HUMAN PERFORMANCE TESTING FOR COGNITIVE SCIENCE-INFORMED INFORMATION PROVISION FOR INTERNATIONAL NUCLEAR SAFEGUARDS INSPECTORS

Z.N. Gastelum
Sandia National Laboratories
Albuquerque, USA
Email: zgastel@sandia.gov

L.E. Matzen
Sandia National Laboratories
Albuquerque, USA

H.A. Smartt
Sandia National Laboratories
Albuquerque, USA

M.C. Stites
Sandia National Laboratories
Albuquerque, USA

Abstract

International nuclear safeguards inspectors have access to more potentially-relevant safeguards information than ever before. Traditional safeguards data sources including State declarations, previous inspection results, and inspector observations are complemented with myriad open sources including news media, overhead satellite imagery, trade data, scientific publications, and even social media information. However, cognitive science and anecdotal evidence agree that the mere availability of more information is not necessarily useful and can result in confusion, errors, frustration, or other symptoms of information overload. The presentation of safeguards information for inspectors working in the field should enable, rather than distract or overwhelm. If successful, the presentation of information for inspectors working in the field should facilitate more timely, accurate, and situationally aware inspection activities. In the paper, the research team describes human performance studies conducted at Sandia National Laboratories which were informed by research in the domains of cognitive science and international nuclear safeguards. Sandia's human performance experiments targeted three areas: visual inspection, wayfinding, and knowledge transfer for safeguards. The research team will describe the motivation, methods, and results of our initial human performance experiments, and outline proposed follow-on human performance experiments that will allow us to make broader recommendations for information provision for in-field international nuclear safeguards inspections.

1. INTRODUCTION

Due to the global expansion of nuclear fuel cycle activities and the increasing volume of potentially safeguards-relevant information from open sources or from additional reporting under the Additional Protocol, International Atomic Energy Agency (IAEA) nuclear safeguards inspectors have access to more information than ever before. While much of this information may be relevant and useful to safeguards inspectors working in the field, the mere provision of such information to inspectors is not sufficient and may result in information overload. Providing potentially-relevant information to safeguards inspectors working in the field should be done in a way that simultaneously makes the most relevant information available for inspectors and reduces cognitive load from unnecessary tasks.

To understand how to best provide information to safeguards inspectors, a research team at Sandia National Laboratories conducted a task analysis of inspection activities and developed a corresponding information model describing the information available to safeguards inspectors for each task they perform in the field. Then the research team cross-validated the task analysis and information model with a detailed literature review into cognitive science fields relevant to international safeguards inspection activities [1]. The team prioritized the resulting areas of interest and came up with three research themes to explore more deeply using human performance testing: visual inspection, wayfinding, and knowledge transfer. The team used safeguards inspection-like tasks to develop experiments that add to the corpus of scientific knowledge in the cognitive psychology domain and provide insight to the nuances of international safeguards inspection activities [2]. The following sections will provide additional information about current activities and results in each of the three areas, and describe additional work the team would like to pursue.

2. VISUAL INSPECTION EXPERIMENT – LIST COMPARISON

One of the research areas prioritized in this project is visual inspection, specifically related to how inspectors use lists representing state declarations or inventory lists to verify either: other lists (e.g., shipment and receipt records) or physical inventories (e.g., sealed containers). The research team presented participants with lists in different formats to assess how the presentation of the information in the list impacted visual inspection performance.

2.1. Visual Inspection Experimental Methods

The research team has completed the first of three visual inspection experiments. In the first experiment, the researchers presented participants with two electronic lists displayed side-by-side on a wide computer screen. The right list was considered the facility list and did not change in its presentation. The left list was considered the "inspector list" and was presented in six different ways. The six list presentations included each permutation of three order-based schemes (random order, numerical order, or generally matching the order of the facility list) and two color-coding schemes (no color coding, or color coding based on the item's general location within the facility list).). See Figure 1 for examples of the experimental stimuli. In addition to items that directly matched the inspector list, the facility list contained items that differed from the inspector list in various ways, such as: partially correct items (first numerical sequence correct, second alphanumerical sequence incorrect), missing items, items with a transposition of two adjacent numbers without a correct match, and items with a transposition of two adjacent numbers in which there was also an item that represented a correct match.

The participants had two tasks. In the first task, participants compared their "inspector" list to the facility list and marked each item as correct, partially correct, missing, or other. The "other" category was used to account for a suspected transposition of text, among other potential issues. In the second task, participants were asked to identify changes of the background screen color behind the lists. Participants were notified that the color would change between zero to four times for each list presentation activity they completed.

The experimental population included 15 staff members of Sandia National Laboratories with normal or corrected-to-normal vision (three other participants' data was discarded due to poor eye tracking or failure to follow the procedure).

Figure 1.

Figure 1 shows two of the six list presentation conditions as well of two of the background colors that were used in the situational awareness task. The left panel shows the inspector list presented in random order with no color coding. The right panel shows the inspector list presented in numerical order, with color coding that indicates which column each seal should appear in on the facility list.

2.2. Visual Inspection Experimental Results

The research team collected both behavioral data (how fast and where the participants clicked the mouse, the accuracy of their selections) and eye tracking data (indicating search strategy between the lists).

2.2.1. Reaction Time

The visual inspection experiment showed that presentation of the list had a large impact on the time that it took participants to compare the inspector list to the facility list. The color-coded list presentation that indicated the general spatial location where the item should be located on the facility list (which, for safeguards inspections, could be interpreted as a location within a facility, such as a certain material balance area, room, etc.), improved search speed but the largest improvement was seen from providing the two lists in approximately the same order. While knowing the precise order of the facility list is not always going to be possible, in some cases where inventory is relatively static such as long-term storage, these location queues could be significant time savers.

Some inspector lists were presented in numerical order. The numerical order condition is useful for comparing lists only when the participant shifts the search strategy from starting with the inspector list to starting with the facility list to make the comparison. Only four participants showed this change in strategy, as indicated by their eye tracking data and the order in which they marked their list as completed. The participants who took advantage of the numerical order completed their inspections significantly faster than those who did not for this condition.

2.2.2. Accuracy

While improving reaction time is an advantage, doing so without compromising accuracy is an important balance in the cognitive science research community. Becoming faster at a task does not have the same benefit if accuracy falls significantly. There was no significant difference in accuracy among the six list conditions. However, the researchers did observe that in the fastest conditions in which the inspector lists were presented in approximately the same order as the facility lists, the error rate for detection of transposed item numbers was higher, indicating that as participants expected the items on the facility list to match, they were less likely to notice subtle differences.

Participants did not show a significant difference in their ability to detect change in the background color across the six list presentations, with accuracy ranging between 40 and 60 percent. The accuracy numbers largely reflect individual differences, with about half of the test population noticing most of the background color changes, and half of the test population noticing very few to no color changes. This is reflective of other “situational awareness” experiments in which about half of the population notices anomalies that are “hidden in plain sight” [3].

Researchers collected real-time and end-of-block reporting on the number of color changes observed, with the end-of-block reporting showing higher accuracy which can be attributed to misremembering the number of observed changes with a bias towards higher numbers, or participants post-facto realizing they didn’t notice as many color changes as they had hoped and over reporting at the end.

2.3. Visual Inspection Next Steps

In the experiment above, participants were tasked with comparing one list to another, with the expectation that most or all of the items from the inspector list would match items in the facility list. As a second step, the research team is currently deploying an experiment in which the inspector list represents only a fraction of the facility list, such as an inspector would find in a statistical sampling as part of an interim inventory verification rather than the full physical inventory verification. By providing only a portion of the facility list, participants may have to change their task completion strategy.

The research team is currently running two additional visual inspection experiments. The first shifts the in-field safeguards verification scenario from a book audit or other records examination as in the experiment described above to an item verification activity. In this experiment, participants will be presented with different list conditions which they will compare to a series of numbered seals and containers which are presented one at a time. This would be cognitively similar to walking through a room in order to check seals, one at a time. The presentation of one item at a time for comparison to the list is expected to encourage different search strategies. As with the previous experiment, participants will be asked to complete their list verification activity and note color-changes occurring in the background.

In the final visual inspection experiment, the research team is adding an interactive and navigational component in which participants will electronically navigate between rooms to verify single items (sealed containers) against a partial list. In some conditions, an interactive task tracker will indicate which containers in a room have been verified and which remain pending. In this experiment, the list presentation encompasses randomly ordered or ordered-by-room items, the inclusion or exclusion of the room number on the list, and a static map or a dynamic map which shows the containers that have been verified.

3. WAYFINDING EXPERIMENT

The second research focus area identified by the research team is wayfinding. While there have been numerous studies within the cognitive science community about how people navigate both indoors and outdoors using GPS navigation, maps, written or verbal directions, landmarks, and more, the team did not identify any research within the cognitive science corpus relating to a specific condition that safeguards inspectors are often in – escorted indoor navigation. This scenario represents a situation in which a safeguards inspector is escorted from one area to another in a nuclear facility, and the inspector is responsible for knowing if they are in the correct location, and should also be aware if they took a circuitous route or appeared to be avoiding certain areas of the facility. The team identified this escorted navigation activity as the target of a series of human performance experiments in which the participants are provided with one of several map conditions, ranging from having no map, to having detailed or 3D maps with them in the facility.

3.4. Wayfinding Methods

In this experiment, participants were guided on a tour through a complex, industrial facility including two levels/stories, during which time eight landmarks within the facility were pointed out. Each participant received one of three map conditions: the ability to study but not carry a facility map, the ability to study and carry the facility map on the tour, and not having access to a map at all. The map was a simple computer-aided drafting (CAD) drawing of the facility that indicated walls, doors, and other primary structural features without much detail. Following the tour, participants were given a battery of tests to measure their facility, route, and landmark awareness. Previous cognitive science research in the domain of spatial knowledge has identified these three areas as being distinct types of spatial knowledge that a person can develop about an environment, and so the experiment included tasks specifically designed to test each type of knowledge to fully assess the impact of map study on the development of spatial awareness. These tests included a verbal pointing task in which participants described the location of the eight landmarks based on a 360 degree circle (survey knowledge), a shortcut task in which participants were directed to find the shortest path between two of the landmarks which involved accessing parts of the facility not covered in the guided tour (hybrid survey/route knowledge), a landmark recognition test in which participants were asked whether or not they saw 24 items from the facility (only a selection of which were visible from the guided tour; landmark knowledge), and a map completion task in which participants traced the route from the tour onto a map and located the landmarks (hybrid route/landmark knowledge). They were also asked to do a self-assessment of their sense of direction using the Santa Barbara Sense of Direction Scale (SBSOD).

The participant population included 60 Sandia staff members.

3.5. Wayfinding Results

All wayfinding tests resulted in one or more measures of accuracy. The memory test, because it was administered electronically, also resulted in participant reaction times.

3.5.3. Pointing Task

Participants who had no access to a map showed the highest degree of error in the pointing task, with a significant difference between the no map condition and the map carry condition. The participants who studied but did not carry the map had numerically lower error rates than the no map group, but this difference was not statistically significant. These results indicate that participants with access to a map, either to study before the guided tour or to also carry with them on the tour, were more accurately able to point to the landmarks indicated on the tour, indicating better survey knowledge of the building.

3.5.4. Shortcut Task

For the shortcut task, the research team used two measures to define accuracy: the ability of the participant to successfully locate the intended landmark, and difference between the shortest distance between two landmarks and the participant's path. Both of these measures indicated that there was no effect of map condition on error.

3.5.5. Landmark Recognition Test

The landmark recognition test evaluated the accuracy for participants' recognition of landmarks (pictures of the landmarks that the participants were instructed to learn), incidental landmarks (pictures of items in the building that the participants were not instructed to learn) and distractor images that the participants would not have seen during the activity. As might be expected, the accuracy for recognizing landmarks that the participants were instructed to learn was higher than for incidental landmarks across map conditions. While there was no significant difference in accuracy between conditions, the map carry group did have the lowest accuracy scores for incidental landmarks which could indicate that those participants were studying their map rather than looking around the facility.

This trend was also observed in the reaction time results. In general, response times were significantly longer for incidental landmarks than for the learned landmarks, and this was especially so for participants from the map carry condition compared to the no map condition. Pairwise tests showed that response times for the incidental landmarks were significantly longer for the map carry condition relative to the no map condition, and marginally longer for the map study versus no map condition, but there was no statistical difference between the map carry and map study conditions. The large difference in reaction time for the incidental landmarks likely explains the lack of difference in accuracy for the groups, and suggests that the map carry group was less efficient at creating/retrieving memory traces of the incidental landmarks, relative to the groups that were not carrying maps on the route. This suggests lowered situational awareness during the route learning portion when participants could actively refer to their map, indicating a possible tradeoff between more accurate survey knowledge afforded by the map condition and lowered situational awareness. The marginal difference between the no map and map study group may be attributable to the effect of cognitive load and/or attentional allocation during route learning: the map study group may have been attempting to map the building layout onto their mental representation of the map during route learning, which may have lowered their available attentional resources to attend to additional aspects of the environment.

3.5.6. Map Completion

The map completion exercise was scored similarly to the shortcut task, with one score for correctly locating the learned landmarks on the map, and another for the accuracy of the route drawn versus actual route taken. Partial scoring credit was given for marking a landmark name with no location (and vice-versa), and for drawing a correct directional change in a slightly incorrect location (i.e., drawing a turn in the correct direction but one hallway too soon). The results showed no difference in scores between participants from different map conditions for the total map accuracy score and when the route score was considered separately from the condition score.

3.5.7. Self-reported Sense of Direction

The research team hypothesized that participants' self-reported sense of direction via the Santa Barbara Sense of Direction (SBSOD) survey would impact individual performance on the wayfinding tasks regardless of the map condition a participant received. When SBSOD was included as a covariate, the SBSOD had a significant effect on the pointing task, the landmark recognition test (for learned landmarks, but not for incidental landmarks), and the map completion exercise (including both the route and landmark location activities).

3.6. Wayfinding Next Steps

The interim findings suggest that receiving a map to study before and during being guided on a route through an indoor environment can improve one's survey knowledge of the environment, perhaps at the detriment of the individual's situational awareness. The research team is now deploying three additional map conditions for the facility – a highly detailed CAD drawing, a three-dimensional (but printed in two dimensions) map created in the Sketchup software from the simple CAD drawing, and a three-dimensional (but printed in two dimensions) map in color that was created from a point cloud using the FARO Focus three-dimensional scanner. Receiving a more detailed map will help participants develop even better spatial knowledge of the building, but may further harm situational awareness by drawing attention away from building surroundings during navigation.

4. KNOWLEDGE TRANSFER

The third experimental domain for this research is knowledge transfer. Knowledge transfer refers to the transfer of information between individuals or teams. For international safeguards, knowledge transfer is relevant for inspectors and inspection teams to report their activities to headquarters, and provide any relevant information for the next inspection team that will be going to a facility. It is also relevant for passing expertise from staff who are rotating out or retiring to the next generation of safeguards inspectors. While knowledge transfer has been studied closely in the medical community or other professions that require transfer of information in-person or over short time periods, the longer time frame between inspectors working in the field and reporting or sharing their experiences at headquarters makes it a unique field of study.

The research team is currently preparing a series of experiments that will assess how notetaking and other observational methods such as using digital photography or voice recording can support change detection in a complex environment. Participants will be presented with a series of abstract, technology-like computer-generated images and will be asked to record their observations using one or several of the observational methods mentioned above. Then the participants will have a break or unrelated mental task such as putting together a jigsaw puzzle while experimenters change aspects of the computer-generated images (shape, color, texture, shading, etc.). Then participants will be presented with the images again and asked to document any changes in the images. The research team hypothesizes that the different note taking methods used by the participants will impact their ability to quickly and correctly identify modifications to the images.

5. FUTURE WORK

The current research focuses on human performance during safeguards in-field verification activities, using information presentation as the experimental manipulation. This initial foray into human performance testing for international nuclear safeguards has indicated that the presentation of information can have, in certain situations, significant impact of human performance on safeguards-like tasks and has opened multiple potential areas for additional research.

5.1. Three-Dimensional, Virtual and Augmented Data Presentation

The current research examines only two-dimensional information presentations, though some of the pending map experiments are starting to approach three-dimension visualizations. Further work in this area could include digital three-dimensional information presentations such as interactive virtual maps, as well as augmented reality presentations that could impose list presentations or notes onto the real-world environment.

5.2. Analyst Environments

Safeguards inspectors working in the field experience unique environments that have not been previously studied within the cognitive science domain. While the research team has conducted extensive analysis of in-field safeguards verification activities, a similar analysis of the headquarters-based analyst environment including data collection, processing, analysis, and integration across multiple domains has not been conducted. The research team aspires to more closely examine the safeguards analyst environment and identify opportunities for additional human performance testing.

5.3. Experimentation with Experienced Populations

The human performance subjects in this research have come exclusively from the Sandia National Laboratories population. While in some cases Sandia staff members may represent similar educational and professional backgrounds as safeguards inspectors, the research team would like to conduct testing directly with inspector-experienced populations to better tune our understanding to their unique training and biases potentially introduced by those who self-select for safeguards inspector position.

5.4. Solutions and Recommendations

Sandia's current research in the area of cognitive science for international nuclear safeguards focuses on the presentation of information for safeguards inspection activities. The team has reached only limited conclusions on how the results of these experiments might be operationalized into recommendations that can be used directly to inform how safeguards inspectors are trained to use and interact with information. A logical next step to this research is to examine training approaches that can be incorporated into in-field activities to support timeliness, accuracy, and situational awareness of inspectors working in the field.

5.5. Other Nonproliferation Domains

Other domains within the nuclear nonproliferation field might also benefit from this work, including nuclear disarmament, nuclear test bans, or other treaties limiting the scale and scope of nuclear activities in a country. Depending on their unique operational environments, the human performance studies could be modified to examine specific questions or issues resulting from other on-site inspection or analyst environments related to nuclear nonproliferation.

5.6. Response to Direct Requests

Finally, this team hopes to be able to respond directly to statements of concern, issue, or requests that can inform the underlying research questions of this work. The research team previously identified research foci based on the team's assessment of safeguards challenges and their intersection with cognitive science research gaps. However, research might also be directed by specific requests resulting from challenges, errors, or other issues.

ACKNOWLEDGEMENTS

This work was funded by Sandia National Laboratories' Laboratory-Directed Research and Development (LDRD) program. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2018-8395 C

REFERENCES

[1] GASTELUM, Z.N., MATZEN, L.E., SMARTT, H.A., HORAK, K.E., MOYER, E.M., ST. PIERRE, M.E. "Brain Science and International Nuclear Safeguards: Implications from Cognitive Science and Human Factors Research on the Provision and Use of Safeguards-Relevant Information in the Field." ESARDA Bulletin 54 (2017) 62-29.

[2] GASTELUM, Z., MATZEN, L., SMARTT, H., HORAK, K., SOLODOV, A., MOYER, E., ST. PIERRE, M., HAAS, J. “Testing human performance in simulated in-field safeguards information environments (Proc. Inst. Nuc. Mat. Man., Indian Wells, California, USA, 2017)

[3] CHABRIS, C., SIMONS, D. The Invisible Gorilla, Crown Publishers, New York (2010).