Two-component model of security inspection: application and findings

K.M. Ghylin[a], C.G. Drury[a], A. Schwaninger[b]

[a] *Department of Industrial and Systems Engineering,, University of Buffalo, Buffalo, NY USA*
[b] *Department of Psychology, University of Zurich, Zurich, Switzerland*

Abstract

Security inspection is a form of visual inspection that presents a changing set of potential targets, i.e. threats. However it has not been shown how the unique features of security inspection fit into more general models of industrial and other inspection tasks. Spitz and Drury [2] developed equations based on modeling the inspection task as containing both search and decision components. The current research tested the applicability of this two-component model to x-ray screening, and utilized the model to determine how a training program (ie. X-Ray Tutor) affected the search and decision parameters. Performance data from 416 airport screeners evaluating 40 x-ray images of passenger bags was analyzed. The analysis found model vs. data r^2 values averaging above 0.8, showing the applicability of the model to a security x-ray task. [Note: Due to security reasons, all detection measures have been multiplied by an arbitrary constant.]

Keywords: security inspection, x-ray, training, inspection model, screening

1. Introduction

Transportation security has become increasingly important in recent years with numerous events sparking new found-interest in the field. Currently only a few theories or bodies of knowledge specifically applicable to security inspection exist. If the unique features of security inspection can fit into the more general inspection and visual search models, then potential new information may quickly become available to researchers tasked with improving security performance. However, few studies have been conducted directly investigating human performance of security screening (see [3, 4]).

Portions of the aviation security inspection task can be seen as being similar to a general inspection task [5]. Both can rely on repetitive visual search for multiple target items on multiple images, and the classification of the targets located as either being acceptable or requiring further investigation. However, security inspection differs from some other inspection tasks as it requires inspectors to continually recognize a changing catalogue of potential threats whose general nature is known (i.e. guns, knives, etc.), but whose actual size, shape, and characteristics change over time. Unlike industrial inspection, threats may be deliberately concealed requiring inspectors to have a high skill level for searching images. The occurrence of a threat is very rare, however, the cost of a missed threat is very high, potentially leading to mass casualties and other catastrophic events.

Drury [6] applied a general model of inspection to security using the functions found within the security inspection process. Five basic functions that can occur within any inspection, including security inspection: 1) set-up, 2) present, 3) search, 4) decision and 5) respond. Within the confines of this model, search and decision are traditionally the two areas with the highest occurrence of errors as they rely on the inspectors' individual abilities.

Their error types are unique with a typical search error being missing a threat, and a typical decision error being misclassifying a target image. By relating the proportion of reaction time to the individual search and decision components, we can further clarify why each sub-component fails in certain situations and thus better understand how to improve inspector performance.

Visual search models provide a framework for the general search process in security inspection. Visual search for inspection tasks includes eye movements and can take from several seconds to several minutes to complete [7]. The search process can terminate in finding a part of the visual field that could be a defect, or in deciding not to continue searching. The search component is typically modeled as either a process with perfect memory, i.e. completely systematic search, or with zero memory, i.e. completely random search [8].

The specific task of x-ray screening requires people to recognize various objects within a finite set of categories (i.e. guns, knives, improvised explosive devices, etc.). Once a salient area is identified, screeners are required to recognize the object as one of a class of threat objects and match the perceptual representation of the image to one stored in memory [9]. Object recognition is highly dependent on perceptual experience and training [10], which is of particular importance in x-ray screening [3, 11]. The shift in perception from novices to experts includes the benefits of recognizing common attributes across categories [12] and increasing the speed of performance [9]. This turns the decision process into a memory retrieval process and aids the search process by increasing the saliency of known exemplars. An effective employee training program to recognize the threats is a simple way to provide the opportunity to store more exemplars, thus allowing screeners to recognize objects faster and infer more information about general categories of items presented.

Along with training, past inspection research has considered influences of age, gender and on the job experience as part of the individual differences between inspectors. Age is generally seen as having a negative effect on inspection performance [13, 14, 15]. However, Czaja and Drury [13] showed that an active training program could overcome much of the age deficit. Studies of gender have found mixed results [14, 16] and affects of job experience have also been mixed [3, 17]. How these findings from inspection generalize to the security inspection task has yet to be accurately determined and modeled, especially in relation to the two components of search and decision within the overall security inspection task. It is well known that a speed-accuracy tradeoff (SATO) will occur within any inspection process [1, 7]. This trade-off can be either resources limited or data limited. In a resource-limited process, performance is dependent on the number of available cognitive resources, resulting in an increase in performance with more time on task. Errors would be related to not identifying a threat component within the image, i.e. search failure. In contrast, the inspection process can also be data-limited with no improvement in performance over time due to the limited information being presented. Errors would be equivalent to making the wrong decision given the present information. However, analysis of many inspection tasks [1] has fitted a performance/time curve having the properties of both a resource-limited and data-limited process similar to the curves seen in Figure 1. [See 8]. The curved portion represents the resource limitation, while the asymptote of the curve represents the data-limitation.

In comparison of the Unified Model and SATO theories we find that both the search component and resource-limited processes have errors of missing a threat in an image. Similarly, the decision component and data-limited processes both have errors of misclassifying a threat. Thus we can further identify the two components of the curve as representing the search and decision components within the security inspection task. To validate such a model one could either use the labor intensive experimental method of eye tracking or utilize mathematical fits of accuracy/time data based on visual search knowledge.

If search and decision information could be gathered by utilizing curve fitting instead of the traditional eye tracker, much experimental time could potentially be saved. Drury [1] and Spitz and Drury [2] developed equations that divide the total inspection time into the functional components of search- and decision-time to give an overall speed-error trade-off. The model utilizes an approximately exponential relationship between the time spent searching for a target and the cumulative probability of locating a target [8]. The search function could be modeled either by a random or a systematic search process, although inspection data are typically well-fitted by a random model [8]. Time not spent actively searching, originally called “decision time”, would be more validly labeled as non-search time.

This terminology will be utilized throughout the rest of this paper. Probability of detection (Pd_x), search time (ST_x) and non-search time (NS_x) are calculated from the model equations for both hits and false alarms.

The aim of the current paper is to test the applicability of the two-component inspection model to x-ray screening, and to help determine whether an individually adaptive computer-based training (CBT) program (X-Ray Tutor) affects the search, decision, or both functions within the x-ray screening task. The effects of demographic variables on these processes are also examined.

2. Methods

Untrained (n= 334, 234 females, mean age = 49.7) and trained (n = 82, 56 females, mean age = 51.7) screeners took part in this study[1.] All participants had on-the-job experience of airport x-ray screening. Trained screeners had completed an average of 88 (SD = 30.24) 20-minute training sessions on X-Ray Tutor HBS 1.7 (for details see [18]).

Detection performance was measured using 40 x-ray images of passenger bags, half of which contained an improvised explosive device (IED). Note that IEDs are typically the most difficult threats to detect. Stimuli were created in close collaboration with the security authorities at a major European airport. Participants had to decide whether the bag was OK (did not contain an IED) or whether it was NOT OK (i.e. it contained an IED). Stimuli were displayed on 17'' TFT screens at a distance of about 50 cm so that x-ray images subtended approximately 22 deg of visual angle. The computer program measured outcome (hit, miss, false alarm (FA), correct rejection (CR)) and the time from image onset to final decision key press.

The inspection model was applied separately to hit and false alarm data for each participant. Data were discarded for hit and false alarm responses under two conditions: (1) where participants did not respond within 14 seconds, which provided ample time for the participant to come to a decision after the image was removed. (2) Where less than five reaction times were available for either Hits or False Alarms. This helped ensure sufficient data for valid model fitting for the cumulative exponential distribution. Data elimination occurred in 4.8% of the trials (799 of the 16,640 cases). Final sample sizes were 325 (hit RT) and 302 (FA RT) for untrained participants, and 82 (hit RT) and 32 (FA RT) for trained participants.

An overall probability of hit and false alarm was calculated. Probability of detection at a given time was then calculated for both hits and false alarms utilizing the procedures found by Spitz and Drury [2]. From this output the means of search time and non-search time along with the goodness of fit (r^2) values were obtained. The r^2 values were then examined in order to test the applicability of the two-component model to the x-ray inspection data.

3. Results

Figure 1 shows the overall effects of training on both the search and non-search parameters for each training group (trained vs. untrained screeners).

Goodness of fit values (r^2) values averaging above 0.8 were obtained for both hits (*Mean* = 0.819, *SD* =0.15) and false alarms (*Mean* = 0.803, *SD* = 0.150). Note that 39% and 31%, respectively, exceeded an r^2 value of 0.9. This indicates that the two-component model is applicable to this security inspection task.

The performance measures of hit rate and false alarm rate for each participant were tabulated and inter-correlated. As expected from Signal Detection Theory, hit rate and false alarm rate correlated for

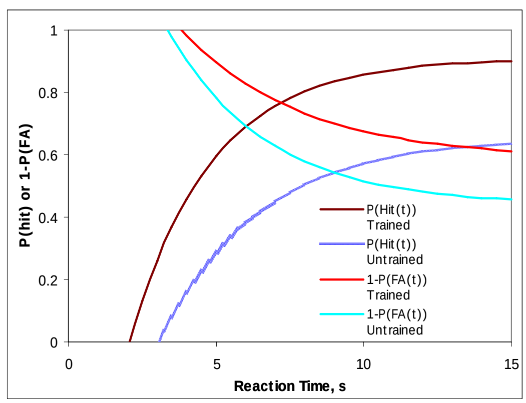

Figure 1: Overall Training Effect on Search and Non-Search Parameters (Note: Probabilites are multiplied by an arbitrary constant for security purposes)

1 Experience, measured by days on job, was utilized but is not reported here since it is security sensitive information.

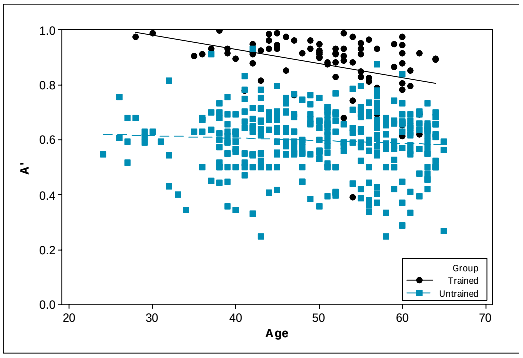

Figure 2: A' Values by Age and Training

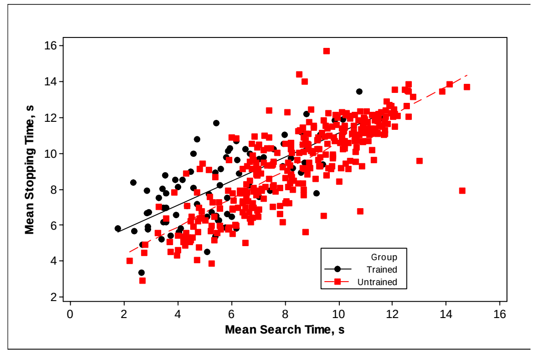

Figure 3: Search and Stopping Time by Training

both untrained and trained participants (r = 0.670, p < 0.001 and r = -0.248, p <0.05, respectively). Although a positive correlation might be expected where participants had about the same level of overall performance (e.g. A') and variation was only between participants in criterion, such trade-offs across participants are not found universally. In particular, it appears that the trained screener group showed no such trade off, implying that their overall performance was not constant.

Detection performance in terms of A' averaged 0.58 for untrained participants and 0.87 for trained participants. A' scores were subjected to a univariate ANCOVA with age, gender and days on job as covariates and training as a between-participants factor (trained vs. untrained group). The effect of training was highly significant, (F (1,407) = 411.296, p < .001, MSE = 5.118, η^2 = 0.505). The covariate effect of age was also significant, (F (1,407) = 4.465, p < .05, MSE = 0.056, η^2 =0.011) and is seen in Figure 2. No other effects were significant.

Average reaction times in seconds were calculated for hits, misses, false alarms, and correct rejections for untrained (Hit: *Mean* = 9.1, *SD* = 3.6; Miss: *Mean* = 9.5, *SD* = 2.8; False Alarms: *Mean* = 9.4, *SD* = 3.2; Correct Rejection: *Mean* = 10.0, *SD* = 3.4) and trained participants (Hits: *Mean* = 5.7, *SD* = 3.6; Miss: *Mean* = 8.9, *SD* = 2.5; False Alarms: *Mean* = 9.5, *SD* = 3.1; Correct Rejection: *Mean* = 8.4, *SD* = 3.0). Performance measures were inter-correlated across participants. All four time measures had inter-correlations above 0.68, all significant at p < 0.001. A factor analysis of the reaction times and performance measures (p(hit) and p(FA)) produced only 2 factors, one with only the performance measures of hit rate and false alarm rate and one with the three time measures, explaining 74% of the variance. Regressions of detection time (hit time) on stopping time (correct rejection time) produced significant results for both training levels and are seen in Figure 3. For untrained participants:

Stopping Time = 2.81 + 0.787 Detection Time
r^2 = 0.670, p < 0.001

For trained participants:

Stopping Time = 4.44 + 0.672 Detection Time
r^2 = 0.521, p < 0.001

In any search task we would expect that the target would be found in less time than a single scan of the whole item [8]. This can easily be seen from the expected relationship of reaction times with Detection Time being smaller than Stopping Time. For a systematic scan the stopping time would be expected to be about twice the detection time, with a somewhat higher multiplier for random search [8]. Here we found multipliers rather lower than this, presumably due to the non-search time added to each.

A highly significant training effect was found for hits (F (1, 410) = 73.47, p < 0.001, MSE = 6.07, η^2 = 0.152) and correct rejections (F (1,410) = 12.86, p < 0.001 MSE = 5.472, η^2 = 0.030). No training effects for false alarms were found. An ANCOVA of detection time with training as a factor and stopping time as a covariate found a significant effect of training (F(1, 410) = 89.01, p < 0.001, MSE = 2.2, η^2 = 0.066). Figure 5 shows these relationships. There was no significant difference between time for correct rejections and time for misses (F (1,816) = 1.95, p = 0.163, MSE = 6.52, η^2 = 0.002).

Since the assumptions of the model were examined and met, individual search and decision parameters were subjected to separate univariate

ANCOVAs with training (untrained vs. trained group) as an independent factor and age, gender and days on job as covariates. A significant main effect of training was observed for search in the hit trials, F (1,406) = 36.829, $p < .001$, MSE = 236608297.88, $\eta^2 = 0.084$. Training also affected the decision in the hit trials F (1,406) = 26.438, $p < .001$, MSE = 65084198.15, η^2 = 0.062. No effects were found for false alarm trials. In all ANCOVAs, no covariates reached significance.

4. Discussion

The two-component inspection model yielded high r^2 values, providing support for its applicability to a security inspection task. Also stopping times were a constant multiplier of detection times, as would be expected for search tasks, showing that this task has similar characteristics to other inspection tasks. Not only did the two-component model fit the data well, but both search and non-search parameters in the model were found sensitive to training.

A large effect on A' was found indicating that the overall performance of participants increased with training. This agrees with the findings previously reported by Schwaninger & Hofer [11]. Age was a determinant of A'; which is similar to results found on age affects in other inspection studies [13, 14, 15]. It should be noted however that the training effects (η^2 = 0.505) were 50 times larger than the age affects (η^2= 0.011). Age is often debated in industry with older inspectors being looked at for replacement due to a decrease in visual acuity and other cognitive abilities [13]. However, according to our research, proper training can more than compensate for any cognitive degradation that occurs with age, at least up to the age limit of this study (65 years). Thus, with proper training, age affects are no longer a major issue, further confirming Czaja and Drury's findings [13]. Note that no Gender effects were found, again typical of inspection research results.

A reduction in the reaction times for hits but not for false alarms or correct rejections was also found following training. The reduction for hits was found for both the search and decision components, showing that training helped decrease the time needed to both locate and identify a true threat.

For false alarms, no individual affects on either the search or non-search parameters was found. This is to be expected since the training program helps participants learn to detect threat items and not to recognize non-threat items. This can also provide further evidence for exemplar theories. The increase in probability of hit and the decrease in the hit reaction time for trained screeners, paired with the relative consistency of false alarm data for both groups, indicate that the training system is very effective. Based on most object recognition theories, our findings indicate that the training program is effective at providing more exemplars of the threat images, which are stored in visual memory. Common attributes of the threat images became easier to recognize with the training program, (c.f. [12]), resulting in increases in the speed and accuracy of performance, (c.f. [9]).

A large improvement in the decision parameter for hits was found as seen in other studies [4]. Through training, more effective visual memory representations of threats were created aiding participants in understanding whether particular features were part of a threat or not. This could be related to novice vs. expert effects (c.f. [9] & [12]) and is consistent with earlier findings of Schwaninger and Hofer [11].

Improvement in the search parameter for hits with training was also found in the current study, and is a unique output of this research. It may lead to the question of whether new feature maps can be created that help guide visual search more efficiently based on better preattentive processing.

5. Conclusions

The individually adaptive computer based training (X-Ray Tutor) was found to be effective by both dramatically increasing sensitivity measures and reducing time taken for the task. The two-component model provided evidence in extension to previous findings [11] of large increases in detection performance and substantial reductions of response time. With training, identification of items dramatically increased. A small increase in search performance for threats within an image also occurred. These initial results provide evidence that training creates more exemplars of images in visual memory to help aid in the decision process for threats, thus providing an improved ability to recognize threats. More research is needed to help explain the increase in search efficiency found through training, and the affects of training on search and decision for false alarms. However, trained participants, regardless of age, gender or days on job were able to both identify and find threat items more efficiently and effectively than their untrained counterparts.

Through this study a successful application of the two-component inspection model to x-ray inspection has been shown. It aids in discerning between the search and non-search, or decision, processes by analyzing the speed-accuracy trade off. It may allow the individual parts of the inspection process to be further analyzed and modelled to provide a better idea of what steps within the inspection processes are changing and why. By having knowledge of the individual components within inspection, then the technologies, methods and procedures that current security inspectors utilize can be better honed to enhance the strengths of the human-system interface and improve their weaknesses.

Acknowledgements

This work was made possible through the cooperation of the Transportation Security Administration (TSA) and the University of Zurich.

References

[1] Drury CG. Inspection of sheet metal materials: model and data. Human Factors, 17 (1975) 257-265.

[2] Spitz G and Drury CG. Inspection of sheet materials – test of model predictions. Human Factors, 20(5) (1978) 521-528.

[3] Schwaninger A, Hardmeier D and Hofer F. Measuring visual abilities and visual knowledge of aviation security screeners. IEEE ICCST Proceedings, 38 (2004) 258-264.

[4] McCarley JS, Kramer AF, Wickens CD, Vidoni ED and Boot WR. Visual Skills in Airport Security Screening. Psychological Science, 15(5) (2004) 302-306.

[5] Drury CG. Human Factors. In: Assessment of Technologies Deployed to Improve Aviation Security, National Research Council, National Academy Press, Washington, D.C., Report # NMAB-482-5, 2004, 48-55.

[6] Drury CG. A unified model of security inspection. In: Proceedings of the FAA's Third International Aviation Security Technology Symposium, November 27-20 (2001) Atlantic City, NJ.

[7] Drury CG. Human Factors in Test and Inspection, Chapter 71. In: Salvendy G (Ed) Handbook of Industrial Engineering, Third Edition. John Wiley & Sons, New York. 2001, pp 1887-1920.

[8] Morawski T, Drury CG and Karwan MH. Predicting search performance for multiple targets. Human Factors, 22(6, (1980) 707-718.

[9] Palmeri TJ, Wong AC-N and Gauthier I. Computational approaches to the development of perceptual expertise. TRENDS in Cognitive Science, 8(8) (2004) 378-386.

[10] Wallis GM and Bülthoff HH. Learning to recognize objects. Trends in Cognitive Sciences 3 (1999) 22-31.

[11] Schwaninger A and Hofer F. Evaluation of CBT for increasing threat detection performance in X-ray screening,. In: Morgan K and Spector MJ (Ed) The Internet Society 2004, Advances in Learning, Commerce and Security. Wessex: WIT Press, pp 147-156.

[12] Sigala N. Visual categorization and the inferior temporal cortex. Behavioural Brain Research, 149 (2004) 1-7.

[13] Czaja SJ, Drury CG. Aging and pretraining in industrial inspection. Human Factors, 23 (1981) 485-494.

[14] Dollinger SM and Hoyer WJ. Age and skill differences in the processing demands of visual inspection. Applied Cognitive Psychology, 10(3), (1996) 225-239.

[15] McPhee LC, Scialfa CT, Dennis WM, Ho G and Caird JK. Age differences in visual search for traffic signs during a simulated conversation. Human Factors, 46(4) (2004) 674-685.

[16] Wang MJJ, Lin SC and Drury CG. Training for strategy in visual search. International Journal of Industrial Ergonomics, 20(2) (1997) 101-108.

[17] Wenner C, Spencer, F and Drury CG. The impact of instructions on aircraft visual inspection performance: A first look at overall results. Proceedings of the Human Factors and Ergonomics Society, Santa Monica (2003) 51-55.

[18] Schwaninger A. Computer based training: a powerful tool to the enhancement of human factors. *Aviation Security International,* FEB/2004, 31-36.