# Effects of Night Work, Sleep Loss and Time on Task on Simulated Threat Detection Performance

Mathias Basner, MD, MSc, MSc[1]; Joshua Rubinstein, PhD[2]; Kenneth M. Fomberstein, BA[1]; Matthew C. Coble[1]; Adrian Ecker[1]; Deepa Avinash, MS[1]; David F. Dinges, PhD[1]

[1]Division of Sleep and Chronobiology, Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA; [2]Department of Homeland Security's Transportation Security Laboratory Human Factors Program, Atlantic City, NJ

**Study Objectives**: To investigate the effects of night work and sleep loss on a simulated luggage screening task (SLST) that mimicked the x-ray system used by airport luggage screeners.

**Design**: We developed more than 5,800 unique simulated x-ray images of luggage organized into 31 stimulus sets of 200 bags each. 25% of each set contained either a gun or a knife with low or high target difficulty. The 200-bag stimuli sets were then run on software that simulates an x-ray screening system (SLST). Signal detection analysis was used to obtain measures of hit rate (HR), false alarm rate (FAR), threat detection accuracy (A'), and response bias (B"$_D$).

**Setting**: Experimental laboratory study

**Participants**: 24 healthy nonprofessional volunteers (13 women, mean age ± SD = 29.9 ± 6.5 years).

**Interventions**: Subjects performed the SLST every 2 h during a 5-day period that included a 35 h period of wakefulness that extended to night work and then another day work period after the night without sleep.

**Results**: Threat detection accuracy A' decreased significantly (P < 0.001) while FAR increased significantly (P < 0.001) during night work, while both A' (P = 0.001) and HR decreased (P = 0.008) during day work following sleep loss. There were prominent time-on-task effects on response bias B"$_D$ (P = 0.002) and response latency (P = 0.004), but accuracy A' was unaffected. Both HR and FAR increased significantly with increasing study duration (both P < 0.001), while response latency decreased significantly (P < 0.001).

**Conclusions**: This study provides the first systematic evidence that night work and sleep loss adversely affect the accuracy of detecting complex real world objects among high levels of background clutter. If the results can be replicated in professional screeners and real work environments, fatigue in luggage screening personnel may pose a threat for air traffic safety unless countermeasures for fatigue are deployed.

**Keywords**: Sleep; sleepiness; fatigue; performance; signal detection, luggage screening, time-on-task

**Citation**: Basner M; Rubinstein J; Fomberstein KM; Coble MC; Avinash D; Dinges DF. Effects of Night Work, Sleep Loss and Time on Task on Simulated Threat Detection Performance. *SLEEP* 2008;31(9):1251-1259.

ACCORDING TO THE DEPARTMENT OF HOMELAND SECURITY (DHS), MORE THAN 700 MILLION PIECES OF BAGGAGE OF PEOPLE TRAVELING ON COMMERCIAL aircraft are being screened for potential threats in the U.S. every year.[1] At the same time, the spectrum of possible threats has widened, liquid explosives being the most recent example. Therefore, high levels of threat detection performance by trained personnel are crucial for air traffic safety. The continuous requirement for detecting weak and infrequent signals among high levels of background clutter requires high and sustained levels of vigilance.

Fatigue may be caused by a variety of factors, including intrinsic and extrinsic sleep disorders, as well as work and lifestyle related changes in sleep schedules. Research has shown that fatigue from night work and sleep loss impairs vigilance performance sooner and more dramatically than most other cognitive functions. These deficits in attention appear to result from "wake state instability" manifesting as increased variability in endogenous alertness and compensatory effort as a result of increases in both errors of omission (i.e., lapses) and errors of commission (i.e., incorrect responses).[2-4]

Although the effects of sleep loss have been studied on a wide variety of vigilance and attention tasks,[4-6] there are no published experimental reports documenting the effects of night work and sleep loss on accuracy and speed of simulated threat detection performance. Thus, the major goals of this study were to determine the effects of fatigue induced by night work, sleep loss, repeated performance shifts, and time-on-task on speed and accuracy of object recognition and target search in nonprofessional subjects during a simulated luggage screening task in which threat detection was the primary performance. We hypothesized that night work and sleep loss would increase both errors of omission (missed threats) and errors of commission (false alarms) in threat detection performance.

## METHODS

We systematically evaluated threat detection performance on a simulated luggage screening task (SLST). Data of N = 24 subjects (mean age ± SD = 29.9 ± 6.5 years, range 22-40 years, 13 female) were obtained and analyzed. Participating subjects

**Figure 1**—Examples of simulated x-ray images of threat bags with typical hit rates. A: gun with low target difficulty in the center (HR was 75%), B: knife with low target difficulty in upper right corner (HR was 56.5%), C: gun with high target difficulty in lower right corner (HR was 50%), D: knife with high target difficulty in lower left corner (HR was 32.5%)
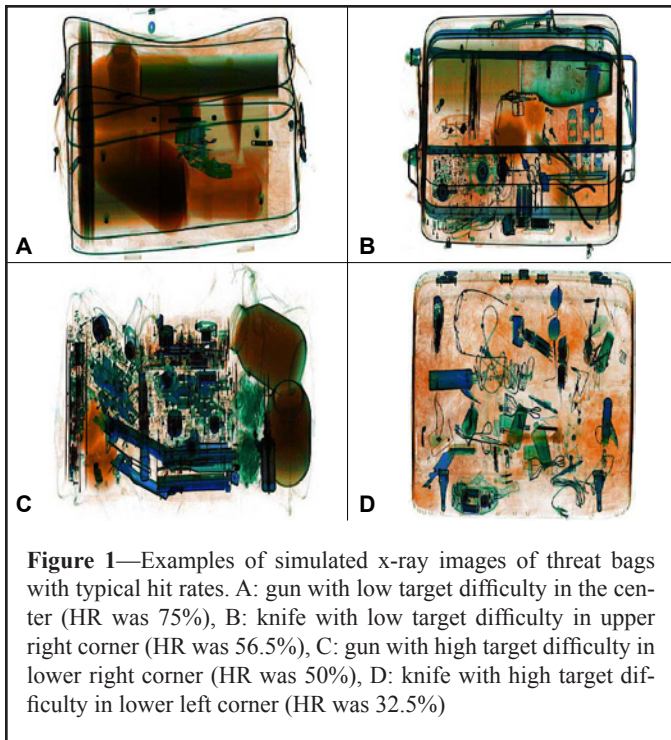
were not shift workers. Their sleep-wake behavior was monitored with sleep diaries and actigraphs in the week prior to the start of the experiment. During this period, subjects were instructed to adhere to their normal bedtimes. According to the Morningness-Eveningness Scale,[7] 15 subjects were evening types and 9 subjects were neither morning nor evening type. Participants were informed about potential risks of the study, and a written informed consent and HIPAA consent were obtained prior to the start of the study. The study was approved by the University of Pennsylvania IRB.

For the SLST, we developed an electronic luggage database that included a large number of x-ray source images of bags, clothing, etc., and threats (guns and knives only), provided by the Transportation Security Laboratory (TSA), Department of Homeland Security. These images were used by us to produce more than 5,800 unique simulated x-ray images of luggage, organized into 31 stimulus sets of 200 bags each, with 50 bags of each 200-bag set containing single threats that varied in type (gun or knife) and target difficulty (high or low). Thus, the a priori probability of a threat in any unique 200-bag set of stimuli was 0.25. Four typical examples are shown in Figure 1 A-D.

Subjects were oriented to the task and trained on 2 separate days. On both days, examples of separate clutter and threat images as well as complete bags with and without threats were shown to them. Subjects were informed that bags never contained multiple threats, and that threats were either knives or guns. They were also informed about the appearance of organic (orange) and metallic (blue/black) materials on the screen. An SLST with 30 bags (orientation day 1) and 200 bags (orientation day 2) was simulated and discussed with the subjects. Subjects were asked to identify a possible threat on the screen, and cases of misses or false alarms were discussed. Study participation did not depend on threat detection performance levels on orientation days.

Study participants stayed in the research lab for 5 consecutive days, which included a 35 h period of wakefulness (i.e., performance testing during the daytime, followed by performance testing at night, followed by performance testing during the day after a night without sleep). The study started at 08:00 on day one and ended at 08:00 on day 5. During 1 of every 2 h awake on all days, subjects performed a computerized neurobehavioral test battery (NTB) that lasted approximately 25-30 min, followed by an SLST performance bout (i.e., a unique set of 200 bags to screen).

On the first day of the study, all subjects performed 7 training bouts of the SLST (see below), followed by an 8-h sleep period. As a time in study (learning) effect was anticipated, the same protocol was repeated for half of the group, followed by a 35-h period of sleep deprivation, while the other half of the group underwent the sleep deprivation condition first (see Table 1). With this crossover design, the time in study effect was reduced for certain comparisons. Altogether, subjects performed 31 SLST performance bouts (7 training, 24 work) during the study. As the composition of each SLST differed according to type and target difficulty of threats, the 24 unique SLST test bouts were block randomized in a Latin square design (i.e., each SLST 200-bag stimulus set appeared once in each position of the 24 test bouts).

The 200-bag stimuli sets for each SLST test bout were run on software that simulates an x-ray screening system. Subjects had to press the space bar (colored green) for safe bags and the letter "D" (colored red) for threat bags. Except for the first training session, the threat-detection task timed out after 7 s, in which case threat bags were considered a miss, while safe bags were considered a correct rejection. A blank screen was shown for 1 s between presentations of 2 consecutive bags. During 3 of the 7 training sessions, detailed feedback was given to incorrect answers—that is the display presented when subjects missed a threat was "ERROR: weapon was present," and when they wrongly classified a safe bag as a threat bag it was "ERROR: NO weapon present." During all 24 test bouts during the experimental conditions (i.e., during the day following sleep, during night work, and during the day following no sleep), subjects were only informed about their overall percentage of hits and false alarms at the end of each 200-bag trial—no other details about their performance were provided. A text message also reminded them that the main goal of the task was to keep the threat detection rate high, while the secondary goal was to keep the rate of false alarms low, and that they should keep trying to attain a perfect score.

Hit rate (HR, true positive rate) and false alarm rate (FAR, false positive rate) were used to compute A' and $B''_D$, nonparametric signal detection theory measures of sensitivity and response bias.[8,9] Sensitivity A' reflects detection accuracy and reveals the extent to which subjects are able to differentiate signal (threat bags) from noise (safe bags). A' varies between 0.5 (signals cannot be distinguished from noise, performance at chance level) to 1.0 (complete separation of signal and noise, perfect accuracy). A' can also be interpreted as the proportion of times subjects would correctly identify the signal if signal and noise stimuli were presented simultaneously.[8] A' is unaffected by response bias (i.e., a subject's general willingness for responding "threat bag" versus "safe bag"). $B''_D$ is a measure of this response bias and ranges from −1 (liberal bias, yes to all) to

| Table 1—Study Design | | |
|---|---|---|
| **Time of Day** | **Group D1** | **Group D2** |
| Day 1 | | |
| 09:30 | T1 | T1 |
| 11:30 | T2 | T2 |
| 13:30 | T3 | T3 |
| 15:30 | T4 | T4 |
| 17:30 | T5 | T5 |
| 19:30 | T6 | T6 |
| 21:30 | T7 | T7 |
| 00:00 | Sleep 0:00 - 8:00 | Sleep 0:00 - 8:00 |
| Day 2 | | |
| 09:30 | **W1** | W1 |
| 11:30 | **W2** | W2 |
| 13:30 | **W3** | W3 |
| 15:30 | **W4** | W4 |
| 17:30 | **W5** | W5 |
| 19:30 | **W6** | W6 |
| 21:30 | **W7** | W7 |
| 23:30 | **W8** | Sleep 0:00 - 8:00 |
| 01:30 | **W9** | Sleep 0:00 - 8:00 |
| 03:30 | **W10** | Sleep 0:00 - 8:00 |
| 05:30 | **W11** | Sleep 0:00 - 8:00 |
| 07:30 | **W12** | Sleep 0:00 - 8:00 |
| Day 3 | | |
| 09:30 | **W13** | **W8** |
| 11:30 | **W14** | **W9** |
| 13:30 | **W15** | **W10** |
| 15:30 | **W16** | **W11** |
| 17:30 | **W17** | **W12** |
| 19:30 | Sleep 20:00 - 8:00 | **W13** |
| 21:30 | Sleep 20:00 - 8:00 | **W14** |
| 23:30 | Sleep 20:00 - 8:00 | **W15** |
| 01:30 | Sleep 20:00 - 8:00 | **W16** |
| 03:30 | Sleep 20:00 - 8:00 | **W17** |
| 05:30 | Sleep 20:00 - 8:00 | **W18** |
| 07:30 | Sleep 20:00 - 8:00 | **W19** |
| Day 4 | | |
| 09:30 | W18 | **W20** |
| 11:30 | W19 | **W21** |
| 13:30 | W20 | **W22** |
| 15:30 | W21 | **W23** |
| 17:30 | W22 | **W24** |
| 19:30 | W23 | Sleep 21:00 - 8:00 |
| 21:30 | W24 | Sleep 21:00 - 8:00 |
| 23:00 | Sleep 23:00 - 8:00 | Sleep 21:00 - 8:00 |

Simulated luggage screening task training bouts (T1-T7) and work bouts (W1–W24) are shown for the group that received the sleep deprivation condition first (D1) and for the group that received the condition after a second night of eight hours sleep (D2). The 35-hour sleep deprivation period is in bold type.

(1) **Night work effects**: Day work (D1: W18-W23, D2: W8-W13) was compared to night work (D1: W7-W12, D2: W14-W19).
(2) **Sleep loss effects**: Day work after at least 8 h time in bed (D1: W18-W22, D2: W8-W12) was compared to day work after a night without sleep (D1: W13-W17, D2: W20-W24).
(3) **Time in study effects**: Day work at the beginning of the study (D1 and D2: W1-W7) was compared to day work at the end of the study (D1: W18-W24, D2: W8-W14).

This choice assured that (a) circadian phase did not differ between conditions for sleep loss and time in study effects, and (b) that the experimental condition was preceded by the control condition in one group while it was followed by the control condition in the other group for night work and sleep loss effects, and therefore partially adjusting for possible time in study effects. Each subject contributed 10 (sleep loss), 12 (night work), or 14 (time in study) data points to the analysis.

Mixed effects regression models with random intercepts and random slopes for bout number within each condition with unstructured covariance were used for comparisons between conditions (Proc Mixed, SAS version 9.1, SAS Institute Inc.). An indicator variable was used to differentiate between experimental (night work, day work after a night without sleep, day work at the end of the study) and control conditions. A variable indicating work bout number was included in the model, adjusting for any residual differences in time in study between conditions. Finally, a group indicator variable adjusted for any differences between groups D1 and D2. Models were checked for interactions between bout number (the repeated measure) and condition (experimental vs. control). The interaction term was kept in the model if $P < 0.05$. Mixed model least square means and their differences are reported in the text and shown in Table 2 and Figure 2.

All 24 work bouts of the 24 subjects contributed to the analysis of a "time-on-task" effect. Here, each SLST 200-bag set was divided into 10 sets of 20 consecutive bags (i.e., 1-20, 21-40, …, 181-200). Thus, each subject contributed 10 data points to each analysis, each data point consisting of 24*5 = 120 threat bags and 24*15 = 360 safe bags. A mixed model with random intercepts and random slopes for centered bag set number with unstructured covariance was used to analyze the data for the dependent variables HR, FAR, A', B"$_D$, and response latency. Additionally, it was tested whether night work, sleep loss, or time in study interacted significantly with time-on-task. For the latter analysis, the restricted data set that was described in detail above was used.

Facilitating data from all 24 work bouts, average HR, FAR, A', and B"$_D$ were calculated for each subject depending on type (gun or knife) and difficulty (low or high) of threats. Average response latencies were calculated for threat bags and safe bags for each subject. For the threat bags, response latency was calculated for all threats, hits, and misses, depending on type and difficulty of threats. Mixed models with random subject effects were used to analyze the data. Degrees of freedom were adjusted according to Satterthwaite method. Tukey-Kramer adjustment was used for post hoc tests.

+1 (conservative bias, no to all), with 0 indicating no response bias in either direction.

## Statistical Analyses

For each of the 24 subjects, HR, FAR, A', B"$_D$, and trial duration were calculated for each of the 24 work bouts. Subsets of the 24 work bouts were used for the comparison of each of 2 conditions (see Table 1):
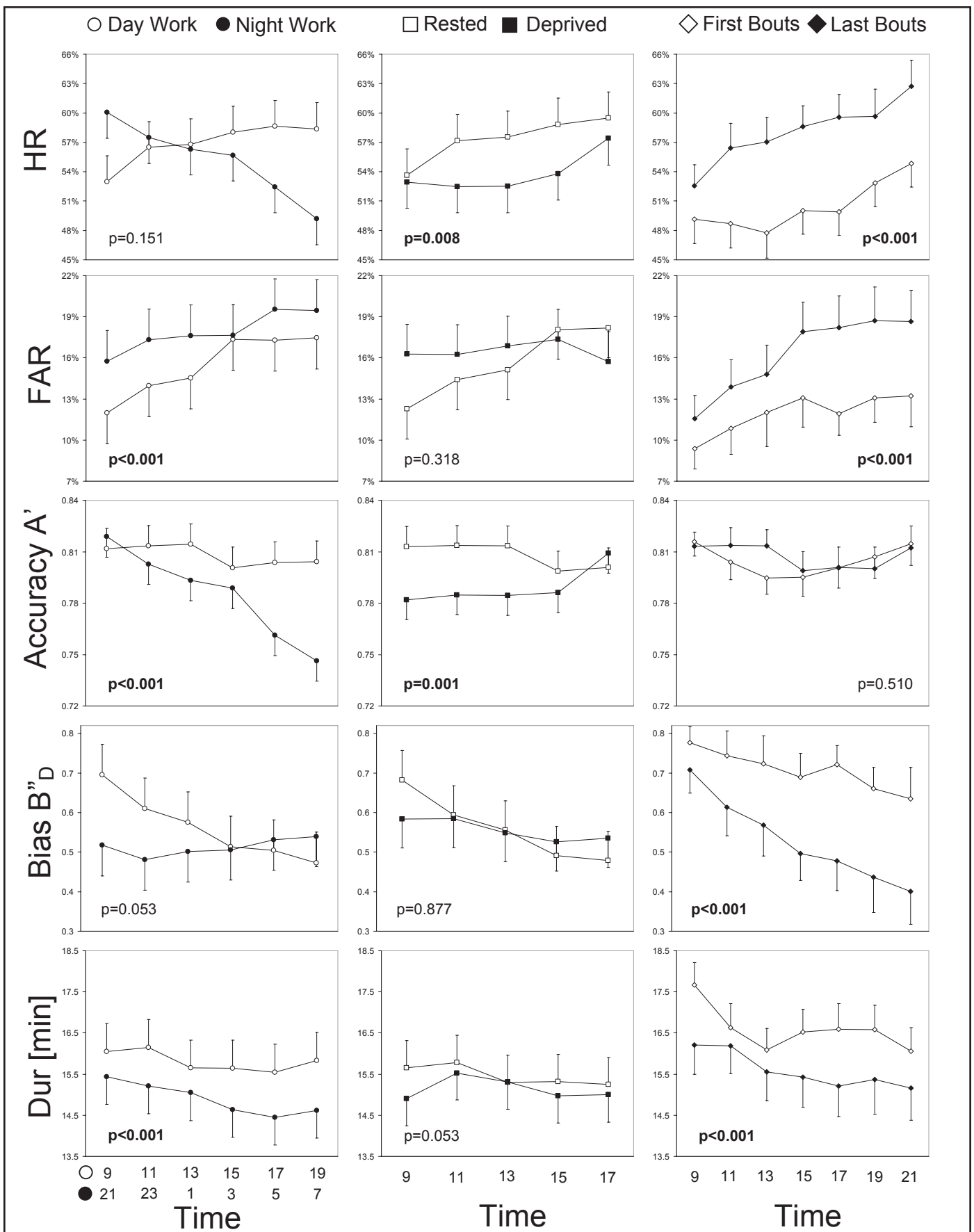
**Figure 2**—Effects of night work, sleep loss and time in study on HR, FAR, A', B''D, and bout duration (Dur) depending on time of day. Mixed model estimates of least square means are shown together with standard errors. P-values indicate differences between experimental and control conditions, i.e., between night work and day work, between deprived and rested subjects, and between last and first bouts in the study.
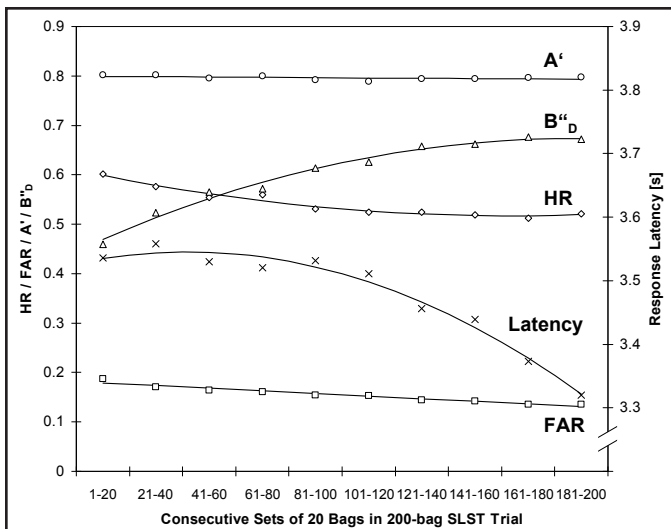
**Figure 3**—Effects of time-on-task on HR, FAR, A', B"D, and response latency are shown. Each 200-bag SLST bout was divided in 10 consecutive sets of 20 bags for all 24 work bouts of each of the 24 subjects. Each point represents the average of 24 individual values within a given bag set.



**Figure 4**—Effects of category (threat bag, safe bag), type of threat (gun or knife), and target difficulty (high difficulty HD or low difficulty LD) on average response latency plus standard errors. Figure 4A shows average response latencies irrespective of how subjects classified the bag. In Figure 4B response latencies are differentiated for bags classified as threats and safe bags.

## RESULTS

The effects of night work, sleep loss, and time in study on SLST performance are shown in Figure 2 and summarized in Table 2.
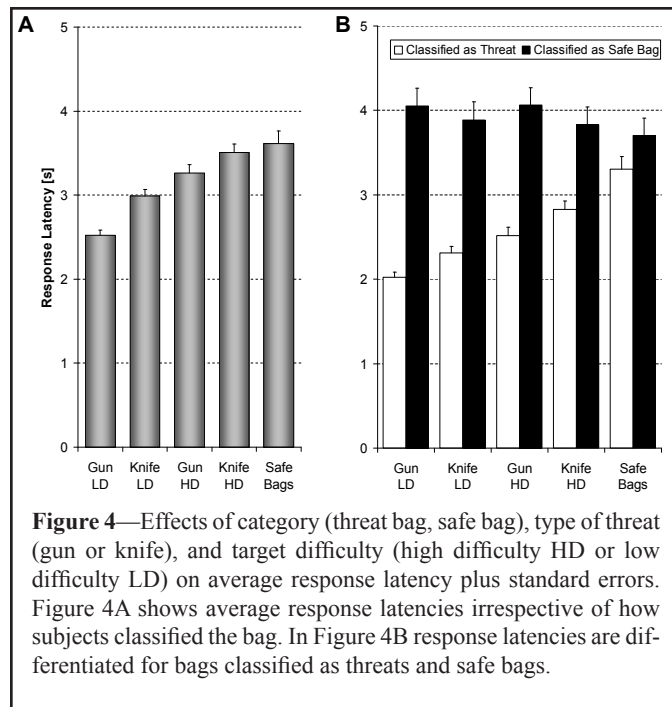
### Night Work Effects

Average HR decreased from 56.9% to 55.2% during night work (P = 0.151), while average FAR increased significantly from 15.4% to 17.9% (P < 0.001), leading to a significant decrease in average A' from 0.808 to 0.785 (P < 0.001). Response bias $B"_D$ decreased nonsignificantly from 0.562 to 0.513 (P = 0.053). Average bout duration decreased significantly from 15 min 49 s to 14 min 54 s during night work (P < 0.001). Averaging over all subjects and all work bouts, subjects needed 15 min 35 s to complete the task (range: 6 min 43 s to 24 min 25 s).

### Sleep Loss Effects

Average HR decreased significantly from 57.3% to 53.8% under the influence of sleep loss (P = 0.008), while average FAR increased from 15.6% to 16.5% (P = 0.318), leading to a significant decrease in average A' from 0.808 to 0.789 (P = 0.001). Response bias $B"_D$ decreased nonsignificantly from 0.561 to 0.556 (P = 0.877). Average bout duration decreased from 15 min 28 s to 15 min 8 s during sleep loss (P = 0.053).

### Time in Study Effects

Both average HR and FAR increased with simultaneously increasing study duration; average HR increased significantly from 50.4% at the beginning of the study to 58.1% at the end of the study (P < 0.001), and average FAR increased significantly from 11.9% to 16.3% (P < 0.001). This was caused by a significant decrease in response bias $B"_D$ from 0.707 to 0.529

(P < 0.001) —subjects were more willing to classify both safe and threat bags as threats towards the end of the study. Signal detection performance increased nonsignificantly from 0.805 at the beginning of the study to 0.808 at the end of the study (P = 0.510). Average bout duration decreased significantly from 16 min 36 s at the beginning of the study to 15 min 35 s at the end of the study (P < 0.001), i.e., subjects completed the task on average 1 min earlier at the end of the study.

### Time-On-Task Effects

Time on task effects on HR, FAR, A', and $B"_D$ are shown in Figure 3. Both HR ($P_{linear}$ and $P_{quadratic}$ < 0.001) and FAR (P < 0.001) decreased significantly with time-on-task. The decline in HR was more prominent at the beginning of the task. It decreased from 60.2% at the beginning to 52.2% at the end of the task. FAR decreased in a linear fashion from 18.8% at the beginning to 13.5% at the end of the task. SLST performance A' remained unchanged (P = 0.198) during time-on-task. Therefore, the simultaneous decrease in HR and FAR was caused by a significant ($P_{linear}$ < 0.001, $P_{quadratic}$ = 0.002) shift in response bias $B"_D$ towards more conservative criteria. Like HR, the increase in $B"_D$ was more prominent at the beginning of the task. Response latency decreased significantly ($P_{linear}$ < 0.001, $P_{quadratic}$ = 0.004) with time-on-task. In contrast to HR and $B"_D$, the decline in response latency was more prominent toward the end of the task. Average response latency decreased from 3.53 s per bag at the beginning to 3.32 s per bag at the end of the task. There were no significant interactions between time-on-task and night work, sleep loss, and time in study for HR, FAR, A', $B"_D$ or response latency at alpha = 0.05.

### Effects of Type and Difficulty of Threat

We sought to determine whether the effects of night work and sleep loss on threat detection were associated with varia-

| | Night Work (Night Work–Day Work) | Sleep Loss (Deprived–Rested) | Time in Study (Last Bouts-First Bouts) |
|---|---|---|---|
| **Hit Rate** | **−0.017** (−0.040, +0.006) | **−0.035**\*\* (−0.061, −0.009) | **+0.076**\*\*\* (+0.058, +0.094) |
| **False Alarm Rate** | **+0.025**\*\*\* (+0.010, +0.039) | **+0.009** (−0.009, +0.026) | **+0.043**\*\*\* (+0.030, +0.057) |
| **Accuracy A'** | **−0.023**\*\*\* (−0.034, −0.012) | **−0.019**\*\* (−0.030, −0.008) | **+0.003** (−0.006, +0.012) |
| **Response Bias B"$_D$** | **−0.049** (−0.100, +0.001) | **−0.005** (−0.067, +0.057) | **−0.179**\*\*\* (−0.223, −0.135) |
| **Bout Duration [s]** | **−54.8**\*\*\* (−73.2, −36.4) | **−19.1** (−38.4, +0.2) | **−60.3**\*\*\* (−78.6, −42.1) |

Point estimates of absolute changes (95% confidence limits) are given.
\*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001 (H0: no difference between groups)

tion in threat type (i.e., guns versus knives), or with threat difficulty (high versus low difficulty), or with a combination of these 2 factors. Based on a survey in our own lab, threats were classified into 4 categories: guns or knives with high or low target difficulty. An analysis based on pooled data of all 24 work bouts showed that HR was higher for guns than for knives (P < 0.001), and for threats with low target difficulty (P < 0.001) than those with high target difficulty, but there was no significant interaction between type and target difficulty of threat (P = 0.751). Average HR declined in the following order: HR = 75.3% for guns with low target difficulty, HR = 56.9% for knives with low target difficulty, HR = 51.6% for guns with high target difficulty, and HR = 32.5% for knives with high target difficulty. Post hoc tests with Tukey-Kraemer adjustment showed that all categories differed significantly from each other (P < 0.05). Further analyses showed that there were no significant 2- or 3-way interactions between type of threat, threat difficulty, and sleep loss, night work, or time in study (i.e., HR was non-differentially influenced by sleep loss, night work, and time in study for all types of threat).

As the estimation of FAR, which was 15.5% on average, is based on safe bags only, the effects of type and difficulty of threat on A' and B"$_D$ depend on changes in HR only, and A' and B"$_D$ were therefore not calculated.

The effects of type and difficulty of threat on response latency are shown in Figure 4. If hits and misses were not differentiated, the order observed for HR was reversed for response latency (Figure 4A). Average time used to scan each threat bag increased in the order of guns with low target difficulty (M = 2.52 s), then knives with low target difficulty (M = 2.99 s), then guns with high target difficulty (M = 3.27 s), and knives with high target difficulty (M = 3.51 s). Response latency for safe bags (M = 3.61 s) was higher compared to all threat bags. Response latency was lower for guns (P < 0.001) and for low difficulty threats (P < 0.001), with a significant interaction between type and target difficulty of threat (P = 0.019). There were no other significant 2- or 3-way interactions between type of threat, threat difficulty, and sleep loss, night work, or time in study. Post hoc tests with Tukey-Kraemer adjustment showed that all categories differed significantly from each other (P < 0.05) except for safe bags and knives with high target difficulty.

Time out rate (response latencies >7 s) was generally low and paralleled response latency. For example, time out rate increased in the following order: guns with low target difficulty (3.2%), knives with low target difficulty (4.3%), guns with high target difficulty (5.4%), and knives with high target difficulty (6.4%). Time out rate for safe bags was identical to bags with high target difficulty knives (6.4%).

Response latencies were also calculated separately for bags classified as threat bags and bags classified as safe bags. Figure 4B shows that response latencies were markedly shorter for bags classified as threat bags compared to those classified as safe bags. For correctly identified threats (hits), response latency increased in the same order that was observed for all threats (i.e., response latency was shorter for guns [P < 0.001] and low difficulty targets [P < 0.001] without significant interaction [P = 0.958]). Post hoc tests showed that all 4 categories differed significantly from each other and from safe bags wrongly classified as threat bags (i.e., false alarms, all adjusted P < 0.001). Response latency for incorrectly rejected threats (i.e., misses) was significantly higher for guns compared to knives (P < 0.001), but did not differ for threats with low and high target difficulty (P = 0.231), and there was no significant interaction (P = 0.123). Additionally, response latencies were significantly higher for misses in all 4 threat categories compared to correctly rejected safe bags (all adjusted P < 0.001).

## DISCUSSION

This is the first study investigating the effects of night work and sleep loss on threat detection performance with a simulated luggage screening task.

We could find only 2 studies that addressed sleep deprivation effects on signal detection performance, and both used auditory vigilance tasks.[10,11] Each study reported significant decreases in HR and detection accuracy, while FAR and response bias were either nonsignificantly increased or remained unchanged. Our results for a visual threat detection task following a night of sleep deprivation are consistent with these earlier auditory detection studies. A few recent studies addressed the effect of sleep loss on object recognition memory. In these studies, specific objects were learned during a learning phase. After a night

with or without sleep, the previously learned objects had to be recalled and differentiated from objects that did not belong to the learning set. In this way, it was shown in both mice[12] and in man[13] that sleep deprivation negatively affects memory consolidation for object recognition.

Ours was the first study using a visual search task to investigate the effects of sleep loss on signal detection performance. In contrast to the object recognition studies, only a few typical examples of threats were shown to the subjects prior to the experiment in this study, and subjects had to detect never before seen threats among high levels of background clutter in situations with varying degrees of sleep deprivation. This may explain the relatively low hit rates found in this study; but in combination with the high fidelity of the simulated luggage screening task, it guaranteed a high ecologic validity of the study results. Low motivation could be another reason for low hit rates, but we found no evidence for this. In fact, increasing response latencies with simultaneously increasing target difficulty, as well as increased response latencies for bags judged to be safe (i.e., no threat) suggests that subjects took a longer time to find threats. A reduced motivation argument would predict that they would hurry to end the task and in so doing have reduced response latencies for all bag types.

Both night work and sleep loss showed detrimental effects on signal detection performance using SLST threat detection stimuli. HR decreased and FAR increased during both night work and sleep loss, leading to a significant decrease in detection accuracy. A' remained more or less constant during the first 16 hours of wakefulness and deteriorated quickly with further sleep deprivation (Figure 2). This finding is consistent with a considerable amount of research demonstrating that performance decrements become evident only after wakefulness is extended 16 hours or more,[14] or into the circadian nadir. Worst SLST performance was found at 07:00 after 23 hours awake, which is consistent with the interaction of the homeostatic drive for sleep and endogenous circadian phase.[6] Studies suggest that the highest risk of sleepiness related traffic accidents is observed around this time of the day.[15,16]

Although the absolute changes in HR and FAR induced by sleep loss appear to be relatively modest, these performance reductions could have serious consequences in security operations. Because of the high amounts of luggage screened at U.S. airports (700 million per year) even minor changes in HR and FAR may cause relevant absolute increments in the number of missed threats, especially in the number of bags unnecessarily subjected to a thorough investigation. These findings emphasize the necessity of well-rested and vigilant luggage screeners to support and promote high levels of threat detection performance and air traffic safety.

It is unknown whether the adverse effects on threat detection performance we observed during night work and after sleep loss would be mitigated by more experience with the task. Following 2 orientation days and one day of initial training on the task, there was no significant further training effect (i.e., SLST performance did not improve towards the end of the study). One possible reason for the lack of a continued training effect is that there was no feedback on actual performance accuracy for each individual bag. Instead, subjects were informed about their overall performance only at the end of each 200-bag set.

Thus, it was impossible for them to learn from previous successes or mistakes.

However, response latency decreased significantly towards the end of the study, which in itself could be interpreted as an improvement, especially since threat detection accuracy did not decrease simultaneously. On the one hand, decreases in response latency without relevant improvements in detection accuracy could be explained by improving visual scanning without improving object recognition. This hypothesis is corroborated by findings of McCarley et al., who tracked eye movements during a simulated threat detection task.[17] Their subjects were quicker to localize and fixate targets with increasing practice, but were not more likely to do so. On the other hand, both HR and FAR were shown to increase significantly towards the end of the study. As hits and false alarms were shown to be associated with shorter response latencies compared to misses and correct rejections, the use of more liberal criteria (i.e., indicating a threat was present more frequently) alone could suffice to explain the overall decrease in bout duration towards the end of the study. The reminder after the completion of each 200-bag set that the primary goal was to achieve high detection rates and the secondary goal was to keep false alarm rate low may have contributed to the significant shift in response bias to more liberal criteria towards the end of the study.

The well known "vigilance decrement" was replicated for HR in our study.[18] In accordance with classical findings,[19] HR decreased prominently during the first minutes of the task, but deteriorated no further during the last third of the task. This decrease in HR was accompanied by a prominent time-on-task effect on response bias. With increasing time-on-task subjects applied more conservative decision criteria (i.e., they were less likely to say "threat" to both threat bags and safe bags). The same result was reported by Deaton et al.[10] Again, the change in response bias was strongest during the first half of the work bout, which may reflect subjects' increasing awareness of low threat prevalence, which was 25% in our study.[20,21] If the above findings were reproduced for professional luggage screeners, higher miss rates would become more and more likely with increasing time-on-task.

Additionally, time used to scan bags decreased with increasing time-on-task, especially toward the end of the bout. This could have been caused by the shift in response bias, as more conservative decision criteria may not only influence the tendency to say "no threat," but also the time needed to come to this decision. However, this is unlikely, as the greatest shift in response bias was observed during the first half of the task, while the greatest shift in response latency was observed during the second half of the task. Alternatively, improving target recognition or visual scanning towards the end of the task may have caused the decline in response latency.[17] This is a more likely explanation than declining motivation or increasing fatigue, as threat detection accuracy A' did not deteriorate with time-on-task in the whole data set.

Threat detection performance decreased in the expected order, with guns being detected more often and faster than knives, and with threats with low target difficulty being detected more often and faster than threats with high target difficulty. Night work and sleep loss decreased performance in the same degree for all types of threats.

In 1972, Emmerich et al.[22] showed in an auditory signal detection task where, after each trial, subjects had to rate the confidence in their decision that short response latencies were associated with higher levels of confidence in the decision. Therefore, stimuli close to the subject's decision criterion (in our case safe bags with suspicious clutter or threat bags with inconspicuous threats) may elicit longer response latencies compared to stimuli distant from the criterion. These findings were corroborated by the results of this study. Average response latency increased with target difficulty and was highest for knives and high target difficulty threats. Response latency associated with hits was up to 2 s shorter than response latency associated with misses. When subjects decided to classify a threat bag as a safe bag, they needed significantly more time to reach this decision in guns compared to knives. Additionally, subjects needed significantly more time to wrongly classify a threat bag as a safe bag than to correctly classify a safe bag as a safe bag.

## Limitations

The results of our study may not simply be transferred to real life conditions. First, we did not use professional airport luggage screeners in our study, who receive a special training that is refreshed on a regular basis. Threat detection performance presumably would have been higher in professionals, but it is unknown whether the effects of sleep loss would have differed from those observed in our subjects. In other professions that have been studied (e.g., pilots, physicians in training, professional truck drivers), the effects of sleep loss in the laboratory have generalized to the operational environment, although the magnitude of the effects can be smaller in a well trained professional cohort than in unselected laboratory subjects.[23-25]

Another reason our results may not readily generalize to the airport security environment is because the 25% threat prevalence we used was higher than found in security operations, at least for guns and knives. If all prohibited items (e.g., bottles, pocket knives, nail files/clippers, lighters, etc.) are taken into account, 25% threat prevalence is, in fact, not unusual at airport checkpoints, and there are times when the combined rate of different classes of prohibited items can exceed 25%. We did not use a lower threat prevalence because huge amounts of stimuli have to be presented in low prevalence bouts to achieve reliable estimates of HR. HR is known to decrease with simultaneously decreasing threat prevalence.[26] Therefore, to determine the extent to which our findings of the effects of night work and sleep loss on threat detection performance have ecological validity, studies would have to be conducted on trained professional screeners with more realistic prevalence rates for guns and knives. That need notwithstanding, our findings do suggest that threat detection performance is likely to be degraded by night work and inadequate sleep in professional screeners.

Our study work shifts were selected to evaluate simulated threat detection performance for approximately the same cumulative duration of time as occurs at the airport during a standard work shift. That is, performance on an airport x-ray machine can be anywhere between 20 and 30 min duration before workers are rotated to another work activity. Because screeners rotate between positions, it is possible for them to experience three to five 20 to 30-min x-ray shifts in an 8-h work shift, which is

what subjects in our experiment experienced when they performed the SLST every 2 h.

Keeping subjects active and performing for longer periods of time (i.e., throughout the night and after a night without sleep) than occurs with standard airport screener work shifts was done to address a concern TSA had about what happens to threat detection performance when screeners hold secondary jobs that cause them to perform their 8-h work shift for TSA following a work shift at another job. This extra long period of active work-time was accurately simulated in our study, but we also recognize that such prolonged experimental work periods do not generalize to the planned work shifts of airport screeners. Thus, we were not attempting to create work-rest schedules that precisely mimicked those of TSA workers (which can vary among airports), but rather, to answer basic questions of how fatigue from night work and sleep loss affected threat detection performance on a high-fidelity simulated threat detection task.

## Conclusions

The results of this experiment suggest that night work and sleep loss adversely affect performance of nonprofessional subjects on a task that simulates threat detection demands of airport screeners. Thus, if the results were to be replicated in professional screeners and real work environments, fatigue in luggage screening personnel could potentially pose a threat for air traffic safety unless countermeasures for fatigue are deployed. In the future, methods should be developed to predict signal detection performance based on brief fitness-for-duty tests or objective monitoring of screener alertness during the screening task, in an effort to assure high levels of vigilance and detection performance.

## REFERENCES

1. Department of Homeland Security's website (www.dhs.gov/xtrvlsec/), last visited on 02/02/2007. 2007
2. Doran SM, Van Dongen HP, Dinges DF. Sustained attention performance during sleep deprivation: evidence of state instability. Archives Italiennes de Biologie: A Journal of Neuroscience 2001;13:1-15.
3. Dorrian J, Rogers NL, Dinges DF. Psychomotor vigilance performance: a neurocognitive assay sensitive to sleep loss. In: Kushida CA, editor. Sleep deprivation: clinical issues, pharmacology and sleep loss effects. New York, NY: Marcel Dekker, Inc., 2005:39-70.
4. Durmer JS, Dinges DF. Neurocognitive consequences of sleep deprivation. Semin Neurol 2005;25:117-29.
5. Dinges DF, Kribbs NB. Performing while sleepy: effects of experimentally-induced sleepiness. In: Monk TH, ed. Sleep, sleepiness and performance. Chichester, United Kingdom: John Wiley and Sons, Ltd., 1991:97-128.
6. Van Dongen HP, Dinges DF. Circadian rhythms in fatigue, alertness and performance. In: Kryger MH, Roth T, Dement WC, eds.

Principles and practice of sleep medicine. Philadelphia: W.B. Saunders, 2000:391-399.

7. Horne JA, Ostberg O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol 1976;4:97-110.

8. Stanislaw H, Todorov N. Calculation of signal detection theory measures. Behav Res Methods Instrum Comput 1999;31:137-49.

9. Donaldson W. Measuring recognition memory. J Exp Psychol Gen 1992;121:275-77.

10. Deaton M, Tobias JS, Wilkinson RT. Effect of sleep deprivation on signal detection parameters. Q J Exp Psychol 1971;23:449-52.

11. Horne JA, Anderson NR, Wilkinson RT. Effects of Sleep-deprivation on signal-detection measures of vigilance - implications for sleep function. Sleep 1983;6:347-58.

12. Palchykova S, Winsky-Sommerer R, Meerlo P, Durr R, Tobler I. Sleep deprivation impairs object recognition in mice. Neurobiol Learn Mem 2006;85:263-71.

13. Wagner U, Kashyap N, Diekelmann S, Born J. The impact of post-learning sleep vs. wakefulness on recognition memory for faces with different facial expressions. Neurobiol Learn Mem 2007;87:679-87.

14. Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 2003;26:117-26.

15. Pack AI, Pack AM, Rodgman E, Cucchiara A, Dinges DF, Schwab CW. Characteristics of crashes attributed to the driver having fallen asleep. Accid Anal Prev 1995;27:769-75.

16. Akerstedt T, Kecklund G, Horte LG. Night driving, season, and the risk of highway accidents. Sleep 2001;24:401-6.

17. McCarley JS, Kramer AF, Wickens CD, Vidoni ED, Boot WR. Visual skills in airport-security screening. Psychol Sci 2004;15:302-6.

18. See JE, Howe SR, Warm JS, Dember WN. Metaanalysis of the sensitivity decrement in vigilance. Psychol Bull 1995;117:230-49.

19. Teichner WH. The detection of a simple visual signal as a function of time of watch. Hum Factors 1974;16:339-53.

20. Szalma JL, Hancock PA, Warm JS, Dember WN, Parsons KS. Training for vigilance: using predictive power to evaluate feedback effectiveness. Hum Factors 2006;48:682-92.

21. Craig A. Is vigilance decrement simply a response adjustment towards probability matching. Hum Factors 1978;20:441-46.

22. Emmerich DS, Gray JL. Response latency, confidence, and ROCs in auditory signal detection. Percept Psychophys 1972;11:65-72.

23. Rosekind MR, Boyd JN, Gregory KB, Glotzbach SF, Blank RC. Alertness management in 24/7 settings: lessons from aviation. Occup Med 2002;17:247-59, iv.

24. Philibert I. Sleep loss and performance in residents and nonphysicians: a meta-analytic examination. Sleep 2005;28:1392-402.

25. Philip P, Akerstedt T. Transport and industrial safety, how are they affected by sleepiness and sleep restriction? Sleep Med Rev 2006;10:347-56.

26. Wolfe JM, Horowitz TS, Kenner NM. Cognitive psychology: rare items often missed in visual searches. Nature 2005;435:439-40.