

Evaluating the two-component inspection model in a simplified luggage search task

ALAN W. J. WALES

*Loughborough University, Loughborough, England
and University of Zurich, Zurich, Switzerland*

CLARE ANDERSON AND KATHERINE L. JONES

Loughborough University, Loughborough, England

ADRIAN SCHWANINGER

University of Zurich, Zurich, Switzerland

AND

JAMES A. HORNE

Loughborough University, Loughborough, England

Visual inspection of X-ray images of luggage is a time-pressured task that typically shows large initial training effects, but there exists a paucity of models capable of evaluating performance and speed concurrently. In the present study, visual inspection ability during learning was modeled using Drury's two-component inspection model (TCM; Drury, 1975) in a laboratory experiment involving 12 younger (mean age = 20.8 years) and 12 older (mean age = 60.0 years) naive participants undertaking a simplified luggage search task. Model fits and assumptions were found to be reliable and accurately reflected improvement with training for decision time, although neither search nor decision components of the model individually showed a significant effect of age. The decision component of the model showed larger improvement with training than did the search component, and stopping-time policy accurately reflected the improvements found between ages and within training levels. The TCM is a useful supplement to detection theory when speed of performance is a factor.

Speed and accuracy are two readily measured components of visual search tasks that are interlinked but are often analyzed separately, because of difficulties in conjoining both through a valid theoretical model (Ratcliff & Rouder, 1998). Many paradigms throughout the visual and inspection field yield high accuracy values, meaning that experimental manipulations can often be witnessed only in response-time changes (Ratcliff, 2001), particularly when tasks are participant paced. Resultant data are thereby split into inferential tests of speed or accuracy, with no clear or meaningful method of evaluating both simultaneously.

Visual inspection of X-rayed airline passenger luggage is a complex task that has implications for passenger security and safety; it is also capable of providing realistic simulation in the laboratory. Studies have shown that computerized luggage search tasks are predictive of on-the-job threat-item detection (Schwaninger, Hofer, & Wetter, 2007) and are highly sensitive to training (Bolfing, Halbherr, & Schwaninger, 2008; Koller, Hardmeier, Michel, & Schwaninger, 2007; Schwaninger et al., 2008). However, the mediating effect of time taken to inspect each item is

rarely evaluated as part of the performance metric. Only one prior study (Ghylin, Drury, & Schwaninger, 2006) has used an inspection model for luggage search, where professional screeners were shown 40 images that used improvised explosive devices as targets.

This study sought to apply Drury's two-component inspection model (TCM; Drury, 1975) to data from a sample of younger and older participants in order to evaluate speed-accuracy learning curves in a simple luggage search task, and to detail how this model is fit to two-choice data. Conceptual comparisons are drawn to sequential sampling models (SSMs), which share a similar goal of resolving speed and accuracy measures of detection theory (DT) and have many similarities with the TCM. The TCM was designed by ergonomists to analyze real-world data, whereas SSMs are rooted mainly in theory and are not suitable for use with stimuli that require more than 1 sec to respond to. However, the rich SSM literature is a valuable adjunct to the TCM literature when evaluating the complications that arise from combining speed and detection measures into a performance model.

Detection Theory

DT (see Macmillan & Creelman, 1991) is widely used throughout the visual search literature, providing two major indexes: a detection measure of sensitivity (i.e., d' , A') and a bias measure (i.e., c , B''). The detection measure is a function of responses to items that contain threats ("hit" or "miss"), and those that do not contain a threat ("false alarm" or "correct rejection"). *Bias* represents the tendency of a participant to favor either response between hits or false alarms, so that an anxious participant may require only a small amount of uncertainty to select "threat item present." Performance improvement is shown as an increase in the detection measure value, although changes in bias may not always result in a change in detection performance. Confidence ratings can also be used in DT to generate theoretical operating curves known as *receiver-operating characteristic curves*. Taken together, a detection and a bias measure are sufficient to summarize performance in purely accuracy terms for two-choice tasks, such as luggage search.

However, DT is limited by its inability to account for speed (or its reciprocal, *time*) by assuming a fixed sampling interval (Smith, 2000). Using DT alone neglects the influence of speed measures on performance; Drury (1994) stated that "speed measures are so basic to performance evaluation that the need for them is rarely discussed" (p. 748). To account for reaction time (RT) in a unified model, SSMs (Wald, 1947)—in particular, random walk and diffusion theory (Link & Heath, 1975; Ratcliff, 1978)—were created to account for this major limitation, although confidence ratings are notable by their absence in both SSMs (Pleskac & Busemeyer, 2007) and Drury's models.

Drury's Models

Drury's TCM has been used widely throughout ergonomics to estimate model parameters for inspection tasks by combining visual search and decision theory. The model predicts that, as the probability of a hit increases with time allowed to inspect an item, the probability of a false alarm increases concomitantly—although not necessarily proportionally (Spitz & Drury, 1978)—with the relative differences, resulting in improvement of overall performance. Using formulas derived from these early studies, Ghylin et al. (2006) successfully applied the inspection model to the X-Ray Tutor training program being used at hundreds of airports (Koller et al., 2007). Because Ghylin et al. used professional screeners and complex targets, the general application to naive participants for simple visual search was evaluated in the present study.

Of Drury's five proposed stages of inspection (Drury, 1994), it is the search and decision components that are most relevant to existing visual search models (i.e., Hoffman, 1979) and form the two principal aspects of the TCM. In these models, the search stage is largely bottom-up, rapid, and global; it is serially linked to the decision stage, which is more considered and involves both bottom-up and top-down processes that are mediated by existing knowledge and expertise (Schwaninger, Hardmeier, & Hofer, 2004). The decision component has its complement as the criterion acquisition time in SSMs (Smith, 2000), which is a function of rate of information

uptake (v), bias (z), and conservatism (a) (Voss, Rothermund, & Voss, 2004). The search component is termed as t_0 , or the nondecisional component of response time, which includes a motor component that is estimated rather than assumed to be equal, as is the case with the TCM.

Drury's (1994) speed/accuracy operating curve (SAOC), which hypothesizes that the probability of correct target detection increases with search time, does not necessarily have to be a trade-off if data are from the learning stage of a task. As in the SSMs, there exists in Drury's model a dichotomy between fixed-time search and unconstrained search; the latter poses the most difficulty to model assumptions. Given that airport screeners typically have between 3 and 6 sec to inspect an item (Schwaninger, Michel, & Bolting, 2007; Singh & Singh, 2004), the luggage search task used for this study limits image exposure to 4 sec and is therefore considered a fixed-time task.

Search can be modeled as either random or systematic (Wang, Lin, & Drury, 1997), differentiating between perfect and nonperfect search strategy (Morawski, Drury, & Karwan, 1980). Since the random-fixation process receives the most support in the literature, the two-component model is based on memoryless random search (Drury, 1994). Search time can alternatively be measured by using eyetracking techniques (i.e., fixations into area of interest; Liu, Gale, Purdy, & Song, 2006), although the TCM can be applied for search time estimation when eyetracking monitors are unavailable. The stopping policy in the TCM is functionally similar to the stopping criterion in SSMs, which represents the termination of evidence accumulation (Smith & Vickers, 1988) in favor of a response (the final stage in Drury's inspection model).

SSMs

SSMs attempt to reconcile the deficiencies of DT by linking detection performance and underlying RT distributions in two-choice tasks (Smith, 2000) while accurately estimating the speed-accuracy trade-off effects on the basis of an accumulation of evidence (Ratcliff & Smith, 2004). SSMs account for the different distributions and scales that speed and detection accuracy have (Ratcliff, 2001). SSM techniques fork, depending on whether a relative or absolute stopping rule is hypothesized, splitting into random walk models (Link & Heath, 1975) and recruitment models (LaBerge, 1962). From the literature of simple two-choice tasks, diffusion models are by far the most prominent methods for evaluating data that are split into detection times and RTs (Ratcliff & Tuerlinckx, 2002).

SSMs are rooted in neuropsychological studies (Smith & Ratcliff, 2004) that are sound, theoretically, for short, single-stage decision process tasks (mean RT \approx 1,000–1,500 msec; Ratcliff, 2006; Ratcliff & McKoon, 2008), but their application can also be relevant to impaired (i.e., aphasic and dyslexic; Ratcliff, Perea, Colangelo, & Buchanan, 2004) and older (aged 75–90 years; Ratcliff, Thapar, & McKoon, 2007) populations, who typically exhibit RTs slower than the recommended test maxima, as well as to simple visual search tasks (Schwarz, 1993). However, when responses over 1 sec are recorded, conclusions are likely to be based on multiple failed decision

attempts (Ratcliff & Rouder, 1998), resulting in a series of unpredictable diffusion processes operating at various times as various evidences are consumed and rejected. For these reasons, SSMs are unsuitable for luggage search tasks; therefore, only the TCM will be applied to our present data, although these issues remain pervasive when one considers applied data from visual inspection tasks.

Aging

Research into older screeners reveals mixed results. Older adults show diminished memory ability when tested on computerized tasks and are less efficient in difficult visual search conditions, as compared with a younger matched sample (Laguna & Babcock, 2000; Wiegmann, McCarley, Kramer, & Wickens, 2006). However, it has been shown that, with routine training, older workers can overcome these deficits, and that any age differences are small in comparison with other factors (Riegelnic & Schwaninger, 2006). Accordingly, we sought to determine whether age was a factor in laboratory performance on the X-ray object recognition task (X-ray ORT; Hardmeier, Hofer, & Schwaninger, 2005).

In summary, luggage search differs from traditional laboratory visual search tasks because each search typically takes several seconds; and, although SSMs are more sophisticated than the TCM, this fact alone renders SSMs inappropriate for use in applied settings. However, the deeper theoretical understanding of performance evaluation provided by SSMs is important to consider when determining the TCM's efficacy. It is also advantageous to understand how the mechanisms that control performance in these tasks work, given that X-ray luggage screening is a multibillion-dollar security industry (Butler & Poole, 2002). This study provides performance norms for the two limits of working age (younger, older) for untrained and trained participants, and model fitting guidelines are provided.

METHOD

Participants

The 12 younger participants (mean age = 20.8 years, $SD = 1.4$; 6 female) were recruited via advertisement on campus; the 12 older participants (mean age = 60.0 years, $SD = 3.7$; 7 female) were recruited from the University of the Third Age. The participants were offered payment of £15 for the day and were required to abstain from alcohol and drugs. None had previous experience of luggage or security surveillance, and all had normal or corrected-to-normal vision. The participants were screened for recreational drug use via urine sample (6 Drug MultiTest 1; SureScreen Diagnostics, Derby, England) and completed a number of personality tests to ensure normality, including the State-Trait Anxiety Inventory for Adults (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), the Arousal Predisposition Scale (Coren & Mah, 1993), and the Eysenck Personality Questionnaire (Eysenck, Eysenck, & Barrett, 1985). Normality was deduced by using the normal value limits given in each of these tests as reference.

Materials and Procedure

The data presented comprise the first and third of a three-session training day. The second session showed performance improvements between those of the first and third sessions; it has been omitted, allowing us to focus on the training effect before and after the day's training. The participants undertook the X-ray ORT consecutively on PCs running Microsoft Windows XP, on 17-in. TFT monitors,

with a 5-min break between sessions. The participants were seated in a soundproofed, light-controlled room using blackout blinds during the midday and were provided a light meal prior to testing. Each session lasted between 20 and 30 min.

The X-ray ORT is a computer-based program that presents participants with X-ray images of simulated luggage items, 128 with threat objects and 128 without threat objects. Threat objects were either knives or guns (8 of each) of different rotations (easy and difficult) within the luggage and were shown prior to the test to aid familiarization. Bag complexities were kept constant, as judged by human raters. Before each session, participants were presented with 8 practice images, given response feedback, and shown all the threat items used in the test, to remove knowledge effects.

The relatively simple interface and demands of the X-ray ORT make it ideal for modeling learning effects in naive participants who are unaccustomed to X-ray luggage screening, as compared with using unusual threats, such as improvised explosive devices, which measures knowledge of threats as much as it does visual search. For modeling purposes, we are only interested in visual search ability, not knowledge of different threats.

A model of estimated speed-accuracy trade-off can be obtained by applying the formulas originally derived by Spitz and Drury (1978). Data could not be modeled for participants producing fewer than five false alarms. For each participant, mean RT was sorted in ascending order and assigned an incremental ordinal number per threat-item hit, starting from 1; the probability of a hit at each threat-item response $\{P[\text{Hit}(x)]\}$ was generated by dividing the ordinal by total threat-item count, which essentially adopts Drury's (1994) $p(\text{error})$ formulation. These probabilities were mapped to the corresponding RTs and used as the ordinate in the equation below:

$$y = \log\{1 - P[\text{Hit}(x)]/Pdh\}, \quad (1)$$

where $P[\text{Hit}(x)]$ at any given RT (x) is the cumulative probability of a hit for each RT and Pdh is the overall probability of a hit derived by dividing the sum of total hits by the total number of threat items. This gives a y -value for each RT (x) that can be graphed for each participant's session, so that a best-fitting linear equation (judged by r^2) can be fit to the various y and x (RT) points. The slope (gradient) and x -intercept of the linear equation represent the search time (STh) and the nonsearch time (NSTh), respectively. To graph the overall SAOC using these values, a second formula can be used:

$$P(\text{Hit}) = \{1 - \exp[-(x - \text{NSTh})/\text{STh}]\} * Pdh, \quad (2)$$

where $P(\text{Hit})$ is the y -value that is generated by graphing the function that uses the values of the NSTh, STh, and Pdh (see Figure 1), and this formula is similar to other speed-accuracy curve equations (i.e., Wickelgren, 1977). This method is then repeated for false alarms by following the exact same method, although, to calculate the ordinal only, the nonthreat items should be considered. The parameters from each participant are then averaged to form the average SAOC shown in Figure 1 and the data given in Table 1. The stopping-time policy is obtained by linear correlation of average hit time per person (detection) compared with average correct rejection time (stopping policy) to give an account of speed for purely accurate judgments.

The sensitivity measure A' was chosen over d' because it makes no assumptions about signal and noise distribution normality. Overall A' detection values, B'' bias, and mean RTs were calculated for younger and older participants using the formulas below, where H is hit rate (number of hits divided by number of threat items) and F is the false alarm rate (number of false alarms divided by number of nonthreat items):

$$\begin{aligned} \text{If } H \geq F, \\ A' &= 0.5 + [(H - F)(1 + H - F)]/[4H(1 - F)] \\ B'' &= [H(1 - H) - F(1 - F)]/[H(1 - H) + F(1 - F)] \\ \text{If } H < F: \\ A' &= 0.5 - [(F - H)(1 + F - H)]/[4F(1 - H)] \\ B'' &= [-H(1 - H) + F(1 - F)]/[H(1 - H) + F(1 - F)] \end{aligned}$$

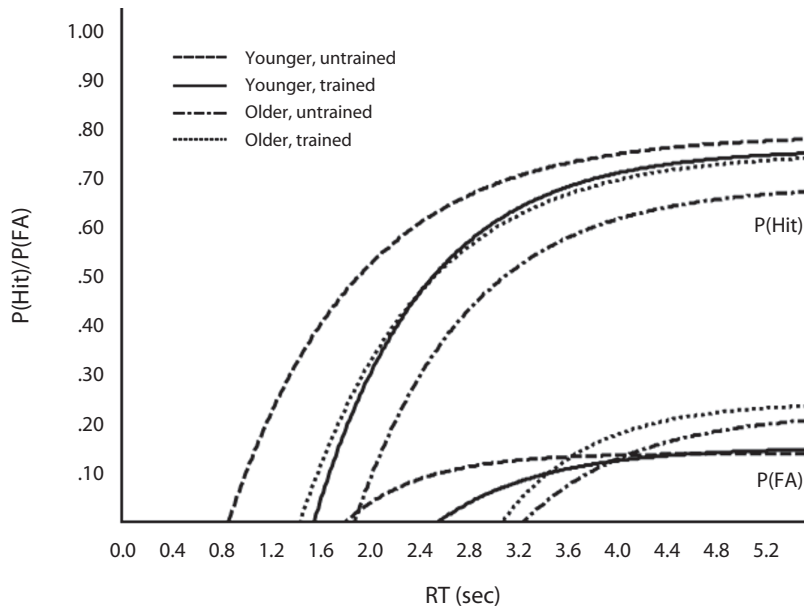


Figure 1. The speed–accuracy operating curve. The younger trained group was faster and more accurate than any other group. There is a clear training effect for speed and accuracy.

Even though there is no formal model linking A' to Grier's B'' (Macmillan & Creelman, 1991), B'' is one of the most useful determinants of bias and remains valid when participants are forced by program design to respond to each image. Significance levels for detection ability were analyzed using a repeated measures ANOVA with group as the between-participants factor, and generalized η^2 was calculated according to Olejnik and Algina (2003), as recommended by Bakeman (2005).

RESULTS

The SAOC

Figure 1 shows the fits from model parameters for younger and older participants on their first and third sessions. Data were eliminated in 0.38% of the trials as outliers (excessively slow RTs, measured by the mean plus

2 SDs), and 3 participants had fewer than five false alarms for which data could not be reliably modeled. The r^2 values for model goodness of fit were higher for hits than for false alarms (Table 1), although only nonsearch time improved significantly with training (Table 2), and age was found not to be a significant factor for any individual SAOC component. All interactions were insignificant, so only main effects are shown in Table 2.

The stopping-time policies for each participant can be mapped as shown in Figure 2, with the effects of training and age being combined to produce a linear relationship ($r^2 = .573$). This shows that detection time and stopping policies are proportionally linked and reflects slowing due to both training and age.

Table 1
Model Parameters for Each of the TCM Parameters

| | Hits | | | | | | | | False Alarms | | | | | | | |
|-------------|----------|-----------|----------|-----------|----------|-----------|----------|-----------|--------------|-----------|----------|-----------|----------|-----------|----------|-----------|
| | Pd | | NST | | ST | | r^2 | | Pd | | NST | | ST | | r^2 | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Younger (u) | .703 | .089 | 1.087 | 0.259 | 1.002 | 0.340 | .913 | .055 | .151 | .102 | 2.552 | 1.165 | 0.837 | 0.448 | .747 | .145 |
| Younger (t) | .785 | .084 | 0.849 | 0.225 | 1.049 | 0.187 | .965 | .024 | .139 | .106 | 1.800 | 1.014 | 0.615 | 0.098 | .763 | .126 |
| Older (u) | .683 | .088 | 1.872 | 0.554 | 0.932 | 0.478 | .905 | .061 | .223 | .124 | 3.239 | 0.822 | 0.903 | 0.505 | .795 | .090 |
| Older (t) | .751 | .113 | 1.431 | 0.276 | 1.003 | 0.333 | .935 | .037 | .271 | .214 | 2.803 | 0.880 | 0.884 | 0.347 | .705 | .081 |

Note—Pd, probability of hit/false alarm; NST, nonsearch time; ST, search time; u, untrained; t, trained.

Table 2
Nonsearch Time (NST) and Search Time (ST) for Hits and False Alarms, by Age and Condition

| | Hits | | | | False Alarms | | | |
|----------|------------------------------------|--|---------------------------------|--|-----------------------------------|--|---------------------------------|--|
| | NST | | ST | | NST | | ST | |
| Age | $F(1,22) = 1.774, \eta^2 = .07$ | | $F(1,22) = 1.463, \eta^2 = .02$ | | $F(1,19) = 1.625, \eta^2 = .06$ | | $F(1,19) = 1.879, \eta^2 = .07$ | |
| Training | $F(1,22) = 16.887, \eta^2 = .05^*$ | | $F(1,22) = 0.861, \eta^2 = .03$ | | $F(1,19) = 5.631, \eta^2 = .07^*$ | | $F(1,19) = 2.640, \eta^2 = .03$ | |

* $p > .05$.

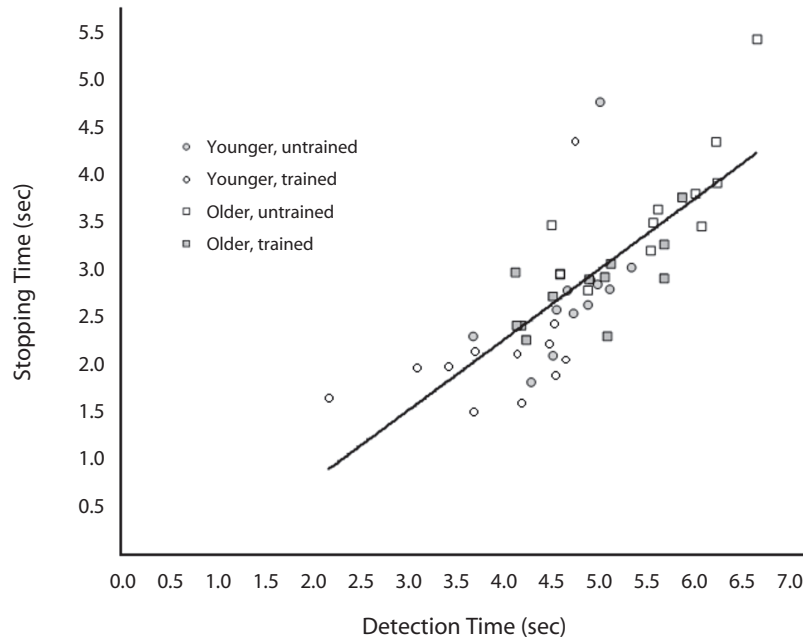


Figure 2. Stopping-time policy between the groups. There is a clear linear trend as age and training are shown in the stopping-time policy.

Detection Results

For the DT measures, overall detection ability (A') improved with training for both younger and older participants [$F(1,22) = 27.172, p < .05, \eta^2 = .15$], and there was a significant between-participants effect of age [$F(1,22) = 8.243, p < .05, \eta^2 = .27$]. These A' values and corresponding average item RTs are graphically summarized in Figure 3; trained participants completed the task quicker and more accurately than did untrained participants. It took three sessions for the older group (mean $A' = .858, SD = .026$; mean RT = 4.06 sec, $SD = 0.469$) to become about as accurate and as fast as the younger group had been during their first session (mean $A' = .852, SD = .051$; mean

RT = 4.09 sec, $SD = 0.481$). During the second session, the younger group averaged an A' of .885 ($SD = .03$) with an RT of 3.53 sec, whereas the older group averaged an A' of .846 ($SD = .05$) and an RT of 4.27 sec. Participants' B'' s did not significantly change after training, and age was found not to be a factor in bias differences.

DISCUSSION

High r^2 values were obtained for individual fits to the TCM, confirming that the TCM is a suitable method for deducing SAOCs in a simple luggage search task (Figure 1). Overall performance values, measured by A' ,

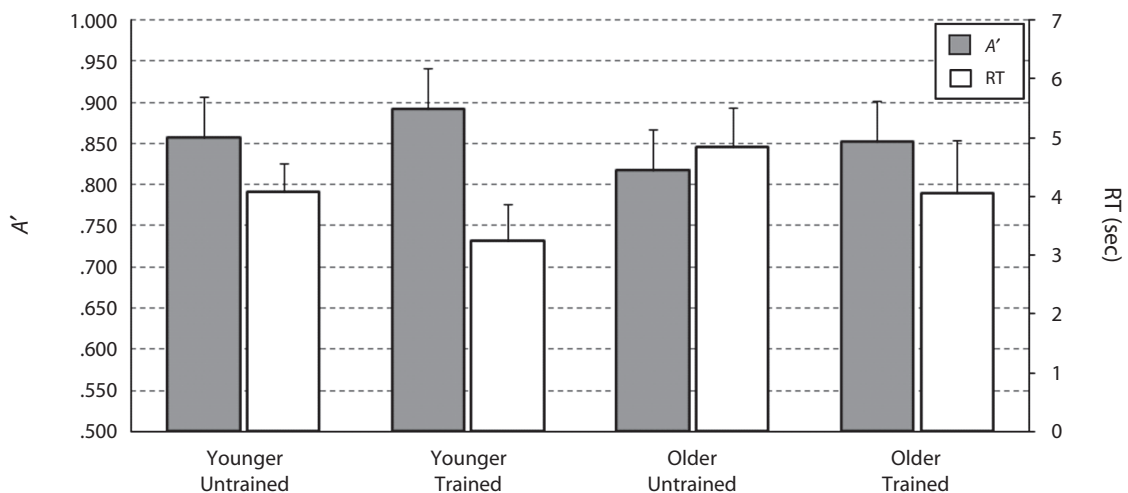


Figure 3. The signal detection values (A') and reaction time (RT) trends reflect the speed-accuracy operating curve trends found in the two-component inspection model, when analyzed separately.

increased significantly with training; RTs fell for both younger and older participants. This pattern of speed and accuracy improvement with training has been shown in previous luggage search task studies (Liu, Gale, & Song, 2007; McCarley, Kramer, Wickens, Vidoni, & Boot, 2004; Schwaninger & Hofer, 2004). Given that B'' did not alter with training, it can be deduced that the participants did not alter their strategy during the course of learning.

Stopping-time policy decreased in conjunction with detection time, indicating that behavior was altered toward speed for both threat and nonthreat items with training. The results also confirm that older adults in the late working-age bracket (55–65 years) were initially much slower and less accurate than their younger counterparts and required training encompassing over 750 images in order to perform at comparable levels (RT = 0.03 sec faster for the older group, and A' of only .006 more accurate in this study). Workers in this age group may be able to overcome initial deficiencies with enough training, but there will be a cost implication to consider.

The statistically significant training effect for the decision component of the model for hits—but not for the search component—accords with previous findings (McCarley et al., 2004). Liu et al. (2007) reported that practice improved focused search into specific areas of interest within the visual scene, as well as visual dwell time, which can be approximated to both search and decision time in the TCM. These findings also mimic Ghylin et al.'s (2006) report that the nonsearch component is improved between experienced screeners and new screeners. Of interest is the finding that no singular component of the SAOC achieved experimental significance when we compared younger with older participants, but overall differences are evident from Figure 1.

There are certain areas where the model can be improved. First, it is biased toward average performance, since exceptional performances have to be omitted for failing to provide enough points for linear regression of false alarms, which is a difficulty that can be traced back to the model's inception (Drury, 1975, p. 261). For this reason, it has been the convention until recent years to measure correct rejections as opposed to false alarms. There remains a large difference, which could be estimated from confidence ratings, between average probability of a hit for easy targets and that for difficult targets—particularly in the time domain—that shows as fast RTs for easy targets and slow RTs for difficult targets. Also, the given formulas are stepwise in nature, owing to a cumulative ordinal value, leading to an inflated linear model having correlations with points restricted around the ordinal value. However, these detractors are offset by the utility of the model for approximating search and decision time without the need for cumbersome apparatuses.

In conclusion, the TCM is a useful tool for gauging the speed–accuracy trade-off, and it accurately reflects differences in response between the upper and lower working-age limits, as well as within-group training effects. Thus, continuing research into both the TCM and SSMs could complement each other toward a unified visual search and

inspection model suitable for lengthier duration tasks, as are found in real-world settings.

AUTHOR NOTE

The second author is now at the Division of Sleep Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA. This study was supported by a research grant from the Economic and Social Research Council, U.K., Grant RES-000-23-1583. Correspondence concerning this article should be addressed to A. W. J. Wales, University of Zurich, Visual Cognition Research Group, c/o CASRA/APSS, Thurgauerstr. 39, 8050 Zurich, Switzerland (e-mail: alan.wales@access.uzh.ch).

REFERENCES

- BAKEMAN, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*, 379-384.
- BOLFING, A., HALBHERR, T., & SCHWANINGER, A. (2008). How image-based factors and human factors contribute to threat detection performance in X-ray aviation security screening. In A. Holzinger (Ed.), *HCI and usability for education and work* (pp. 419-438). Berlin: Springer.
- BUTLER, V., & POOLE, R. W., JR. (2002, July). *Rethinking checked-baggage screening* (Policy Study 297). Los Angeles: Reason Foundation.
- COREN, S., & MAH, K. B. (1993). Prediction of physiological arousal: A validation of the Arousal Predisposition Scale. *Behaviour Research & Therapy*, *31*, 215-219.
- DRURY, C. G. (1975). Inspection of sheet metal materials: Model and data. *Human Factors*, *17*, 257-265.
- DRURY, C. G. (1994). The speed–accuracy trade-off in industry. *Ergonomics*, *37*, 747-763.
- EYSENCK, S. B. G., EYSENCK, H. J., & BARRETT, P. (1985). A revised version of the psychoticism scale. *Personality & Individual Differences*, *6*, 21-29.
- GHYLIN, K. M., DRURY, C. G., & SCHWANINGER, A. (2006, July). *Two-component model of security inspection: Application and findings*. Paper presented at the 16th World Congress of Ergonomics, Maastricht, The Netherlands.
- 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. In *Proceedings of the 39th Annual IEEE International Carnahan Conference on Security Technology* (pp. 189-192). Piscataway, NJ: IEEE Press.
- HOFFMAN, J. E. (1979). A two-stage model of visual search. *Perception & Psychophysics*, *25*, 319-327.
- KOLLER, S. M., HARDMEIER, D., MICHEL, S., & SCHWANINGER, A. (2007). Investigating training and transfer effects resulting from recurrent CBT of X-ray image interpretation. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 1181-1186). Austin, TX: Cognitive Science Society.
- LABERGE, D. A. (1962). A recruitment theory of simple behavior. *Psychometrika*, *27*, 375-396.
- LAGUNA, K. D., & BABCOCK, R. (2000). Computer testing of memory across the adult life span. *Experimental Aging Research*, *26*, 229-243.
- LINK, S. W., & HEATH, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, *40*, 77-105.
- LIU, X., GALE, A., PURDY, K., & SONG, T. (2006). Is that a gun? The influence of features of bags and threat items on detection performance. In *Contemporary Ergonomics 2006: Proceedings of the International Conference on Contemporary Ergonomics* (pp. 17-22) London: CRC Press.
- LIU, X., GALE, A., & SONG, T. (2007). Detection of terrorist threats in air passenger luggage: Expertise development. In *Proceedings of the 41st Annual IEEE International Carnahan Conference on Security Technology* (pp. 301-306). Piscataway, NJ: IEEE Press.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- MCCARLEY, J. S., KRAMER, A. F., WICKENS, C. D., VIDONI, E. D., & BOOT, W. R. (2004). Visual skills in airport-security screening. *Psychological Science*, *15*, 302-306.
- MORAWSKI, T., DRURY, C. G., & KARWAN, M. H. (1980). Predicting search performance for multiple targets. *Human Factors*, *22*, 707-718.

- OLEJNIK, S., & ALGINA, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, **8**, 434-447.
- PLESKAC, T. J., & BUSEMEYER, J. (2007). A dynamic, stochastic theory of confidence, choice, and response time. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 563-568). Austin, TX: Cognitive Science Society.
- RATCLIFF, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.
- RATCLIFF, R. (2001). Diffusion and random walk processes. In *The international encyclopedia of the social and behavioral sciences* (Vol. 6, pp. 3668-3673). Oxford: Pergamon.
- RATCLIFF, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, **53**, 195-237.
- RATCLIFF, R., & MCKOON, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, **20**, 873-922.
- RATCLIFF, R., PEREA, M., COLANGELO, A., & BUCHANAN, L. (2004). A diffusion model account of normal and impaired readers. *Brain & Cognition*, **55**, 374-382.
- RATCLIFF, R., & ROUDER, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, **9**, 347-356.
- RATCLIFF, R., & SMITH, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, **111**, 333-367.
- RATCLIFF, R., THAPAR, A., & MCKOON, G. (2007). Application of the diffusion model to two-choice tasks for adults 75-90 years old. *Psychology & Aging*, **22**, 56-66.
- RATCLIFF, R., & TUERLINCKX, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, **9**, 438-481.
- RIEGELNIG, J., & SCHWANINGER, A. (2006). The influence of age and gender on detection performance and the criterion in X-ray screening. In *Proceedings of the 2nd International Conference on Research in Air Transportation* (pp. 403-407). Belgrade: ICRAT.
- SCHWANINGER, A., BOLFFING, 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. In *Proceedings of the 3rd International Conference on Research in Air Transportation* (pp. 317-324). Fairfax, VA: ICRAT.
- SCHWANINGER, A., HARDMEIER, D., & HOFER, F. (2004). Measuring visual abilities and visual knowledge of aviation security screeners. In *Proceedings of the 38th Annual IEEE International Carnahan Conference on Security Technology* (pp. 258-264). Piscataway, NJ: IEEE Press.
- SCHWANINGER, A., & HOFER, F. (2004). Evaluation of CBT for increasing threat detection performance in X-ray screening. In K. Morgan & J. M. Spector (Eds.), *Advances in learning, commerce and security* (pp. 147-156). Southampton, England: WIT Press.
- SCHWANINGER, A., HOFER, F., & WETTER, O. E. (2007). Adaptive computer-based training increases on the job performance of X-ray screeners. In *Proceedings of the 41st Annual IEEE International Carnahan Conference on Security Technology* (pp. 117-124). Piscataway, NJ: IEEE Press.
- SCHWANINGER, A., MICHEL, S., & BOLFFING, A. (2007). A statistical approach for image difficulty estimation in X-ray screening using image measurements. *Applied Perception in Graphics & Visualization*, **253**, 123-130.
- SCHWARZ, W. (1993). A diffusion model of early visual search: Theoretical analysis and experimental results. *Psychological Research*, **55**, 200-207.
- SINGH, M., & SINGH, S. (2004). Image segmentation optimisation for X-ray images of airline luggage. In *Proceedings of the IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety* (pp. 10-17). Piscataway, NJ: IEEE Press.
- SMITH, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, **44**, 408-463.
- SMITH, P. L., & RATCLIFF, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, **27**, 161-168.
- SMITH, P. L., & VICKERS, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, **32**, 135-168.
- SPIELBERGER, C. D., GORSUCH, R. L., LUSHENE, R. E., VAGG, P. R., & JACOBS, G. A. (1983). *State-Trait Anxiety Inventory for Adults*. Menlo Park, CA: Consulting Psychologists Press.
- SPITZ, G., & DRURY, C. G. (1978). Inspection of sheet materials: Test of model predictions. *Human Factors*, **20**, 521-528.
- VOSS, A., ROTHERMUND, K., & VOSS, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, **32**, 1206-1220.
- WALD, A. (1947). *Sequential analysis*. New York: Wiley.
- WANG, M.-J., LIN, S.-C., & DRURY, C. G. (1997). Training for strategy in visual search. *International Journal of Industrial Ergonomics*, **20**, 101-108.
- WICKELGREN, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, **41**, 67-85.
- WIEGMANN, D., MCCARLEY, J. S., KRAMER, A. F., & WICKENS, C. D. (2006). Age and automation interact to influence performance of a simulated luggage screening task. *Aviation, Space, & Environmental Medicine*, **77**, 825-831.

(Manuscript received November 11, 2008;
revision accepted for publication March 9, 2009.)