This is the author's version of a work that was submitted/accepted for publication in the following source:


This file was downloaded from: [http://eprints.qut.edu.au/94790/](http://eprints.qut.edu.au/94790/)

© Copyright 2015 Human Factors and Ergonomics Society

**Notice:** Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:

When and why threats go undetected:

Impacts of event rate and shift length on threat detection accuracy during airport baggage screening

Renata F. I. Meuter, Queensland University of Technology (QUT), Brisbane, Queensland, Australia

Philippe Lacherez, Queensland University of Technology (QUT), Brisbane, Queensland, Australia

Word Count Text: 4,281 (excluding title, abstract, tables, acknowledgements, key points and biographies)

PRÉCIS: We investigated threat detection performance in cabin baggage screeners at a major international airport, using their detection of Fictitious Threat Items as the performance measure. Threat detection was negatively impacted by increased screening shift length and the number of bags processed per minute (event rate). Performance decrements emerged within the first 10 minutes of a busy, high event rate shift. The findings suggest shorter shift rotations would ensure high vigilance in screeners.
ABSTRACT

Objective: We aimed to assess the impact of task demands and individual characteristics on threat detection in baggage screeners.

Background: Airport security staff work under time constraints to ensure optimal threat detection. Understanding the impact of individual characteristics and task demands on performance is vital to ensure accurate threat detection.

Method: We examined threat detection in baggage screeners as a function of event rate (i.e., number of bags per minute) and time on task across 4 months. We measured performance in terms of the accuracy of detection of Fictitious Threat Items (FTIs) randomly superimposed on X-ray images of real passenger bags.

Results: Analyses of the percentage of correct FTI identifications (hits) show that longer shifts with high baggage throughput result in worse threat detection. Importantly, these significant performance decrements emerge within the first 10 minutes of these busy screening shifts only.

Conclusion: Longer shift lengths, especially when combined with high baggage throughput, increase the likelihood that threats go undetected.

Application: Shorter shift rotations, although perhaps difficult to implement during busy screening periods, would ensure more consistently high vigilance in baggage screeners and, therefore, optimal threat detection and passenger safety.

Corresponding author: Renata Meuter, School of Psychology and Counselling, Queensland University of Technology (QUT), Victoria Park Rd, Kelvin Grove campus, Kelvin Grove, Brisbane Q 4059, Australia. Ph: (61) 7-3138 4641, email: r.meuter@qut.edu.au

Running Head: When and Why Threats Go Undetected

Key words: Airport security, vigilance, threat image projection, detection performance
WHEN AND WHY THREATS GO UNDETECTED

WHEN AND WHY THREATS GO UNDETECTED:

IMPACTS OF EVENT RATE AND SHIFT LENGTH ON THREAT DETECTION

ACCURACY DURING AIRPORT BAGGAGE SCREENING

The smuggling of dangerous materials is a constant threat in our airports, carrying immediate security risks for travelers and the wider community. This threat is managed by security personnel who handle a tremendous volume of passenger throughput and work under extreme pressure, not least because of the risks associated with failing to detect a dangerous item. Vigilance or sustained attention is central to screeners’ ability to effectively monitor X-rayed passenger bags for threats that are ill-defined and unpredictable in their occurrence (Warm, Parasuraman, & Matthews, 2008). Since Mackworth’s (1948, 1950) demonstrations that, over time, we become less efficient at detecting untoward events (a phenomenon known as the vigilance decrement), research has focused on tracking the vigilance decrement in laboratory-based vigilance tasks and simulations, often using highly monotonous tasks across extended time periods (cf. Scerbo, 2001).

A debate persists regarding the underlying mechanism that accounts for the vigilance decrement (Helton & Warm, 2008; MacLean et al., 2010). Two main competing classes of theories exist, namely resource (overload) theories and mindlessness (underload) theories. Resource theory suggests that lapses occur because the need to sustain attention across time depletes already limited cognitive resources (Helton & Warm, 2008; Matthews, Warm, Reinerman-Jones, et al., 2010), a contention supported by strong correlations with perceived mental fatigue (Helton & Warm, 2008; Szalma, Hancock, Dember, & Warm, 2006) and increased detection errors with increased task difficulty (See, Howe, Warm, & Dember, 1995). Mindlessness theory holds that, when critical events are infrequent and separated by long time intervals (as occurs in monotonous tasks), observers actively disengage from the task, resulting in detection errors (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997).
They become bored, as indicated by an increase in task unrelated thoughts (Smallwood et al., 2004). Unrelated cues could mitigate against this by inducing a more mindful mode (Cheyne, Solman, Carriere, & Smilek, 2009; Dockree et al., 2004; but see Helton & Russell, 2011; Helton & Russell, 2012; Manly et al., 2004).

Although some recent laboratory-based studies favor resource theory as the explanation for the vigilance decrement (Helton & Russell, 2011, 2012), the debate is as yet unresolved (Warm, Finomore, Vidulich, & Funke, 2015; see also Thomson, Besner, & Smilek, 2015). Whether the decrement occurs similarly in workplace settings is also a subject of ongoing debate (e.g., Hancock, 2014; Mackie, 1987; Parasuraman & Giambra, 1991; Parasuraman, Warm, & Dember, 1987; Pigeau, Angus, O'Neill, & Mack, 1995). Airports internationally have endeavored to combat the potential mindlessness of baggage screening through the implementation of Threat Image Projection (TIP) technology. This technology embeds Fictitious Threat Images (FTIs) in X-ray images of some bags at random intervals and provides performance feedback (Hancock & Hart, 2002). This serves the purpose of increasing the proportion of bags with salient items (capable of capturing attention) which may reduce the potential for attention to wander, and the knowledge that performance is monitored using the inserted images (FTIs) may also make screeners more aware of possible attentional lapses (e.g., Schwaninger, 2003). The task of baggage screening, although repetitive, does offer ongoing variety: Passenger bags, when scanned, are often visually complex, and each bag differs from the previous one. The task is also cognitively challenging: In order to identify the unknown threat (and do so efficiently and accurately), screeners have to recognize it as a threat item even if it is unusually rotated or partially occluded and surrounded by other objects.

The length of time across which vigilance must be maintained is a key factor impacting performance efficiency. Guidelines in place at the airport where the data presented
herein were collected stipulated a maximum of 20 minutes continuously per screening shift (Brisbane Airport Corporation, personal communication, 27 June 2006). We wished to establish whether these guidelines ensured optimal performance. Furthermore, a high passenger volume together with the demands for optimal vigilance results in a high workload, more marked at peak travel times (Frederickson & LaPorte, 2002; Seidenstat, 2004). Although the effects of high workloads on performance in various work environments have been discussed (e.g., Bezerra & Ribeiro, 2012), to date these have not been quantified for security screeners in predicting threat detection performance. Screening at busier screening points with higher passenger volume and concomitantly great baggage throughput may engender higher stress levels, which could be better tolerated by some (e.g., Cox-Fuenzalida, Angie, Holloway, & Sohl, 2006; Matthews & Campbell, 2009).

The impact of other factors, such as the threat’s visual characteristics (Smith, Redford, Washburn, & Taglialatela, 2005), screeners’ prior and on-going training (Hardmeier, Hofer, & Schwaninger, 2006; Koller, Hardmeier, Michel, & Schwaninger, 2008; Schwaninger & Hofer, 2004), individual differences in visual ability or knowledge (Hofer & Schwaninger, 2005; McCarley, Kramer, Wickens, Vidoni, & Boot, 2004; Smith et al., 2005), and age and gender characteristics (e.g., Riegelnig & Schwaninger, 2006; Schwaninger, Hardmeier, Riegelnig, & Martin, 2010), has been described. Regarding gender, Riegelnig and Schwaninger (2006), for example, found a higher proportion of hits (correct FTI identifications) for female screeners. Higher FTI hit rates are believed to reflect better detection rates of real threat items but female screeners also showed higher false alarm rates, which could be interpreted as a leniency in responding to avoid the risk of passing a dangerous item undetected through the screening process.

Screeners typically range in age from young adult to late middle age or older. Age-related cognitive slowing occurs in various cognitive abilities essential in baggage screening,
such as vigilance or sustained attention and working memory (Salthouse, 2003), but not procedural and semantic memory (Craik & Bosman, 1992). Some aspects of perception are negatively impacted also, including mental rotation (Morrow & Schriver, 2007). Indeed, Morrow and Schriver (2007) showed better performance in males overall, a difference that tended to persist with age. Schwaninger et al. (2010) demonstrated that older adults with many years of job-specific screening experience nonetheless showed a marked performance decline.

We asked whether the progressive performance decrement across time seen in the laboratory occurs similarly in baggage screening, a real-world vigilance task that is characterized by safety-critical decisions frequently made under high pressure to ensure both passenger safety and optimal passenger throughput. We did so by analyzing screener performance data, collected at passenger screening points as part of a standard performance efficiency tracking and training protocol at a major Australian international airport. We wished to determine the impact of age, gender, work status (full-time, part-time, or casual), time on task (shift length), and event rate (measured as number of bags per minute) on screeners’ performance. All these aspects are important to quantify because, to ensure screeners can maintain optimal vigilance, clear empirical data are required on the optimal workloads and shift lengths. Because our data were obtained by the airport and not in controlled laboratory conditions, shift lengths varied, as did the number of shifts each individual screener experienced in the 4 months we analyzed. Such performance data provide a rich source of information on actual screener performance in the workplace.
WHEN AND WHY THREATS GO UNDETECTED

METHODS

Participants

During the 4-month data collection period, 170 security screeners worked on the cabin baggage X-ray screening machines at an Australian international airport. Because the accuracies across the sample were generally very high (see below), data were analysed only for those who experienced at least 30 FTIs across this period (N=123) to enable sufficient presentations to observe misses. Employee records were interrogated on our behalf to obtain age, gender, and work status (full-time, part-time, or casual) for each participant. Table 1 shows the screeners’ demographic characteristics. Part-time screeners were on an ongoing employment contract but on a part-time basis. Casual screeners were hired on a short-term basis only.

<table>
<thead>
<tr>
<th>Demographic Characteristic</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>38</td>
<td>12.20</td>
</tr>
<tr>
<td>Gender (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Work Status (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Part-time</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>Casual</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

Materials and Procedure

The data examined herein were acquired using TIP technology (see also Hofer & Schwaninger, 2004, 2005; Schwaninger, 2003). The airport used TIP technology to monitor threat detection performance as follows. During screening, the software randomly embedded FTIs in X-ray images of some bags at a rate of one FTI per 25 bags on average and,
occasionally, also presented a fictitious image of an entire bag containing an FTI (≤ 25% of all projected images). A minimum of 2 FTIs were projected in a 20 minute period; shorter shifts did not always contain FTIs. FTIs were randomly sourced from the TIP image library. In all, 950 individual images were presented (featuring bombs 30%, knives 32%, and guns 37% in different orientations). The screeners knew that FTIs could occur and that their FTI detection performance was recorded to measure their efficiency. They pressed a marked console button only if they believed an FTI was present. A correct FTI detection within a prescribed time period (5s) constituted a hit. If the FTI was embedded in a real bag or the screener indicated incorrectly that an FTI was present, they re-screened the bag. On-screen feedback was provided about correct FTI detections, misses, and incorrect identifications. On completion of each shift, feedback about their overall performance accuracy (based solely on FTI detection) was also provided. Dense areas in the display were marked to make it easier for screeners to zoom into them. If the screener believed a real threat was present, a security protocol was triggered to identify the threat. For each screening shift, the recorded data included (1) the identity of the FTIs, (2) whether they were correctly identified, (3) time to identification, (4) shift length, (5) screeners’ identification numbers, and (6) the event rate for the shift. We examined 36,920 FTI presentations across 4 months (archival airport data), sampling across 13,787 screening shifts across several screening points. For a subset of shifts (N = 5,094 shifts comprising 20,245 FTIs, and collected across 123 screeners), there was sufficient information to determine the number of bags seen per shift. The analyses pertaining to these shifts are reported here. Log-on and log-off times, recorded for each shift, were used to establish shift length. Performance was assessed as a function of the length of time screeners were stationed at the X-ray screen before the first FTI appeared and the event rate during the session. We considered performance in terms of the percentage of correct detections (measured as accuracy) and response time (the time taken to signal that an FTI was
detected). Because of technology limitations at the airport, false alarms could not be accurately measured. Specifically, it was not possible to verify whether a true threat may have existed in any of the physical bags. A positive response in the absence of FTI may either have been a false alarm or a real threat not captured by the system. A positive response to an inserted image (indexed as a hit), however, is not in question, as this is a measurable response to a verified image. For this reason, we focused our analysis on hits.

Data Analysis

To examine the effects of age, gender, and work status on accuracy and response time (RT), parametric and non-parametric between-groups analyses and correlation analyses were performed. Both accuracy and RT were highly skewed because of general high levels of performance and were therefore analysed non-parametrically. Because the standards for performance of the TIP training require that the participant consult the display for no longer than 5 seconds, only trials less than or equal to 5 seconds in duration were analysed. To examine the effects of event rate and time on task, we employed Generalized Estimating Equations (GEE), with each individual trial as a case and screener identity as a random factor. Such an analysis permits modelling of continuous covariates (here event rate and time on task) without the need to group and average responses, therefore improving sensitivity and allowing better modelling of the effects. Additionally, GEE permit unbalanced designs, and therefore do not require all participants to have had similar shifts. As recommended (Ballinger, 2004), models were compared using several covariance structures (independent, autogressive, and exchangeable) and the model providing the best fit (according to the Quasi Akaike Information Criterion) was interpreted.
RESULTS

Overall accuracy

Overall, there was considerable variation in overall accuracy (of all FTI presentations; percentage of correctly identified FTIs). The mean accuracy was 93.5% (SD = 4.27%). However, accuracies for some screeners were low, between 80 - 90% (N = 17); one screener had only a 71% accuracy.

Relationships between overall accuracy and the screeners’ age and gender were examined using non-parametric tests to account for the skewness of the distribution. A Spearman rank-order correlation revealed no significant association between age and the overall accuracy (ρ = -.026, p = .711). Men (M = 93.53%), and women (M = 93.22%) did not differ in terms of their accuracy overall (Wilcoxon rank-sum test W = 4627, p = .658). Full-time (M = 93.6%), part-time (M = 93.4%) and casual employees (M = 93%) also appeared equally accurate overall (Kruskal-Wallis \( \chi^2 \) (2) = .352, p = .839). See Table 2 for a summary of the data.

Relationships of response time and accuracy

Mean RTs showed considerably less variation in overall performance. With invalid trials (those > 5 sec in duration) removed and all trials averaged for each screener, no screener had a trimmed mean RT greater than 2s (mean RT = 1.18s, SD = .22). Overall accuracy correlated with RT (Spearman’s ρ = -.295, p = .001). On average slower screeners were less accurate than faster screeners.
TABLE 2: Mean ± Standard Deviation accuracy (percent correct detection) and RT (seconds) as a function of gender and work status

<table>
<thead>
<tr>
<th>Gender</th>
<th>Accuracy</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>93.53 ± 4.87%</td>
<td>1.14 ± 0.23</td>
</tr>
<tr>
<td>Male</td>
<td>93.22 ± 5.11%</td>
<td>1.21 ± 0.22</td>
</tr>
<tr>
<td>Work Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Casual</td>
<td>92.99 ± 5.31%</td>
<td>1.27 ± 0.23</td>
</tr>
<tr>
<td>Full-time</td>
<td>93.60 ± 3.43%</td>
<td>1.14 ± 0.16</td>
</tr>
<tr>
<td>Part-time</td>
<td>93.38 ± 5.15%</td>
<td>1.17 ± 0.23</td>
</tr>
</tbody>
</table>

Effects of shift length and event rate

The accuracy of identification for each FTI was examined using a two-way Generalised Estimating Equation with the time from log-on to the present FTI presentation (time to the FTI in question), and event rate during the session as factors, screener as a random variable, and accuracy as the criterion. There was no significant main effect of time-to-FTI, χ²(1) = 2.488, p = .115. However, there was a significant event rate main effect, χ²(1) = 4.999, p = .025, and a significant event rate x time-to-FTI interaction, χ²(1) = 4.715, p = .030 (see Figure 1). To visualize the continuous interaction, Figure 1 was constructed by splitting all trials into those occurring early (below the median of 8.33 min) and late in the shift (above the median), and those which occurred within a shift characterized by high (above the median of 5.4 bags per minute) or low event rate (below the median). As Figure 1 shows, when the number of bags seen was high, performance greatly deteriorated as the shift continued. By contrast, when the number of bags was low, the performance decrement was markedly less.
When and Why Threats Go Undetected

Figure 1. Interactive effects on performance accuracy (percent correct) of time at which FTI appeared on the basis of shift time (early or late; less than or greater than 8.3 minutes) and event rate (low or high; less than or greater than 5.4 bags per minute during the shift).

To follow up the performance deterioration which apparently occurred selectively in sessions characterised by high event rate, an analysis was conducted to examine the effect on accuracy of time-to-FTI only for shifts characterised by a high event rate (above the median), and for shifts shorter than 30 minutes (to remove the possible influence of a few cases exceeding 30 minutes). To examine whether performance declined at a steady rate or exponentially over time, three effects were compared in the same GEE: (1) a linear term, representing the time between the start of a shift (log-on) and the FTI, (2) a quadratic term, calculated as the square of the time between log-on and the FTI, and (3) a cubic term, calculated as the cube of the time between log-on and the FTI. A stepwise analysis, removing each non-significant term until only significant terms remained, revealed a non-significant linear component, $\chi^2(1) = 1.610$, $p = .205$, and a non-significant cubic term, $\chi^2(1) = 1.608$, $p = .205$. The quadratic component alone remained significant, $\chi^2(1) = 12.998$, $p < .001$. Figure 2 shows the exponential effect of increasing time between log-on and FTI presentation on accuracy. There is a pronounced curvilinearity such that, initially, performance declines only slightly from the very first trials for perhaps the first third of the shift, and then begins to decline at an increasing rate. Notably, performance begins to drop as early as 10 minutes into the shift, and drops rapidly after this time.

[Insert Figure 2 about here – Figure heading below]

*************

[Insert Figure 2 about here – Figure heading below]

*************
WHEN AND WHY THREATS GO UNDETECTED

Figure 2. Fitted quadratic trend for the GEE model for accuracy (percent correct) as a function of time on task for high workload shifts. Performance declined exponentially with increasing time on task.

DISCUSSION

The present study demonstrates that, for baggage screeners at X-ray checkpoints, shifts which were characterized by greater busyness (i.e., a greater number of bags seen per minute) were associated with a reduction in performance over time, and that during busy periods such performance decrement could be observed as early as 10 minutes into a shift. This pattern suggests that shift times for X-ray baggage screeners may need to be rotated faster under high passenger throughput conditions to reduce fatigue effects and attentional lapses. Our data suggest a strong cumulative effect of the number of bags screened, such that screeners can maintain vigilance over longer periods when fewer bags go through per minute (low event rate); however, performance declines more quickly when the event rate is high.

This variable performance pattern, dependent on baggage throughput, suggests that the screeners’ workload is the primary determinant of lapses of attention at baggage screening points. Importantly, as the data show (see Figure 1), the drop in performance is only noticeable under conditions of high workload. This effect of higher workload - caused by an increased number of bags needing to be thoroughly checked faster – is consistent with increased task demand producing mental fatigue and resulting in the early vigilance decrement. The pattern of results we observed therefore further supports the argument that, consistent with numerous laboratory-based findings (e.g., Szalma et al., 2004; Warm et al., 2008), the resource demand associated with the screening task (especially acute at busy times) accounts for the vigilance decrement (see also Helton & Russell, 2011; 2012; 2015).

Importantly, we observed this pattern in performance data obtained in a real world screening environment. The early appearance of the vigilance decrement when the event rate is high is
WHEN AND WHY THREATS GO UNDETECTED

less easily reconciled with an underload account, specifically why the decrement occurs only
in higher workload settings. Not only is the screening task less monotonous through the use
of FTIs (Hancock & Hart, 2002), it is unlikely that screeners are habituated or bored so early
into the task. It is worth noting that, in this study, workload was operationalized as the total
number of bags screened in the work session. It may be informative in future studies to
examine in finer detail whether performance at different times during a session (e.g., the most
busy moment compared to the least busy moment within a session) showed similar changes.
Our data did not allow for such an analysis, however they did clearly show that it was during
busier shifts that performance decrements over time were observable.

Airport security staff generally perform at a high level, as is also reflected in our data.
However, even small declines in accuracy (indeed, a single missed threat) could be
catastrophic if the attentional lapse occurred when a dangerous item was smuggled through.
Our data suggest that increases in either passenger throughput, or fatigue beyond the limits
we observed, could be associated with significant reductions in performance if the trends
continued. A combination of higher workload and longer shifts would likely increase
operator stress, thus increasing the negative impact on performance. This prediction is
consistent with the finding that a visual vigilance task performance was negatively affected
by increased workload and length of the task (Szalma et al., 2004) and with reports of
feelings of stress and mental fatigue after carrying out a vigilance task (Szalma et al., 2006).

There were large individual differences in performance accuracy, with a few
individuals contributing a considerable number of errors. Overall, those with higher error
rates were also generally slower to respond. This finding is consistent with previous research
indicating that individual differences may be the primary determinant of performance on
these baggage screening tasks (Hardmeier et al., 2006; Schwaninger, Hardmeier, & Hofer,
2005). Such a pattern might be symptomatic of either a lack of confidence or a lack of ability
in the task and as such, slow responders, as well as those with abnormally low accuracies should be carefully screened to ensure their competence to carry out screening duties. It is worth exploring whether the deleterious effects of work load shown here interact with other personality dimensions, such that individuals with inherently lower ability may be more distractible or fatigue more quickly than more able individuals (see also Szalma, 2009).

No overall gender differences were found in relation to the accuracy of FTI identifications. This broad accuracy measure also showed no effects of age or work status (for example, see Matthews, Warm, Shaw, & Finomore, 2010; Reinerman-Jones, Matthews, Langheim, & Warm, 2011). It is conceivable that differences do obtain when false alarm data can be considered. Because the TIP program coded false alarms exclusively in relation to incorrect FTI identification and no data were available to verify that a false alarm was not, at the same time, linked with the correct identification of an actual threat or a screener’s signaled suspicion that a threat might be present, we are not in a position to address this issue conclusively. Future research of on-site baggage screening should attempt to collect contemporaneous data relating to actual threat detection and baggage re-check occurrences.

Here we examined accuracy solely as a function of workload (conceptualized as the number of bags screened per minute) but other factors in the screening environment could also significantly impact performance. For example, the predominantly speech-based background noise (some directly relevant to their task) could be distracting (Szalma & Hancock, 2011), and screeners are also sometimes distracted by passengers and other staff. Some authors have suggested that auditory distraction can affect performance of a visual task (e.g., Jones, 1995; Schaefer, Tewes, Münte, & Johannes, 2006; but see also Guerreiro & van Gerven, 2011). Perceived stress can also impact negatively (e.g., Hancock & Warm, 2003) but it is largely unknown how performance is impacted by the combination of background noise and the perceived pressure not to miss any real or fictitious threats.
WHEN AND WHY THREATS GO UNDETECTED

FUTURE DIRECTIONS

Our data consisted only of performance on FTIs, shift lengths, age, gender, and work status of screeners. Several variables were unknown and, had they been available, would have allowed further exploration of the data. For example, how much training individuals received to prepare them for their screening responsibilities was unknown, and neither did we know whether screeners (especially those with poor performance) received additional training during the 4-month period we analysed. Such data, when combined with FTI detection, would yield a more comprehensive picture of performance.

Much research has focused on selection criteria for security positions (e.g., Hofer, Hardmeier, & Schwaninger, 2006; see also Matthews et al., 2010; Reinerman-Jones et al., 2011) with the express objective to establish those individual characteristics that make for the ideal screening operator, however much of this research has been confined to laboratory-based studies. Given the variable screening environment (characterised by difference in busyness at screening points) and the impact of individual characteristics (such as extraversion/introversion) on one’s ability to perform vigilance tasks (Cox-Fuenzalida et al., 2006), it would be valuable to determine if the impact of these individual characteristics generalizes to real world security screening settings such as the operational settings that generated our data.

Importantly, our data showed a marked and clear negative impact on screening performance the busier the shift, and a measurable decline in performance accuracy that not only starts early (within 10 min) but accelerates with time. Operational changes (such as shorter shift lengths and paced baggage processing) could mitigate against this performance decline and the associated performance data would provide a valuable source for further investigation. As noted above, it was not possible in this study to distinguish between false alarms and real threat items correctly identified in passengers’ bags. It would be valuable to
be able to link FTI performance data with more detailed information (obtained online and on site) regarding the nature of the false alarms. In so doing, more detailed performance analyses would allow further exploration of the nature of the performance decrement in the screening setting, in particular of the impact of overload on the allocation of attentional resources and strategic behavior.

SUMMARY AND IMPACT ON INDUSTRY

Shift times for baggage screeners should be computed as a function of the workload screeners face at any given time, rather than remaining fixed at recommended time limits. Moreover, shift lengths should be flexible enough to enable screeners to rotate faster as demand at a particular screening point increases. Doing so is likely to ensure high and stable performance across shifts. It is worth investigating whether the same patterns occur in other security settings where time on task and workload both vary according to situational factors. Airports should weigh up the need to ensure fast passenger throughput to meet departure schedules against the increased risk of employing poorly trained or less experienced screeners. The performance patterns we tracked provide further support for targeted personnel selection and training as some of the key measures required to safeguard airport security.

ACKNOWLEDGEMENTS

We thank the staff at Brisbane Airport Corporation for their assistance in providing the data. We further acknowledge the research assistance of Melanie White. This project was funded by the Brisbane Airport Corporation and additional financial support was provided by the Institute of Health and Biomedical Innovation (IHBI), at QUT.
SUPPLEMENTARY MATERIAL

The data were obtained from the Brisbane Airport Corporation (BAC) by special request and access is privileged.

KEY POINTS

- We examined threat detection performance in baggage screeners as a function of shift length and event rate (number of bags per minute).
- Performance was measured in terms of the accuracy of the detection of Fictitious Threat Items (FTI).
- Our analyses of real-world screening data reveals poorer threat detection on longer shifts with a higher workload (i.e., greater bag throughput) occurring as early as 10 minutes into the session.
- Optimal threat detection (and therefore passenger safety) requires shorter shift rotations, especially at busy times.
REFERENCES


WHEN AND WHY THREATS GO UNDETECTED


WHEN AND WHY THREATS GO UNDETECTED


WHEN AND WHY THREATS GO UNDETECTED


WHEN AND WHY THREATS GO UNDETECTED


WHEN AND WHY THREATS GO UNDETECTED


WHEN AND WHY THREATS GO UNDETECTED


BIOGRAPHIES

Renata Meuter (D. Phil., 1995, University of Oxford) is an Associate Professor in the School of Psychology and Counselling and a member of the Institute of Health and Biomedical Innovation (IHBI), Queensland University of Technology (QUT), Brisbane, Australia. Her research focuses on factors determining task performance in different operational settings affected by monotony (such as driving and baggage screening) as well as the processes underlying language selection and cognitive control in multilingual speakers.

Philippe Lacherez (PhD, 2008, University of Queensland) is a lecturer in the School of Psychology and Counselling and a member of the Institute of Health and Biomedical Innovation (IHBI), Queensland University of Technology (QUT), Brisbane, Australia, specializing in perceptual and human factors psychology. His research involves assessment of cognitive and visual function among older drivers, use of in-vehicle devices, as well as sonification and auditory displays.