Simulating a vigilance task: Technology for homeland security research

Robert Hubal  Matthew S. Cain, Stephen R. Mitroff
RTI International Duke University
Research Triangle Park, NC Durham, NC
rhubal@rti.org {matthew.s.cain, mitroff}@duke.edu

ABSTRACT

Airport and border security personnel are working to improve security practices and preparedness to prevent acts of terrorism. While they must guard against false positives that result in needless delays for travelers, they must also balance those against the grave costs of missing a threat. In support of a study investigating individual differences in vigilance tasks, the authors have developed and pilot-tested a controllable simulated airport security screening environment. The simulation enables, in a more realistic setting than typically possible, research focused on understanding issues such as:

- Implications when task instructions differ in detail;
- Implications when task instructions require simultaneous search for multiple classes of objects, or when the classes have differing security implications;
- Implications of distraction (e.g., noise, activity, both expected and not) in the environment.

These factors allow for an examination of distinct contributions to attentional processing, top-down executive control and bottom-up perceptual processing. Both processes need to be considered to adequately address perception of and attention to items in the environment. Ongoing basic studies are comparing use of the simulated environment against traditional techniques. Ongoing work with Government security personnel is directing the application to existing training needs to better determine how training can influence performance.

ABOUT THE AUTHORS

Robert Hubal, Ph.D., a senior researcher in RTI’s Behavioral Health and Criminal Justice division, conducts research focusing on development, presentation, and evaluation of materials and identifying approaches to improve training and assessment effectiveness. He has investigated the interaction effects among cognitive task demands, the role of experience, mental models, and representation of information. He is interested in experimental evaluations of the usability, acceptance, and cost effectiveness of training and assessment systems and their applications to both everyday and specialized domains.

Matt Cain, Ph.D., is a postdoctoral associate at the Duke University Department of Psychology & Neuroscience and Center for Cognitive Neuroscience, where he studies visual attention, executive control, and how context and expertise affect these processes. His research looks at cognitive control and executive attention using behavioral and fMRI methods. In addition to training provided in a lab setting, he is interested in how real-life activities such as speaking multiple languages and playing video games can influence low-level cognitive processes.

Stephen Mitroff, Ph.D., is an Assistant Professor in the Department of Psychology & Neuroscience and Center for Cognitive Neuroscience at Duke University. His research interests focus on the study of visual cognition, examining the nature of visual memory, perception, and attention. He is currently working with the Institute of Homeland Security Solutions, the Army Research Office, and Nike, Inc. to explore human factors of visual cognition, with a focus on individual differences on visual search tasks.
INTRODUCTION

Recent airport security breaches have led to increased concerns over security practices and preparedness. As security personnel work to prevent potential acts of terrorism, they also must guard against false positives that result in needless delays for travelers. In support of a study investigating individual differences in vigilance tasks that might influence the false positive rate, the authors have developed a controllable simulated airport security screening environment. The simulation enables, in a more realistic setting than typically possible, research focused on understanding issues such as:

- Implications when task instructions differ in detail (e.g., detect weapons vs. guns and needles);
- Implications when task instructions require simultaneous search for multiple classes of objects (e.g., detect guns vs. needles), or when the classes have differing security implications (e.g., detect guns and needles vs. water bottles);
- Implications when search provides information in one modality (e.g., an auditory cue) that is dependent on the class of object (e.g., one tone for guns, another for water bottles);
- Implications when cues (e.g., auditory tones) are synchronized not with the class of objects but with the individual’s actions (e.g., one tone when a target is identified, another when missed);
- Implications of the characteristics of these cues (e.g., does it matter if a high tone is used for hits and low for misses?); and
- Implications of distraction (e.g., noise, activity, both expected and not) in the environment.

These factors allow for an examination of distinct contributions to attentional processing, top-down executive control and bottom-up perceptual processing. Both processes need to be considered to adequately address perception of and attention to items in the environment (Costello et al., 2010). The authors are working with Government security personnel to direct continuing studies. An ultimate goal of these studies is to better determine how training can be tailored to influence performance.

BACKGROUND

This paper describes the simulation for a homeland security related task. Since there was not a lot of prior work in developing a simulation as envisioned for this type of vigilance task (Hubal et al., 2010), there was no clear design to start with for the simulation. Thus, several cycles of simulation development took place. The first version of the simulation, for instance, was visually appealing but not suitable to run an experiment, because there were not enough external controls provided to the experimenters and not enough variation in experimental stimuli. As discussed below, subsequent rounds of changes led to an ability to run an initial pilot experiment, but the task turned out to be too easy, with essentially ceiling performance. Each development cycle led to changes not only for development but also for updating experimental controller code.

It is expected that the process and ongoing findings of experimentation, the initial ill-defined requirements and ease of performance and subsequent tweaks leading to a functional system, will be of some interest to an I/ITSEC audience, most of whom are focused on developing not unrelated simulations for military training and assessment. As such, this paper regards simulation development and testing issues of what was wrong or what was unclear early on, what activities were undertaken to overcome issues, how what was done was tested, and future studies that are planned.

BAGGAGE SCREENING SIMULATION

Overview

The simulated environment depicts an airport scene with a baggage conveyor belt in the foreground, a queue of passengers in the centerground, and a stream of semi-ecological distractions in the background (see Figure 1). Each of the items/passengers/events has some preprogrammed attributes and some that are controllable at runtime. The environment is presented...
on a display and the participant’s job is typically to respond with key presses by which reaction time and error rate are measured; no head-mounted displays or other immersion equipment is necessary.

For the baggage conveyor belt, a controller program sends a signal to the environment simulation to add a bag to the conveyor belt. The signal also specifies which of a large library of bag images to display and at what rate the bags are to be released. The rate of movement of the bags on the conveyor belt is also controllable. The conveyor belt has the ability to move backward by a bag to allow for a second look at the last bag. In addition to bags, pairs of shoes, laptop bins, a skateboard, and other items are available to be presented. There is a baggage contents screening camera that is projected on the screener’s display. Currently, each bag has a version with a gun silhouette, versions with several other objects, and one version without any object present.

A set of simulated characters can be called to join the passenger queue by the controller program. Certain elements of the passengers are at the moment preset, such as clothing and facial expression (as they are not components of current tasks), but their styles of movement can vary. It is possible to flag a passenger to be stopped by a simulated Transportation Security Agency (TSA) employee after passing through a metal detector. As with the luggage, there is separate control over adding and removing passengers from the queue. (The passenger queue does not move backward.)

Meanwhile, in the background there is a stream of distractions which are not intended to be task relevant. The onset of these events is initiated by the controller program, and these are fixed-duration events. In addition to these visual elements, a controllable audio track can play announcements, distracting sounds, and general background noise.

The simulation for experimentation has been developed in two forms, both producing the same level of user interaction and feedback. For both forms a server program renders the simulated environment including the security area, luggage, and passengers. The two forms differ in how their client (controller) programs connect to the server: the idea was to enable both standalone operation on standard Windows or Apple operating systems (shown in the figure), either over a network or on a single machine, and program-managed operation to ensure rigorous control over introduction, timing, and behavior of items and events during experimentation. This latter form of operation was used in the pilot experiments described below.

**Details of simulation design, configuration, and implementation**

Before presenting the pilot experiment findings, some detail is given on the design and implementation of the simulation that may be of interest to other researchers.
Initial design decisions
As originally conceived, knowing the intention was to focus on airport baggage screening, the project team envisioned a main view of the display that a screener would normally view, plus peripheral views of typical (persons in the passenger queue putting laptops in bins, taking off coats, placing bins on the conveyor belt) and less typical (kids crying, upset businessman in a hurry, confrontation between other TSA personnel and a passenger) activities. The project team had some discussion whether, for experimentation, this would entail one, two, or even three computer displays, ultimately settling on one. From the beginning the project team wanted variations such as specific appearances of passengers in the queue, specific color bags being placed on the conveyor belt (vigilance to color would be an experimental variable), and specific items in specified bags. To the extent possible, the project team wanted to be able to determine outside of the simulation program itself, via external parameters in an initialization or runtime configuration file, what activities would be upcoming.

During follow-up discussions, the concept was revised to have one central display with the participant’s task on it (e.g., respond using specific actions to certain stimuli) and one peripheral display with the simulated environment, and a line of communication open between programs (i.e., between that managing the task and that rendering the environment). It was anticipated that certain elements of the foreground and centerground might be used as secondary monitoring tasks (e.g., “Press the spacebar if there is a person in a red shirt.”). The primary and any secondary tasks would be controlled by MATLAB, a common off-the-shelf experimental manipulation tool already used by the experimenters in their cognitive lab. The environment would be coded using an open-source platform (initially Flash, later changed to Unity3D) and would be controllable via interpreted code from MATLAB. The virtual environment would be an airport scene with a baggage conveyor belt in the foreground, a queue of passengers in the centerground, and a stream of semi-ecological distractions in the background. Each of the items/persons/events would have some pre-programmed attributes and some that would be controllable at runtime.

For the baggage conveyor belt, the project team envisioned the controller program sending a signal to the virtual environment to add a bag to the conveyor belt and that signal would specify which of a large library of bag images to display. Further, the ability would be present to highlight a bag with a colored or textured overlay (and to specify that highlight color or texture). There would then be a separate signal to remove a bag and shift the remaining bags closer to the participant. These signals could, for instance, couple the removal of a bag with successful performance by the participant on a task, but have the addition of bags be on a different schedule (e.g., one bag every set number of seconds, regardless of performance). The rate of movement of the bags on the conveyor belt would have to be controllable, and would have to have the ability to move backward (as in, “Let’s take a second look at that last one.”) and forward. As first designed the project team sought only bags on the conveyor belt, but later thought to add other items like pairs of shoes and laptop bins.

For the passenger queue, the project team planned for a set of persons that could be called to join the queue by the controller program (“Have person X join the passenger queue.”), preferably with separately specifiable attributes like body type, shirt color, and facial expression, and a range of personalities and behaviors that the persons could display (e.g., obnoxious guy on a cell phone, woman in a rush, or person having trouble passing through the metal detector). The desire was for greater parameterization because signaling from the controller would be eased in contrast to programming dozens of different persons. As with the luggage, the project team envisioned control over adding and removing persons from the passenger queue, but there would be no need to have the passenger queue move backward.

In the background the project team envisioned a stream of task-irrelevant distractions (e.g., have a golf cart with a blinking light populated by old persons drive by as a specific trial started). Further, behind the background layer, there might be a wall enabling colored or textured overlay, with color or texture able to be dynamically set. The onset of these events, figured to be of fixed, known duration, would be controlled as well by the client. The project team envisioned a wide range of events: Groups of persons (Boy Scouts, a K-9 unit, tour groups, families); carts going by (silent, flashing, flashing and noisy, etc.); a person taking flash pictures; emotional outbursts; even, for comparison with results from a different study (Simons & Chabris, 1999), a gorilla that walked out, paused, thumped its chest, and then walked off the other side. In addition to these visual elements, the project team imagined an audio track that could play announcements (boarding calls and security messages such as “Unattended vehicles may be ticketed or towed.”), distracting events (alarms going off), general background noise, or conversations of persons in the passenger queue.
Ongoing design decisions
As development progressed the project team held several interim progress meetings and through these meetings redirected some of the development of the simulated environment. As examples, the project team saw needs such as adjusting the camera position to allow better view of the passenger queue; script control to disallow passengers to walk through each other (the occurrence of which is a key issue in disengagement; Hubal, 2009); adding a capability to pan right to left and up or down; having a longer conveyor belt with luggage having a longer travel animation; creating an animated TSA employee manning a body scanner (standing idle, waving passengers through, or waving passengers to step to the side and idling the passenger queue); and adding red and green lights to the body scanner and enabling the controller program to activate lights turning on, blinking, alternating, or rotating. Among other activities, the project team:

- Scripted luggage to smoothly move along the conveyor belt and not climb on top of one another;
- Created a baggage x-ray that gets projected on the screener’s display and a swappable texture for each bag to appear correctly as it passes through the baggage screener;
- Scripted collision between objects and passengers;
- Blended idle cycles, when characters stopped, and added random behaviors to passengers;
- Scripted the ability to add 3D sounds to pan from left to right;
- Scripted the ability to add individual objects to the background, change out the background (changing travel posters, a background texture swap, was a novel feature), and control the conveyor belt; and
- Created a turn style path for the passengers to walk through before reaching the body scanner.

Further testing yielded timing adjustments. At one point, for instance, the bags could be released once every five seconds and the persons could be released from the passenger queue once every seven seconds. This appeared to be reasonably accurate, but given that one of the tasks given to participants (for this testing, a gun / no gun judgment) was easier than a regular baggage search, pilot participants had been performing too well, so that there was a need to speed up the conveyor belt and passenger queue. The project team thus made the conveyor belt and passenger queue speeds controllable from a command from MATLAB. The project team also added more baggage models and more passenger models.

Later still, it was noted that the x-ray silhouettes for each bag had only two versions, one with a gun and one without. A first research goal was to get participants to miss the gun under certain circumstances, such as having to perform a distractor task, but the pilot testing found the gun / no gun decision to still be too easy for participants since they did not need to sustain attention to the baggage screen display. The conclusion was that if there were an object in every bag and participants had to decide if each object was a gun or not, then that might increase the amount of attention participants had to pay to the baggage screen display. Hence, a number of different items were proposed that could show up in the bags (one item per bag), specified by the controller. Suggestions ranged from a tennis racquet to a rubber duck, recognizable items that would not get confused (unlike, e.g., a drill or hair dryer) with a gun. In the end, the new items created were a syringe, stethoscope, speakerphone, beer bottle (that could be used in later studies of vigilance to banned but non-dangerous items), Walkman/.mp3 player, fire extinguisher (initially red, later made white to match other objects), and computer mouse. To further complicate the participant’s task, new baggage items were created, including a guitar, a skateboard, a backpack, a veggie box, five different shopping bags, a cooler, a melon box, a golf bag, and a boom box.

Client/server interfaces
Some time was spent to both simplify and extend the communication between MATLAB (www.mathworks.com/products/matlab; the controller program) and Unity3D (unity3d.com; the chosen visualization engine). The intent was to be able to run the simulation server in various configurations, including on the same machine running the controller client; on a different machine than that running the controller client, either locally or over a network; or in “demo” (standalone) mode with a graphical user interface for user control over the simulation rather than through MATLAB.

The result is an extensible client/server architecture with the simulation driven by external parameters that other researchers may wish to use for vigilance studies. A Unity3D server renders the simulated environment including the passengers, luggage, security area, and distractions. Two different types of clients, MATLAB for experimentation and Unity for demonstration, connect with and communicate with the server through an IP address, and the latter portrays the graphical user interface and runtime controls for the simulation. The task for the participant and all items/passengers/events...
in the simulated environment are controlled using the client.

- The MATLAB client / Unity3D server configuration requires a Windows-based dynamic link library (.dll) for the client and server to communicate. For this configuration, an initialization file causes the .dll to look on localhost for the Unity3D server, but the server can be moved to any machine (enabling a two-station solution) simply by (i) porting an executable file and data directory to that machine, and (ii) updating the server IP address entry in the initialization file to another IP address or to any resolvable hostname. As stated, the MATLAB client was created specifically because of the desire to rigorously control the introduction, timing, and behavior of the items and events in the simulation, through a preset script, as is needed during experimentation of visual search with secondary tasks.

- The Unity/Unity3D client/server configuration can be run through any browser or as a standalone program on standard Windows or Apple operating systems, either over a network or standalone on a single machine. The Unity client is useful for demonstration and unscripted control; a version was also created to read from an external file to fool a stub .dll into thinking there are two displays and MATLAB (i.e., a two-station Windows environment with a small .dll bridging Unity3D and MATLAB). A graphical user interface provides the means to release baggage or passengers or initiate background events.

Development of all of these components resulted in a highly flexible simulation environment over which either a facilitator or experimenter has detailed control, suitable for demonstration but tailored to the critical timing and scripting demanded by vigilance experiments.

**FIRST PILOT EXPERIMENT**

In the pretesting there were objects in every bag, so that participants should have paid more attention to do the main task; the bags moved relatively slowly and were released less often than needed. The people, similarly, came more slowly and less often than desired. Several pilot experiments have subsequently been run, with two described here.

**Procedures**

**Setup**

Participants were seated at a computer displaying the Unity3D simulation. A second computer, running MATLAB, sent commands over the network to the simulation computer. Participants’ keyboard responses were recorded on this second, control machine. While it is possible to run both the simulation and the control program on the same computer, in pretesting this configuration led to jerky simulation performance and unreliable response timing.

**Calibration**

Because of the open-loop client/server model, the control program only knows the time at which it requested a bag or passenger to be released, but not the time at which the objects reach their target zone. Thus, in order to ascertain correct performance, estimates of the time between controller request and object appearance in the target zone need to be taken empirically.

Four participants from the lab took part in a short calibration experiment. Each object (i.e., all bags and passengers) appeared one at a time and proceeded along its path. No more than one object was on the screen at any time. Participants pressed the space bar when the object first reached the target zone (i.e., became visible on the x-ray screen or passed through the metal detector) and pressed again when the object left the target zone. No target/non-target decision was needed. Averages were taken across these four individuals and some additional time was added to allow for target/non-target decisions. Responses to bags were considered valid for 4.22–5.33 seconds after release; responses to passengers were considered valid for 46.15–55.38 seconds after release.

**Stimuli**

**Bags.** Twelve different bag models were used (e.g., pink backpack, green cooler, black rollerboard, etc.). Bags would travel down a conveyor belt in the foreground of the scene and would pass through an x-ray machine.

A screen on top of the x-ray machine showed a simulated grayscale image of the interior and structure of the bag. X-ray images were artist renderings and were intended to conform to the popular perception of such images. Each bag either contained the silhouette of a handgun or contained nothing.

(It is noted that the visual task is not really like x-ray searches. However, the need for vigilance is similar
between tasks. The distractions being modeled in the background are inspired by the actual distractions baggage screeners would face. Further, real x-ray images were considered, but x-rays require training to read—infeasible to do on a large scale—and use of actual baggage images raises potential regulatory concerns, such as only U.S. citizens being allowed as participants. The key cognitive process being assessed is vigilance, and that can be (and has been) measured with a wide variety of tasks. Vigilance abilities on a simple task inform and predict performance on more complex tasks.)

Passengers. Seven different passenger models were used that spanned a range of ages and ethnicities. Each passenger walked through a switchback queue before passing through a metal detector. No passengers set off the detector.

Conditions
Baseline. Bags came across the conveyor belt toward the scanner. Participants pressed the spacebar if the bag contained a gun silhouette and did nothing if it did not. Gun silhouettes appeared in one-fifth of the bags. Bags were released every 2.22 seconds.

Single-task. Bags came down the conveyor as in the baseline condition and participants responded in the same fashion. Passengers now also passed through the queue, but participants did not have to attend to or respond to the passengers. Passengers were released every four seconds.

Dual-task. At the beginning of the block, one passenger was released in isolation and participants were informed that this was the target passenger for the block and to press the Enter key whenever the target passenger was at the front of the line for the metal detector. The target passenger occurred one-seventh of the time, the same likelihood as any other passenger. The target passenger was visible on the screen several seconds before a response was required. Bags were released and responded to as in the other two conditions.

Predictions
Participants should, in general, do quite well at the baseline, as there is no distraction. The distraction present in the single-task could produce a slight decline in performance compared to the baseline. There should be a major decrement in performance for having to perform two tasks at once in the dual-task condition. Beyond raw performance, there may be individual differences in the performance decrements between conditions. It was predicted that there may be some individuals who do not perform well in the ‘boring’ baseline and easy single-task but may excel at the demanding dual-task condition.

Results
Seven participants who did not perform the calibration procedure took part in this pilot experiment. In short, the experiment was too easy. Four participants made no errors of either omission or commission. Two participants missed one gun silhouette each. One participant made a false alarm to an incorrect passenger.

To make the experiment more difficult so that differences in performance could be observed, several changes were made to both the simulation and the control program.

SECOND PILOT EXPERIMENT

Procedures

Setup
Setup was the same as in the first pilot experiment.

Calibration
The speeds of the bags and passengers were increased, necessitating a new round of calibration. Six participants took part. Responses to bags were considered valid 4.25–6.25 seconds after release; responses to passengers were considered valid 7.5–9.5 seconds after release.

Stimuli
Bags. Eleven bag models were used. Bags had either a handgun silhouette (one-fifth of trials) or one of six other distractor objects (e.g., .mp3 player, megaphone, fire extinguisher). No bags appeared empty.

Passengers. Six passenger models were used.

Conditions
The same three conditions were used. Bags were released every two seconds. Bags traveled 2.3 times faster than in the first pilot experiment. Passengers were released every 2.1 seconds. Passengers traveled three times faster than in the first experiment. Passengers all traveled at the same speed in this experiment, while there had been variation between different passenger models in the first experiment.

Predictions
Predictions were the same as in the first experiment.
Results

Four participants who did not take part in the first experiment or in calibration participated in this pilot experiment. Performance was perfect in the baseline condition. Single-task performance was nearly at ceiling, with only two misses across all four participants. Dual-task performance was generally unimpaired for the baggage task, with rates only slightly higher than the single-task condition (four misses total). Miss errors were now, however, seen in the passenger detection task. Participants were often slower to respond to passengers, not hitting the Enter key until after the target passenger had left the screen.

This experiment, therefore, started to produce errors that would allow the differentiation of good and bad performers. Results are suggestive but not significant, partly due to the low number of participants, but also because the conditions can still be made more difficult. These tests are underway.

SUMMARY

In these pilot experiments results showed some dual-task effects: When participants were searching for both target bags and target passengers they started missing the passengers. Participants were probably prioritizing the bag task (based on the other conditions run with just bags, no passengers). A new pilot experiment is being coded with more balanced expectations to see what happens to baggage and passenger identification accuracy.

Another important modification for a future pilot experiment will be the addition of a single-task block in which participants do not need to respond to baggage, but do need to respond to passengers so that there is not an imbalance between the two tasks. Currently participants favored the baggage task at the expense of the passenger task rather than balancing the two tasks.

In sum, pilot data were successfully collected for a vigilance experiment using the simulation. Participants looked for gun silhouettes in baggage x-rays under three distraction conditions: No passengers, passengers present, and a dual task where they must also respond to a target passenger. In line with previous vigilance work, pilot participants missed more targets when there were passengers present than when there were no such distractions. Further, they missed more still when they were completing a dual task.

Future plans are to collect data and personality measures from participants to see if the differences in accuracy between conditions is lessened or exacerbated by various cognitive and personality measures (see Gunzelmann et al., 2008; McCallum et al., 2005; Shaw et al., 2010). Key traits to be measured are ADHD and autism spectrum symptoms (sub-clinical) and video game and media experience (Clark, Fleck, & Mitroff, 2011).

By providing a richer and more engaging visual scene, the intention is to measure differences in vigilance with a more ecologically valid experimental setup than has been used previously. Other researchers are encouraged to employ the tool in their research.

ACKNOWLEDGEMENTS

The authors wish to thank other members of the Mitroff lab at Duke, and especially thank Ryan DeWitt for his development of the experimental simulation. This project was funded through contract HSHQDC-08-C-00100 by the Institute for Homeland Security Solutions, a research consortium sponsored by the Human Factors Division of the Department of Homeland Security (DHS) established to conduct applied research in the social and behavioral sciences. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the official policy or position of their institutions or DHS. The study is approved for public release.

REFERENCES


