Projection for X-Ray Baggage Screening and Strategies for Improvement

Robin Riz à Porta, Yanik Sterchi & Adrian Schwaninger

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2025.2598113

Evaluating Artifact Effects in Threat Image Projection for X-Ray Baggage Screening and Strategies for Improvement

Robin Riz à Porta[a], Yanik Sterchi[a,b] and Adrian Schwaninger[a,b]

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

ABSTRACT

Many airports worldwide employ threat image projection (TIP) that occasionally projects prerecorded threat items (e.g., bombs) into X-ray images of passenger baggage to enhance and monitor security officers' detection performance. However, some TIP-generated images appear artificial, potentially undermining the effectiveness of TIP. We conducted a simulated screening task with security officers using TIP images from an international airport. We applied path modeling to analyze how TIP artifacts affect threat detection and interact with other image characteristics. We found that threats aligning poorly with other items increased threat detection. Furthermore, artifacts occurred more often when the threat items were projected onto less complex bags or loose items in a tray. TIP should still be effective, because artifacts only occurred in every sixth image and showed a small impact on threat detection compared to other factors. To increase TIP effectiveness, projection onto baggage images with moderate complexity can significantly reduce artifacts while producing more realistic threat images.

KEYWORDS

Human factors; design and evaluation of innovative interactive systems; intelligent tutoring; coaching and debugging systems

1. Introduction

In airports, trained security officers (called screeners) search the X-ray images of passenger baggage for threat items, such as bombs, guns, and knives, to prevent terrorist attacks (Harris, 2002). The screening task entails visual search and decision-making (Koller et al., 2009), presenting several challenges (Biggs & Mitroff, 2015; Donnelly et al., 2019; Schwaninger, 2006). One is that threat items are rarely present, and this reduces the likelihood of detecting them (Buser et al., 2020; Godwin et al., 2010; Wolfe et al., 2007). As a countermeasure, most airports worldwide deploy threat image projection (TIP) systems, which increase target prevalence by projecting fictional threat items (FTIs) onto X-ray images of to-be-screened passenger baggage (Hofer & Schwaninger, 2005; Meuter & Lacherez, 2016; Skorupski & Uchroński, 2016). TIP also provides instant feedback when a screener detects or misses a threat image, thereby ensuring regular feedback that would have been rare (Wolfe et al., 2013). Regular feedback is an important contributor to job motivation (Hackman & Oldham, 1976; Humphrey et al., 2007). TIP systems also record screeners' detection performance, and the generated TIP data are used to monitor screeners' performance (European Commission, 2015; Hofer & Schwaninger, 2005; Meuter & Lacherez, 2016; Skorupski & Uchroński, 2016). To ensure TIP effectiveness, TIP images must be realistic. Screeners have reported that they sometimes detected threat images because they appeared artificial (Bassetti, 2018). This could present a challenge as screeners might miss real threat items by focusing on detecting artifacts. Furthermore, when screeners detect TIP images because of artifacts, monitoring their detection performance from TIP data would be invalid.

CONTACT Robin Riz à Porta robin.rizaporta@fhnw.ch School of Applied Psychology, Institute Humans in Complex Systems, University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland

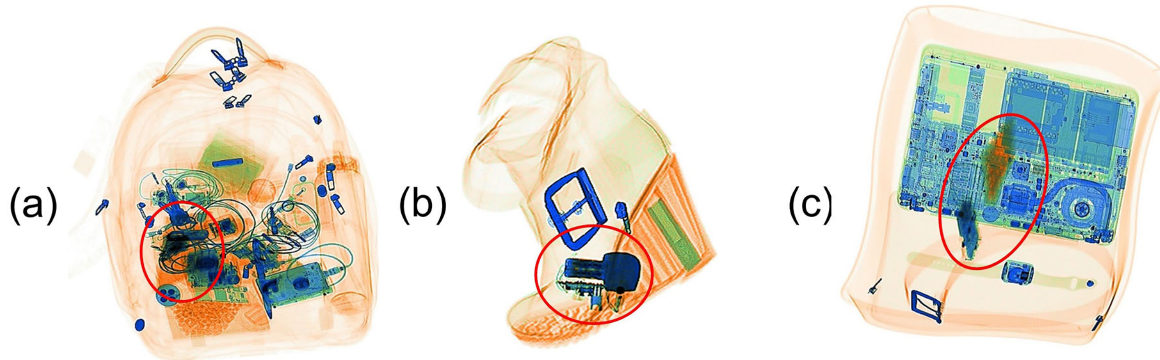

Figure 1. Examples of artifacts in TIP images.
Note: Examples of real TIP images presented to screeners at the security lanes: with a low rating regarding artifacts (a), with a high rating regarding placement artifact (b), with a high rating regarding alignment artifact (c); red circles indicate the locations of the FTIs for the readers' convenience and are not shown in the original TIP images.

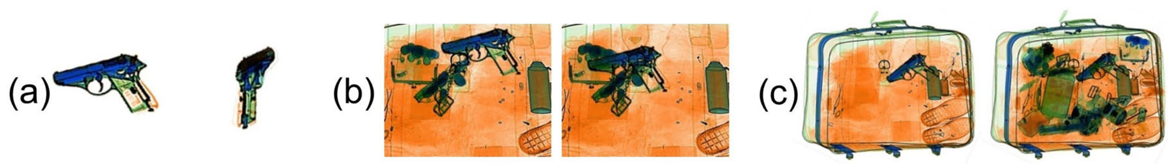

Figure 2. Illustration of image based factors.
Note: Illustration of view difficulty (a), superposition (b), and bag complexity (c). Adapted from Schwaninger et al. (2005).

We previously assessed existing artifacts and their occurrence frequencies by first interviewing screeners, followed by allowing them to rate a sample of TIP-generated threat images for the presence of artifacts (Riz à Porta et al., 2022). We observed the occurrence of two main types of artifacts, which affected approximately one in six images. The first type of artifacts, referred to as placement artifacts, occurs when the projected threat items appear to penetrate other objects in the baggage. Figure 1(a,b) depict a regular TIP-generated image without an artifact and with a placement artifact, respectively (the projected bomb appeared to penetrate the solid heel of a shoe). The second type, referred to as alignment artifacts, occurs when the projected threat items align poorly with other items in the baggage. Figure 1(c) shows an example, where a laptop, watch, and belt are aligned with the tray, whereas the projected bomb exhibits a different orientation.

Artifacts were observed to occur much less in images of bags, suitcases, backpacks, or other pieces of baggage (occurring in approximately 7% of the TIP images; see Figure 1(a) for an example of a threat in a backpack) than in images of shoes, jackets, wallets, keys, or other loose items in a tray (occurring in approximately 45% of the TIP images; see Figure 1(b,c) for examples of shoes or loose items in a tray). Therefore, one strategy for reducing artifacts would be by projecting more threat items onto pieces of baggage. It has been observed that artifact occurrence is associated with the baggage image characteristics that affect the detection of threat items, i.e., the so-called image-based factors, IBFs (Schwaninger et al., 2005). As illustrated in Figure 2, these factors include view difficulty (orientation of the threat item), superposition (the degree to which the threat item is superimposed by other items), and bag complexity, which comprises clutter (background unsteadiness) and opacity (proportion of the image that is opaque). In the previous study, a lower occurrence of artifacts was observed when the view difficulty, superposition, or baggage complexity was high; however, their respective roles in inhibiting artifact occurrence are unclear.

Despite evidence of TIP artifacts, their specific impact on threat detection and their interactions with image-based factors like view difficulty and bag complexity remain poorly understood, necessitating further investigation. Previous findings revealed that artifacts occur, but the extent to which they affect the detection of threat items is still unclear. Additionally, although previous studies revealed the relationship between artifacts and IBFs, the individual role of each IBF in facilitating artifact occurrence is still also unclear. Thus, to bridge both gaps in the literature, we measured the frequency with which

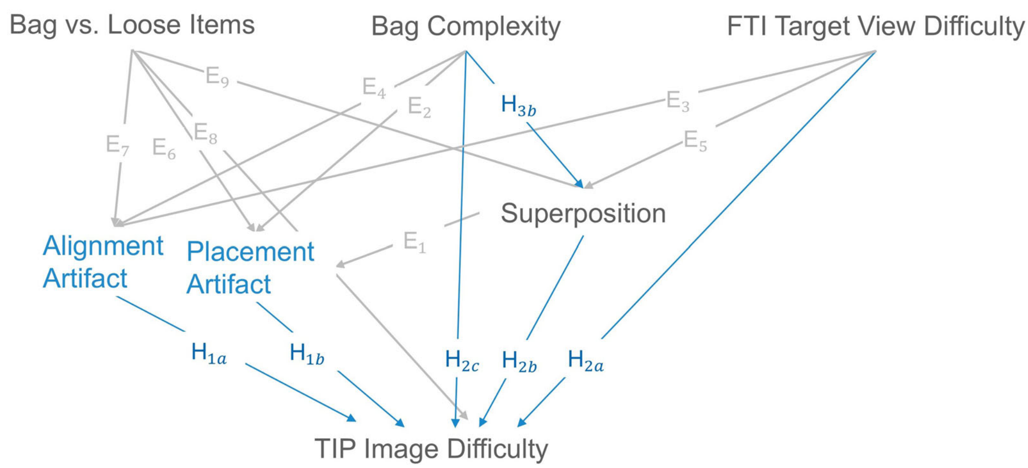

Figure 3. Hypothesized model.

screeners detect or miss threat items in a representative sample comprising 600 TIP images that had been rated regarding the presence of artifacts and IBFs (Riz à Porta et al., 2022). Thereafter, the obtained data were used to estimate a model of how artifacts and other image characteristics interact, as well as how they affect threat image detection.

Figure 3 shows the hypothesized model based on the results of previous studies and the sequence in which FTI merges with the baggage image to generate the projected images. The first layer comprises the view difficulty of FTI (*FTI Target View Difficulty*), the complexity of the baggage image (*Bag Complexity*), and whether FTI ends up in a piece of baggage, e.g., bags, or on loose items in a tray (*Bag vs. Loose Items*). The middle layers comprise superposition, alignment artifact, and placement artifact resulting from the merging of the FTI with the baggage image. The last layer comprises the difficulty of detecting a TIP image as threat detection only occurs after all other characteristics are given.

For the specific effects among the different variables, we hypothesized that screeners detect TIP images better in the presence of placement (H_{1a}) or alignment artifacts (H_{1b}) and poorly when the threat item is viewed from a challenging angle (H_{2a}), is superimposed by other items (H_{2b}), or is projected onto complex baggage (H_{2c}), as reported in prior research (Bolfing et al., 2006, 2008; Hardmeier et al., 2005; Schwaninger et al., 2005, 2007, 2008). We also hypothesized that bag complexity increases superposition (H_{2d}) and indirectly affects the probability of detecting a threat in a TIP image (H_{2e}), as indicated in previous studies (Bolfing et al., 2006; Schwaninger et al., 2007). In these studies, bag complexity influenced threat detection in TIP images, although its effect diminished when superposition was introduced as a predictor. This indicates that bag complexity primarily influences threat detection indirectly.

The other potential effects could not be hypothesized from the existing literature; therefore, they were based on logical deduction– their results will be interpreted as exploratory. We anticipated that superposition would affect the occurrence of placement artifacts considering that the more an FTI is superimposed on other items, the more likely such an FTI might appear to penetrate the items. Furthermore, we expected that bag complexity would affect the occurrence of placement artifacts considering that with high bag complexity, more item layers are stuffed in a bag with more options for an FTI and the other items to be placed in different layers without appearing to penetrate each other. Regarding alignment artifacts, we anticipated that view difficulty would affect their occurrence. That is, guns and knives have a low view difficulty, when they lie flat (Koller et al., 2008; Schwaninger et al., 2008) and when they life flat, they might also be less likely to show unrealistic alignment with other items which tend to lie flat as well. Furthermore, we anticipated that bag complexity would affect the occurrence of alignment artifacts as an FTI might more likely align with some of the other items as bag complexity increases, entailing an increasing number of items and disorders in the bag. Further, we expected that view difficulty would affect superposition, as the image of a threat item is much smaller in certain orientations, making it less likely to be superimposed.

As previously stated, artifacts were much rarer in the images of bags, suitcases, backpacks, and other pieces of baggage than in the images of shoes, jackets, wallets, keys, and other loose items in a tray.

Despite the presence of artifacts, we assumed that the superposition and threat detection were affected by whether an FTI is projected onto a piece of baggage.

2. Methods

2.1. Participants

The participants of this study included 83 professional cabin baggage screeners of a large European airport (47 female and 36 male), and none of them participated in our previous study on artifact occurrence (Riz à Porta et al., 2022). They had between 1 and 32 years of working experience ($M = 8.65$, $SD = 6.21$) and were aged between 22 and 65 years ($M = 44.83$, $SD = 11.93$). The study complied with the American Psychological Association's Code of Ethics and was approved by the Institutional Review Board of the School of Applied Psychology, University of Applied Sciences and Arts Northwestern Switzerland. Data were anonymized to protect participants' identities.

2.2. Materials

We employed 600 single view TIP images from the previous study as stimuli (Riz à Porta et al., 2022). Each image was used two times: once with the projected threat item and once without it. The images were divided into four blocks, each containing 150 images. Further, 450 images without threat items were added to each block, resulting in a target prevalence of 25%. In each block, the same TIP image was only shown once either with the projected threat item or without it. The 600 images were sampled by a stratified random sampling approach from 14 X-ray machines at the airport where the screeners worked. Notably, 300 images were sampled from X-ray images obtained in winter, and 300 were sampled from X-ray images obtained in summer to consider seasonal changes to the baggage content. Furthermore, the missed TIP images were oversampled to increase the overall difficulty level and avoid a lack of statistical power to detect the various effects on the probability that threats are missed in TIP images (TIP image difficulty). Specifically, although only 12% of the threats in TIP images were missed at the airport, we sampled 150 images (25% of our dataset) from such misses – the remaining 450 images (75%) were sampled from hits. For prevalence descriptions of the image characteristics (Section 3), the representative values were reported by correcting for oversampling.

2.3. Procedure

To determine TIP image difficulty, the participants performed a simulated baggage X-ray screening task. Each participant was randomly assigned to one of the four blocks of images ($n = 21$, 21, 22, and 20). Within each block, images were further divided into three subblocks of 200 images. The three image subblocks were presented in a counterbalanced order using a Latin square design and within the subblocks, the images were presented in random order to control for order effects (e.g., screener fatigue). The participants were instructed to analyze the images as they would when working. They were required to indicate the position of a suspected threat item by double-clicking on its location, followed by clicking the "NOT OK" button. In the absence of a suspected threat item, they were required to click the "OK" button. The participants took a 10 min break after every 200 images to mitigate fatigue. They completed the task in 1–1.5 h.

2.4. Measures

Bag vs. loose items represented a dichotomous variable describing whether the FTI was projected onto a piece of baggage (1) or on loose items in a tray (0). Following Schwaninger et al. (2005) approach, *FTI target view difficulty* was calculated as the rate at which airport security officers missed an FTI (Equation (1)) It was calculated using the data from the airport where the participants worked and from a minimum of 29 responses per FTI ($M = 375.53$, $SD = 194.27$).

Table 1. Image characteristics, rating questions, and anchors.

Characteristics	Ratings	Anchors
Alignment	FTI alignment of the FTI seems artificial.	Totally disagree (1) Totally agree (7)
Placement	The location of the FTI is unlikely.	Totally disagree (1) Totally agree (7)
Superposition	Superposition of the FTI by other items	Very low (1) Very high (7)
Bag complexity (clutter)	Clutter in the baggage	Very low (1) Very high (7)
Bag complexity (opacity)	Proportion of the image that is opaque	Very small (1) Very large (7)

$$Fti_{TVD} = \frac{N_{Misses}}{N_{Projections}}. \tag{1}$$

Furthermore, the placement and alignment artifacts, bag complexity, and superposition were adopted from our previous study (Riz à Porta et al., 2022). In the present study, the average rating for each TIP image was calculated from a minimum of 12 professional screeners who had rated the images on seven-point rating scales (Table 1) with excellent agreement (ICC $>$ 0.75; Cicchetti, 1994). Additionally, bag complexity was calculated as the mean of the clutter and opacity (Bolfing et al., 2008; Schwaninger et al., 2008), which had been separately rated. Images with an average rating $\geq$3.5 on this complexity were assumed to be of medium-to-high bag complexity (because the rounded rating is either on the mid-point or on the higher end of the complexity scale). The TIP images with an average rating $\geq$5 on the respective scale were assumed to contain alignment or placement artifacts, as the screeners agreed on the presence of the artifact. TIP image difficulty represents the miss rate of the TIP images calculated from the performance of $\geq$20 screeners (Section 2.3).

2.5. Analyses

The interdependencies among the variables were estimated by path analysis in R (R Core Team, 2016) using the lavaan package (Rosseel, 2012). To satisfy the assumption of path analysis regarding the linear relationships among variables (examined by scatter plots), *bag complexity* and *superposition* were log-transformed, following Liu's (2016) approach. As the assumption of multivariate normal distribution was violated, path analysis was performed with a robust maximum likelihood method, which provides robust standard errors and a corrected χ^2 value (Satorra & Bentler, 1994). As mentioned above, four blocks of images were each presented to a different group of screeners. To control for any potential differences in screener competency between the groups (e.g., due to experience or cognitive abilities), we compared linear regression models with and without group as a predictor, and we found that the model including group should be preferred for TIP image difficulty ($F(3, 590) = 5.92$, $p < 0.001$), alignment artifact ($F(3, 593) = 15.23$, $p < 0.001$), and placement artifact ($F(3, 593) = 49.29$, $p < 0.001$), but not for superposition ($F(3, 593) = 0.46$, $p = 0.17$). The path model controlled for group accordingly. To assess the precision of the prevalence of artifacts in X-ray images and of the difficulty of these images by bag characteristics (bag complexity, and whether FTI is in a piece of baggage), 95% bias-corrected and accelerated (BCa) bootstrap confidence intervals were computed based on 10,000 resamples (Efron & Narasimhan, 2020).

3. Results

Table 2 presents the correlations, means, and standard deviations of the path model variables. The hypothesized model showed good fit indices (CFI = 0.99 and SRMR = 0.02) above recommended cutoffs (Hu & Bentler, 1999). Figure 4 shows the model and the estimated standardized path coefficients, and Note. Standardized path coefficients and explained variance (R2) of variables. Solid lines indicate significant relationships. Blue lines are the results of the main hypotheses and grey lines are the results of the exploratory hypotheses.

Table 2. Means, standard deviations, and Pearson correlations between the variables of the path analysis model.

Variable	*M (SD)*	1	2	3	4	5	6
1. FTI target view difficulty	0.09 (0.11)						
2. Log bag complexity	0.57 (0.49)	0.05					
3. Bag vs. loose items[a]	0.60 (0.49)	0.06	0.60***				
4. Log superposition	0.62 (0.47)	0.08*	0.76***	0.41***			
5. Alignment artifact	3.15 (1.11)	−0.16***	−0.33***	−0.034***	−0.29***		
6. Placement Artifact	2.91 (1.38)	−0.11**	−0.56***	−0.69***	−0.43***	0.65***	
7. TIP image difficulty	0.12 (0.21)	0.59***	0.29***	0.20***	0.37***	−0.35***	−0.29***

Note: [a]Dichotomous variable (0 and 1).
***$p < 0.001$; **$p < 0.01$; **$p < 0.05$.

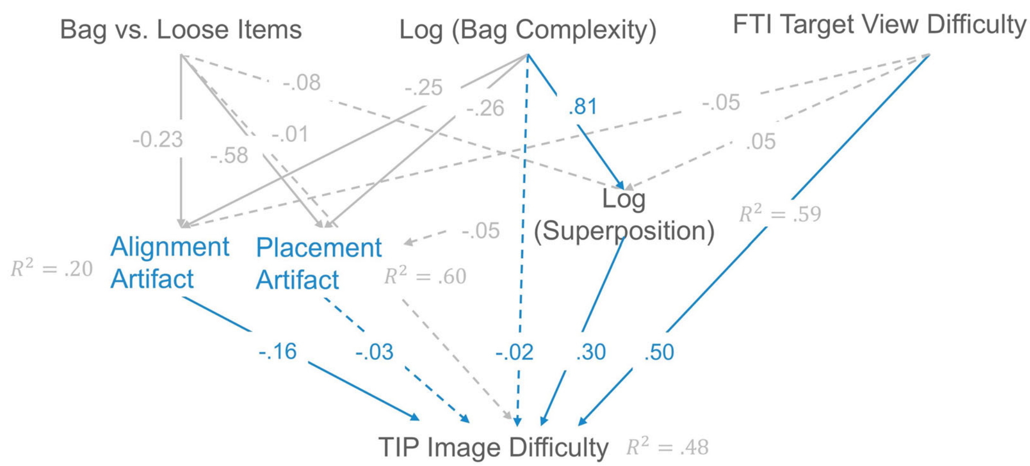

Figure 4. Results of path analysis.
Note: Standardized path coefficients and explained variance (R^2) of variables. Solid lines indicate significant relationships. Blue lines are the results of the main hypotheses and grey lines are the results of the exploratory hypotheses.

Table 3 summarizes the estimated direct effects and shows for each hypothesis whether it was confirmed. The estimates did not confirm the effect of placement artifacts on TIP image difficulty (H_{1a}; $\beta = -0.03$, $p = 0.525$), whereas the effect of alignment artifacts on TIP image difficulty was confirmed (H_{1b}; $\beta = -0.16$, $p < 0.001$), To test H_{2e}, the indirect effect of bag complexity on bag complexity over superposition was tested and found to be significant ($\beta = 0.24$, $p < 0.001$).

To illustrate how alignment artifacts affect TIP image difficulty and, more generally, the effectiveness of the TIP system, we calculated a scenario of how TIP image difficulty would change if no alignment artifacts were present. In the absence of an alignment artifact, our model indicates that TIP image difficulty increases by 6.5 percentage points (calculated as the average difference in the alignment artifact rating of 2.17 between the images with and without artifacts multiplied by the alignment artifact's effect of 0.03, see Table 3). As the alignment artifacts occurred in almost every sixth TIP image, the overall TIP miss rate would be approximately 1.1 percentage points higher (6.5 percentage points divided by 6) without artifacts, translating into an increase of the miss rate from 11.7 to 12.8% at the airport where the TIP images originated.

To further explore how artifact occurrence can be reduced, we compared the prevalence of artifacts in all the TIP images with the prevalence in TIP images of medium-to-high bag complexity, and with the prevalence in TIP images with the FTI in a piece of baggage (Figure 5(a)). We also compared the miss rates (TIP image difficulty) between the three types of images (Figure 5(b)).

4. Discussion

Many airports worldwide employ threat image projection (TIP) that projects X-ray images of prerecorded threat items (bombs, guns, knives and other threats) into randomly selected X-ray images of passenger baggage. TIP improves threat detection, monitors screener performance, and provides performance feedback (Hofer & Schwaninger, 2005; Meuter & Lacherez, 2016; Skorupski & Uchroński,

Table 3. Summary of the analyzed effects.

Type[a]	Effect	*b*	β	CI-95	*p*-Value
H1a	Placement artifact -> FTI image difficulty	0.00	−0.03	[−0.13, 0.07]	0.525
H1b	Alignment artifact -> FTI image difficulty	−0.03	−0.16	[−0.23, −0.09]	<0.001
H2a	FTI target view difficulty -> FTI image difficulty	0.90	0.50	[0.42, 0.58]	<0.001
H2b	Log (superposition) -> FTI image difficulty	0.13	0.30	[0.20, 0.40]	<0.001
H2c	Log (bag complexity) -> FTI image difficulty	−0.01	−0.02	[−0.14, 0.10]	0.759
H2d	Log (bag complexity) -> Log (superposition)	0.78	0.81	[0.76, 0.86]	<0.001
H2e	Log (bag complexity) -> Log (superposition) -> TIP image difficulty	0.10	0.24	[0.16, 0.32]	<0.001
E1	Log (superposition) -> Placement artifact	−0.13	−0.05	[−0.11, 0.02]	0.149
E2	Log (bag complexity) -> Placement artifact	−0.72	−0.26	[−0.33, −0.18]	<0.001
E3	FTI target view difficulty -> Alignment artifact	−0.51	−0.05	[−0.11, 0.00]	0.054
E4	Log (bag complexity) -> Alignment artifact	−0.55	−0.24	[−0.33, −0.16]	<0.001
E5	FTI target view difficulty -> Log (superposition)	0.20	0.05	[−0.01, 0.10]	0.087
E6	FTI_In_Bag -> Placement artifact	−1.63	−0.58	[−0.64, −0.52]	<0.001
E7	FTI_In_Bag -> Alignment artifact	−0.51	−0.23	[−0.32, −0.14]	<0.001
E8	FTI_In_Bag -> FTI image difficulty	−0.01	−0.03	[−0.12, 0.06]	0.504
E9	FTI_In_Bag -> Log (superposition)	−0.08	−0.08	[−0.14, −0.01]	0.015
	Block_2 -> Alignment artifact	0.06	0.02	[−0.07, 0.12]	0.640
	Block_3 -> Alignment artifact	0.71	0.28	[0.19, 0.37]	<0.001
	Block_4 -> Alignment artifact	0.30	0.12	[0.03, 0.20]	0.009
	Block_2 -> FTI image difficulty	0.06	0.13	[0.06, 0.19]	<0.001
	Block_3 -> FTI image difficulty	0.01	0.01	[−0.05, 0.08]	0.732
	Block_4 -> FTI image difficulty	0.05	0.11	[0.03, 0.19]	0.005
	Block_2 -> Placement artifact	0.21	0.07	[0.01, 0.13]	0.027
	Block_3 -> Placement artifact	0.94	0.30	[0.24, 0.35]	<0.001
	Block_4 -> Placement artifact	0.97	0.30	[0.25, 0.36]	<0.001

Note: [a]The effects of Types H and E are related to our hypothesis and exploratory analysis.

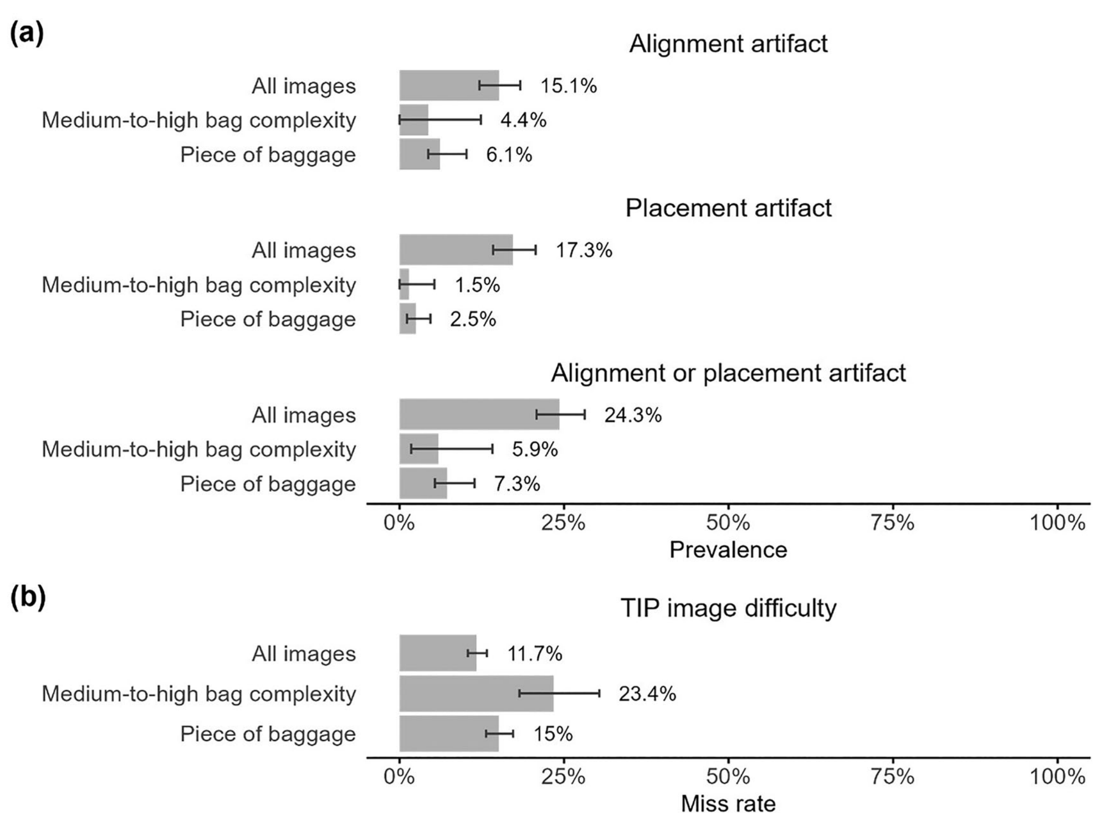

Figure 5. Prevalence of artifacts and image difficulty.
Note: Prevalence of artifacts (a) and image difficulty (b). Error bars represent 95% bias-corrected and accelerated bootstrap confidence intervals (Efron & Narasimhan, 2020).

2016). However, some TIP images appear unrealistic because of artifacts (Riz à Porta et al., 2022), which potentially undermine the effectiveness of TIP. To explore this issue and find solutions to reduce artifacts, we experimented and tested a path model that examined how artifacts influenced detection and their interaction with other image attributes.

4.1. Effects of artifacts and image-based factors on threat detection in TIP

We hypothesized that screeners would more likely detect threats in TIP images if they contain placement and alignment artifacts. Although placement artifacts did not ease detection, alignment artifacts did. This difference may stem from the distinct stages at which these artifacts influence the visual-cognitive process in X-ray image inspection. Perceiving a placement artifact requires the evaluation of whether one item penetrates another, and this requires the screener to distinguish them as two separate items. Therefore, a placement artifact may only become apparent after already detecting the FTI as a distinct object and already likely detected as a threat item. Conversely, orientation represents a basic feature (Wolfe & Horowitz, 2017) and a misaligned FTI can guide screeners' attention at the early stage of visual processing when the threat item has not been identified as a distinct object, see Wolfe and Horowitz (2017) for the discussion on basic features. When an alignment artifact occurs, the projected FTI exhibits a different orientation from those of the surrounding items. An object with a different orientation can draw attention and visually "pop out" pre-attentively (Itti & Koch, 2000; Treisman & Gelade, 1980), thereby facilitating its detection. In short, the lack of an effect from placement artifacts may reflect their detection at later visual processing stages, unlike alignment artifacts, which influence early attention allocation. Future studies could explore this distinction using eye-tracking methods (Donnelly et al., 2019).

Regarding IBFs, view difficulty and superposition increased the TIP image difficulty, as hypothesized ($H_{2a,b}$). Contrary to our hypothesis (H_{2c}), bag complexity did not directly affect TIP image difficulty. However, it affected it indirectly, confirming H_{2c}. The findings in which view difficulty and superposition increased difficulty were consistent with the results of previous studies (Bolfing et al., 2006, 2008; Hardmeier et al., 2005; Schwaninger et al., 2005, 2007, 2008). Moreover, the absence of a direct effect of bag complexity aligns with the findings of two previous studies (Bolfing et al., 2006; Schwaninger et al., 2007), whereas other studies revealed that bag complexity directly affected threat detection (Bolfing et al., 2008; Hardmeier et al., 2005; Schwaninger et al., 2005, 2008). Notably, we lacked the statistical power to determine any small effect (according to Cohen (1988)), as indicated by the confidence intervals.

4.2. Do artifacts impair the effectiveness of threat image projection?

Artifacts subject screeners to the risk of focusing on their detection rather than visually searching for threat items. However, a comparison of the overall effect of alignment artifacts on threat detection with the effect of IBFs revealed that the effect of the latter was approximately five times stronger (as indicated by the standardized effect sizes). Compared with other aspects of the TIP image, alignment artifacts, therefore, seem not to exert a dominant effect on the detection of TIP images. Furthermore, if screeners solely detected threats in TIP images because of alignment artifacts, they would only detect about one out of six threats (as about one out of six TIP images contain alignment artifacts). Regardless, the recorded detection rate at the airport from which the TIP images for this study were sourced was 88%, indicating that the detection of most of the images was guided by reasons other than alignment artifacts. These results strongly indicate that screeners are not primarily focused on detecting artifacts; rather, they are focused on visually searching for threat items.

Furthermore, artifact occurrence in TIP could compromise the validity of TIP data for monitoring screeners' threat detection performance if their hit rates are reflections of their artifact detection performance rather than genuine threat detection performance. Our results indicate that screeners' detection rates would be approximately 1 percentage point lower if TIP did not generate artifacts, implying that the overall distortion due to artifacts was small. Although only slightly affected by artifacts, our results indicate that the TIP detection rates overestimate the detection rate screeners would achieve with real threats. That is, some threats are likely easier to detect in TIP than in real-life scenarios. With TIP, they are sometimes projected into baggage of low complexity, making detection easier compared to baggage of high complexity. However, it is reasonable to assume that terrorists, in practice, would likely attempt to conceal a threat in complex baggage. When TIP images with baggage of low complexity were excluded, the detection rate was about 10 percentage points lower than when all TIP images were considered.

4.3. How to reduce threat image projection artifacts

Our results indicate that artifacts would occur less if the threat items were projected more onto pieces of baggage or images with higher complexity. For example, 6% of the TIP images of at least medium complexity contain artifacts, compared to 24% when all TIP images are considered. TIP should certainly not completely stop projecting threat items onto low complexity baggage. But reducing the frequency of TIP in low complexity baggage would produce more realistic projected threat images that cannot be easily detected. For this, a TIP system requires an algorithm that estimates bag complexity or distinguishes pieces of baggage from loose items in a tray. To project onto pieces of baggage a version of the algorithm developed by Wei et al. (2021) might be used and bag complexity could be measured using the algorithm developed by Bolfing et al. (2008). The algorithm by Bolfing et al. (2008) measures bag complexity by estimating clutter and opacity. For this, grey-scale values of the bag image serve as input and clutter is estimated by applying a high-pass filter in Fourier space whereas opacity is estimated as the relative share of pixels below a certain threshold (e.g., 64). The specific formulae can be found in the original publication. Bolfing et al. (2008) validated the complexity measures by showing that they can predict the detection rate of screeners for images of different complexity. Because the algorithm was developed and validated for one specific X-ray system, it would require calibration with airport-specific X-ray data to ensure accuracy. Validation could involve comparing algorithm outputs against screener ratings, whereby an ICC of 0.75 would be suggested (Cicchetti, 1994).

Another solution is provided by a different approach, where FTIs are pre-combined with baggage images (Hofer & Schwaninger, 2004; Wetter, 2013). These combined threat images (CTIs) can be examined before adding them to a TIP library to ensure that only TIP images without artifacts are used in operation. However, as CTIs are not projected onto baggage on the fly, they can only be used in environments where screeners cannot directly determine that the screened baggage does not correspond to that shown on the TIP image.

4.4. Limitations and future research

We employed real TIP images from an airport to investigate the effects of artifacts. This represents a limitation, as the artifacts, IBFs, and TIP image difficulty exhibit correlational association, making the estimated causal effects depend on the correctness of the assumptions of the path model. A study based on an experimental approach might validate these assumptions. Additionally, future studies could examine X-ray images obtained from various X-ray and TIP systems, as artifacts might be specific to the system producing the images. As newer machines generate multiple views of screened passenger baggage from diverse angles (Mendes et al., 2013) or three-dimensional, rotatable computed-tomography images (Hättenschwiler et al., 2019; Kishan & Prashanth, 2017; Megherbi et al., 2012, 2013; Wang et al., 2020), artifacts and their effects may differ in these systems. Our findings may not fully generalize to newer multi-view or 3D X-ray systems, warranting further research. Addressing this limitation would be particularly interesting as several recent studies have found significant advantages of 3D imaging compared to 2D imaging for threat detection (Godwin et al., 2025; Hättenschwiler et al., 2019; Latscha et al., 2025; Parker et al., 2022). Future studies could also investigate whether the perception of TIP artifacts depends on the screeners' experience, as more experienced screeners might have more skill in detecting deviations from real images, or they might get used to some of the artifacts due to repeated exposure.

4.5. Conclusion

In summary, we observed that alignment artifacts increased threat detection in TIP, whereas placement artifacts did not affect detection. As this increase is generally small, screeners may not rely on artifacts to detect threats in TIP images. Artifacts can be reduced by projecting images more frequently onto pieces of baggage with medium or high complexity, and this would increase TIP image difficulty, presenting more realistic threat scenarios.

Acknowledgments

We thank Marius Latscha for his support in participant acquisition and conducting the performance study. No funding was received.

Author contributions

CRediT: **Robin Riz à Porta**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing – original draft; **Yanik Sterchi**: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing; **Adrian Schwaninger**: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Robin Riz à Porta http://orcid.org/0000-0002-6393-8632
Yanik Sterchi http://orcid.org/0000-0002-7656-182X
Adrian Schwaninger http://orcid.org/0000-0001-7753-106X

Data availability statement

The data is restricted due to anonymization and security concerns, contact: yanik.sterchi@fhnw.ch.

References

Bassetti, C. (2018). Airport security contradictions: Interorganizational entanglements and changing work practices. *Ethnography*, *19*(3), 288–311. https://doi.org/10.1177/1466138117696513

Biggs, A. T., & Mitroff, S. R. (2015). Improving the efficacy of security screening tasks: A review of visual search challenges and ways to mitigate their adverse effects. *Applied Cognitive Psychology*, *29*(1), 142–148. https://doi.org/10.1002/acp.3083

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

Bolfing, A., Michel, S., & Schwaninger, A. (2006). Assessing image difficulty in x-ray screening using image processing algorithms. *Proceedings of the 2nd International Conference on Research in Air Transportation* (pp. 253–258). ICRAT. http://alt.casra.ch/publications/doc/BolMicSch2006a.pdf.

Buser, D., Sterchi, Y., & Schwaninger, A. (2020). Why stop after 20 minutes? Breaks and target prevalence in a 60-minute X-ray baggage screening task. *International Journal of Industrial Ergonomics*, *76*, 102897. https://doi.org/10.1016/j.ergon.2019.102897

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

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

Donnelly, N., Muhl-Richardson, A., Godwin, H. J., & Cave, K. R. (2019). Using eye movements to understand how security screeners search for threats in x-ray baggage. *Vision*, *3*(2), 24. https://doi.org/10.3390/vision3020024

Efron, B., & Narasimhan, B. (2020). The automatic construction of bootstrap confidence intervals. *Journal of Computational and Graphical Statistics*, *29*(3), 608–619. https://doi.org/10.1080/10618600.2020.1714633

European Commission. (2015). Commission implementing regulation (EU) 2015/1998 of 5 November 2015 laying down detailed measures for the implementation of the common basic standards on aviation security (CELEX No. 32015R1998). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015R1998

Godwin, H. J., Liversedge, S. P., Mestry, N., Dewis, H., & Donnelly, N. (2025). Time on task effects during interactive visual search. *Journal of Experimental Psychology Applied*, *31*(1), 40–57. https://doi.org/10.1037/xap0000521

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational behavior and human performance*, *16*(2), 250–279.

Godwin, H. J., Menneer, T., Cave, K. R., & Donnelly, N. (2010). Dual-target search for high and low prevalence X-ray threat targets. *Visual Cognition*, *18*(10), 1439–1463. https://doi.org/10.1080/13506285.2010.500605

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. *Proceedings - International Carnahan Conference on Security Technology* (pp. 189–192). https://doi.org/10.1109/CCST.2005.1594876

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., Mendes, M., & Schwaninger, A. (2019). Detecting bombs in X-ray images of hold baggage: 2D versus 3D imaging. *Human Factors*, *61*(2), 305–321. https://doi.org/10.1177/0018720818799215

Hofer, F., & Schwaninger, A. (2004). Reliable and valid measures of threat detection performance in x-ray screening. *38th Annual 2004 International Carnahan Conference on Security Technology, November 2004* (pp. 303–308). ICCST. https://doi.org/10.1109/CCST.2004.1405409

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

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Humphrey, S. E., Nahrgang, J. D., & Morgeson, F. P. (2007). Integrating motivational, social, and contextual work design features: A meta-analytic summary and theoretical extension of the work design literature. *Journal of applied psychology*, *92*(5), 1332.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506. https://doi.org/10.1016/s0042-6989(99)00163-7

Kishan, K. K., & Prashanth, K. V. M. (2017). Techniques for Detecting and Tracking of Baggages in Airports. *Proceedings - 2017 International Conference on Recent Advances in Electronics and Communication Technology, ICRAECT 2017*, 333–338. https://doi.org/10.1109/ICRAECT.2017.78

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

Latscha, M., Merks, S., Sauer, J., & Schwaninger, A. (2025). Detecting bombs in hold baggage: 3D imaging is better than 2D imaging when image quality is the same. *Applied Ergonomics*, *128*, 104534. https://doi.org/10.1016/j.apergo.2025.104534

Liu, X. (2016). Traditional methods of longitudinal data analysis. In X. Liu (Ed.), *Methods and applications of longitudinal data analysis* (pp. 19–59). Academic Press. https://doi.org/10.1016/B978-0-12-801342-7.00002-2

Megherbi, N., Breckon, T. P., Flitton, G. T., & Mouton, A. (2013, October). Radon transform based automatic metal artefacts generation for 3D threat image projection. In *Optics and photonics for counterterrorism, crime fighting and defence IX; and optical materials and biomaterials in security and defence systems technology X* (Vol. 8901, pp. 94–102). SPIE.

Megherbi, N., Breckon, T. P., Flitton, G. T., & Mouton, A. (2012). Fully automatic 3D threat image projection: Application to densely cluttered 3D computed tomography baggage images. *2012 3rd International Conference on Image Processing Theory, Tools and Applications, IPTA 2012* (pp. 153–159). IEEE. https://doi.org/10.1109/IPTA.2012.6469523

Mendes, M., Schwaninger, A., & Michel, S. (2013, October). Can laptops be left inside passenger bags if motion imaging is used in X-ray security screening? *Frontiers in Human Neuroscience*, *7*, 654. https://doi.org/10.3389/fnhum.2013.00654

Meuter, R. F. I., & 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

Parker, M. G., Muhl-Richardson, A., & Davis, G. (2022). Enhanced threat detection in three dimensions: An image-matched comparison of computed tomography and dual-view X-ray baggage screening. *Applied Ergonomics*, *105*, 103834. https://doi.org/10.1016/j.apergo.2022.103834

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Riz à Porta, R., Sterchi, Y., & Schwaninger, A. (2022). How realistic is threat image projection for x-ray baggage screening? *Sensors*, *22*(6), 2220. https://doi.org/10.3390/s22062220

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage Publications, Inc.

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

Schwaninger, A., Bolfing, A., Halbherr, T., Helman, S., Belyavin, A., & Hay, L. (2008). The impact of image based factors and training on threat detection performance in X-ray screening. *Proceedings of the 3rd International Conference on Research in Air Transportation, October 2015* (pp. 317–324). ICRAT. http://www.casra.ch/uploads/tx_tvpublications/SchBolHalHelBelHay2008_02.pdf.

Schwaninger, A., Hardmeier, D., & Hofer, F. (2005). Aviation security screeners visual abilities & visual knowledge measurement. *IEEE Aerospace and Electronic Systems*, *20*(6), 29–35. https://doi.org/10.1109/MAES.2005.1412124

Schwaninger, A., Michel, S., & Bolfing, A. (2007). A statistical approach for image difficulty estimation in x-ray screening using image measurements. *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization*. ACM Press.

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

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Wang, Q., Megherbi, N., & Breckon, T. P. (2020). A reference architecture for plausible threat image projection (TIP) within 3D X-ray computed tomography volumes. *Journal of X-Ray Science and Technology*, *28*(3), 507–526. https://doi.org/10.3233/XST-200654

Wei, Y., Zhu, Z., Yu, H., & Zhang, W. (2021). A real-time Threat Image Projection (TIP) model base on deep learning for X-ray baggage inspection. *Physics Letters, Section A: General, Atomic and Solid State Physics*, *400*, 127306. https://doi.org/10.1016/j.physleta.2021.127306

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

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

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 0058. https://doi.org/10.1038/s41562-017-0058

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

About the authors

Robin Riz à Porta is a PhD student. His research focuses on human factors, human–machine interaction and deriving actionable insights from applied research.

Yanik Sterchi is a Professor at the FHNW School of Applied Psychology. His research focuses on human–machine interaction, aviation security, and human–AI collaboration.

Adrian Schwaninger is a Professor and head of the Institute Humans in Complex Systems of the FHNW School of Applied Psychology. His research focuses on aviation security human factors and human-machine interaction.