Eye Tracking to Evaluate Proficiency of Crime Scene Investigators with Varying Levels of Expertise: The Ability to Recognize Evidence and Understand its Significance

Renuka Devi Watalingam

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation
https://researchrepository.wvu.edu/etd/6921

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact research.repository@mail.wvu.edu.
Eye Tracking to Evaluate Proficiency of Crime Scene Investigators with Varying Levels of Expertise: The Ability to Recognize Evidence and Understand its Significance

Renuka Devi Watalingam

Thesis submitted
to the Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Master of Science in
Forensic & Investigative Science

Jacqueline Speir, Ph.D., Chair
Robert O’Brien, M.S.
Casey Jelsema, Ph.D.

Department of Forensic & Investigative Science

Morgantown, West Virginia
2016

Keywords: crime scene investigation, expertise, experts vs. novices, eye tracking, significance of evidence, interpretation of evidence, proficiency

Copyright 2016 Renuka Devi Watalingam
ABSTRACT

Eye Tracking to Evaluate Proficiency of Crime Scene Investigators with Varying Levels of Expertise: The Ability to Recognize Evidence and Understand its Significance

Renuka Devi Watalingam

Crime scene analysts are the core of criminal investigations; decisions made at the scene greatly affect the speed of analysis and quality of conclusions that directly impact the successful resolution of a case. If an examiner does not recognize the pertinence of an item on scene, it cannot be used to support his or her theory regarding the crime. Conversely, unselective evidence collection will include irrelevant material, thus increasing a forensic laboratory’s backlog and sending the investigation into an unproductive and costly direction. Therefore, it is critical that analysts recognize and properly evaluate forensic evidence that can support or disprove hypotheses and accurately reconstruct events. With this in mind, the aim of the study was to determine if quantitative eye tracking data and qualitative observations could be used to distinguish investigator expertise. In order to assess this, 32 participants were recruited and categorized as experts or trained novices based on their educational and practical backgrounds. Each volunteer then processed a mock crime scene while wearing a mobile eye tracker, wherein visual fixations, duration, search patterns and qualitative reconstruction accuracy were evaluated. Quantitative eye tracking data were compared using pursuit percentage on areas of interest (AOIs), Earth Mover’s Distance (EMD) and the Needleman-Wunsch (N-W) algorithm. Results indicate significant group differences (Mann-Whitney U test, $\alpha = 0.05$) for two out of 14 AOIs selected for analysis in this study. In addition, significant group differences were also detected for both search duration on specific AOIs (EMD), as well as search sequence (N-W), wherein experts exhibited more dissimilar search durations, but more similar search sequences than their novice counterparts (with possible implications regarding hypothesis-based scene reconstruction). Finally, hierarchical and $k$-means clustering based on multivariate AOIs suggest that latent variables may be present, which is the topic of future research. In addition to the quantitative visual comparisons, each participant’s reconstruction skill was assessed using a 22-point binary system. Significant group difference (Mann-Whitney U test, $\alpha = 0.05$) was detected as a function of total reconstruction accuracy. However, the total number of cases processed does not correlate linearly (or well) with total reconstruction score. Equally of interest was the fact that canonical discriminant analysis (CDA) provided good group separability as a function of participant education (which was found to be more important for this cohort than professional development). Overall, results from this study found significant and interesting
group differences, but likewise revealed the complexity associated with using gaze behavior as a means of assessing cognitive processes.
This work is dedicated to my mom and mummy, who have been my support system. This accomplishment would not have been possible without them.
Acknowledgements

I would first like to thank my thesis advisor, Dr. Jacqueline Speir for providing me with both guidance and funding to complete this study. Dr. Speir was very helpful in leading me to find a topic for my research, and her patience over the past one and half year drove me in completing this study. She consistently allowed this paper to be my own work, but steered me in the right direction whenever I was lost for new ideas. Dr. Speir has helped me since year one in getting through not just research, but several other classes, and I am tremendously grateful for her guidance, encouragement and patience.

I would also like to thank Professor Robert O’Brien, the second reader of this thesis, for assisting me in setting up the mock crime scene and providing me with any additional advice regarding the practical aspect of this project. I would also like to thank Dr. Casey Jelsema for his valuable input in this thesis and his statistical advice.

A very special thanks to both my colleagues, Madonna Nobel and Nicole Richetelli, for helping me complete this project. They have relentlessly proofread all my drafts, given useful suggestions and assisted me with coding the algorithms. Without them, I would have not been able to complete this research in a timely manner. I also appreciate the help I have gotten from both my undergraduate students, Zachary Weadock and Gabrielle Marquis for assisting me with data processing for this project.

I would also like to acknowledge Dr. Jeff Pelz from the Center of Imaging Science at Rochester Institute of Technology (RIT) for allowing me to use their eye tracker. I am indebted to his kindness and the initial guidance provided during my preliminary data collection. I am also thankful for the excellent technical support by Jason Babcock from Positive Science, as well as for their willingness to go above and beyond to accommodate software needs of my research.

Finally, I must express my very profound gratitude to my mom, mummy, brother and partner for providing me with a listening ear. Their unfailing support and continuous encouragement were the wind beneath my wings as I completed my coursework and research.

Thank you.
# Contents

1 Introduction ....................................................... 1
   1.1 Education .................................................. 1
   1.2 Training .................................................... 2
   1.3 Certification .............................................. 2
   1.4 Proficiency Testing ........................................ 3
   1.5 Eye Tracking to Assess Proficiency ................. 4

2 Method ............................................................. 6
   2.1 Equipment ................................................... 6
   2.2 Crime Scene ................................................. 7
   2.3 Participants ............................................... 10
   2.4 Procedure .................................................. 12
   2.5 Data Processing ............................................ 13
   2.6 Data Analysis ............................................... 14
      2.6.1 Reconstruction Accuracy .............................. 14
      2.6.2 Pursuit Percentage on AOIs ......................... 15
      2.6.3 A Priori Classification as a Function of Participant Attributes ........................................ 15
      2.6.4 A Posteriori Classification as a Function of Scene Activity ........................................... 16
      2.6.5 Gaze Similarities within Group ..................... 18
      2.6.6 Search Sequence Similarities within Group .......... 20
      2.6.7 Differences between Groups ......................... 21

3 Results & Discussion ........................................... 23
   3.1 Reconstruction Accuracy .................................. 23
   3.2 Agglomerative Hierarchical Clustering ................ 24
      3.2.1 Comparison of Observed and Expected Groups 31
   3.3 K-Means Partitioning ..................................... 31
      3.3.1 Canonical Discriminant Analysis (CDA) .......... 33
   3.4 Direct Comparison of AOIs and Dwell Times .......... 39
   3.5 Earth Mover’s Distance .................................. 41
   3.6 Needleman-Wunsch (N-W) Algorithm ..................... 42
1. Introduction

When a crime is reported, a criminal investigation can be conducted to help reconstruct the possible sequence of events and narrow down or identify the perpetrator(s). This is initiated by the search, recovery and collection of forensic evidence at the scene by trained crime scene investigators. To be able to do this successfully, examiners must assess the significance of materials in order to determine their pertinence to the case. Items deemed significant will be collected and then sent to the appropriate analytical unit for examination in order to glean potential forensic information that can be used in the resolution of the case. Since analytical efforts are costly and time consuming, evidence cannot be submitted indiscriminately; excessive and irrelevant materials can complicate the reconstruction of events and increase backlogs in the laboratory. Conversely, if an item of importance is not collected, it cannot be used to prosecute an individual in court. Thus, the usefulness of the data obtained from analyzing crime scene materials is contingent on the ability of the investigator to recognize the significance of an item relevant to the crime. An examiner who fails to appreciate the relationship between evidence and the case may improperly steer the direction of the investigation, fail to identify an offender, or at worst, fail to exonerate an innocent individual. Therefore, it is essential to understand the sources of variability present among agencies and their investigators, and how this impacts their ability and the quality of investigations.

1.1 Education

Crime scene investigators can include “uniformed officers, detectives, crime scene technicians, criminalists, forensic scientists and coroners” [1]. In the United States, the job responsibilities of a crime scene technician differ among jurisdictions. Hence, the minimum education requirement to become an investigator also varies based on his or her role in the agency. In some jurisdictions, examiners are expected to have a scientific background and be able to process the scene and/or analyze evidence, whereas in other agencies, uniformed officers act as technicians in which their responsibilities are to attend to crime scenes, collect evidence and submit collected items to the forensic laboratory for analysis [1]. In some organizations, a high school diploma qualifies one to become a crime scene investigator, while others require a university degree in forensic science or another science-based field. The National Academy of Sciences (NAS) report from 2009
entitled “Strengthening Forensic Science in the United States: A Path Forward” claimed that this absence of educational standards contributes to varying crime scene practices and reconstruction accuracy in the United States \[1\].

1.2 Training

The National Institute of Justice (NIJ), in its 2004 special report “Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions, and Students” defines training as a “formal, structured process through which a forensic scientist reaches a level of scientific knowledge and expertise required to conduct specific forensic analyses” \[2\]. There are two types of training: (i) on-the-job training by the hiring agency, and (ii) continuing professional development. Initial internal training is mandatory in the forensic community and typically lasts between six months to two years \[2\]. This training is usually peer-based and supervised in which the trainee is taught the standard operating procedures and basic practices of the agency to which he or she belongs. The second type of training, continuing professional development, is optional for the practitioner as it is only required for certification \[2\], which is also non-mandatory. This is unfortunate since pursuing continual professional development is essential for specialization and exposure to advances in technology, new practices and scientific advancements in the industry.

The Technical Working Group on Crime Scene Investigation (TWGCSI) encourages crime scene personnel to develop and continually update their knowledge and skills in terms of scene processing through training; however, there are no documents to address minimum training requirements for examiners \[3\]. Although there are documents that provide training guides, such as “Crime Scene Investigation: A Reference for Law Enforcement Training,” this 2004 report is used merely as a basis to supplement existing training programs in agencies \[4\]. Since there are no standardized training guidelines, the skills developed by crime scene investigators from these programs differ among agencies, thus impacting the quality of investigations in terms of evidence recovery efforts and the reconstruction of events.

1.3 Certification

To become a certified crime scene investigator by the International Association for Identification (IAI), technicians must have completed a minimum of one year in a crime scene-related profession and 48 hours of Crime Scene Certification Board (CSCB)-approved instruction in crime scene related courses within the last five years before application \[5\]. Since the certification is not dependent on education, it is intended to encourage personnel without formal forensic science educational backgrounds to obtain additional training in order to amass a specialized forensic knowledge base by attending scenes and enrolling in crime scene investigation courses in preparation for the certification exam. Although
most agencies recommend that an investigator be certified within the first two years of employment, certification is not compulsory \[1\]. Without certification, an investigator’s skills and knowledge are limited to endorsement by their respective agency through internal proficiency testing. Since different organizations have disparities in terms of testing standards and expectations, the way in which personnel process a crime scene and the quality of work also vary based on agency.

1.4 Proficiency Testing

In forensic science, proficiency testing is crucial to measure an investigator’s performance, which then allows the agency to identify areas for improvement \[6\]. In addition, such performance evaluations can be extended to assess the operations between different agencies \[6\]. Generally, there are two ways in which these tests can be conducted: (i) random case reanalysis, in which a prior case completed by an examiner is randomly re-analyzed by a supervisor, or (ii) blind testing, in which a test case is given to practitioners as if it were a real case \[1\].

Although these approaches may be suitable for forensic practitioners in laboratories, they are not practical for crime scene investigators. Random case reanalysis by a supervisor can only be based on the documentation taken by personnel while on scene. This is not reflective of how searching and recovery of evidence were performed on scene unless the supervisor is physically present on site to audit the examiner. For instance, if there was a relevant item on the scene hidden underneath a bed but missed by the investigator, the supervisor would not know this since it would not be documented, therefore, making case reanalysis ineffective.

On the other hand, current testing methods for investigators, such as "After the Fact," are based on virtual crime scenes in which examiners complete a simulated walk-through, take photographs, record notes, and collect and package evidence before answering questions related to the simulated scene in a separate written test \[6\]. Although the program is designed to be as realistic as possible, it does not completely resemble crime scene investigation as a field science in which analysts are able to interact with and manipulate objects in order to discover and interpret the significance of evidence. Aside from being conducted in a laboratory setting (which does not simulate the tasks involved with crime scene analysis), investigators are also completely aware of being evaluated when using this program, which defeats the purpose of blind testing.

Failure to implement a proper and applicable proficiency testing program aimed at verifying the quality of an examiner’s work in terms of technical skills, knowledge and abilities supports the 2009 NAS report’s claim that several forensic disciplines, including crime scene investigation, do not have “sufficiently rigorous” proficiency testing methodologies \[1\]. A good proficiency testing program is essential for quality assurance in order to monitor and confirm competent performance, and identify areas where improvement may be needed. Despite the clear need, the challenge to develop a mechanism to test crime scene investigators persists.
1.5 Eye Tracking to Assess Proficiency

Given the constant evolution of technology, proficiency testing for crime scene investigators is due for an overhaul. To enhance proficiency testing methods in the forensic field, it is advantageous to explore the manner in which companion fields have tackled similar problems. One source of knowledge is the medical profession due to its close association with the forensic field in terms of the clinical diagnostic process. Several recent studies have shown that eye tracking data such as movements and pupil size can be used to evaluate the proficiency of pathologists or clinicians viewing relevant medical images \[7,8\]. Eye tracking technology has also been widely used in other fields, such as psychology and performance sports, to get a better understanding of human perception and cognition based on gaze behavior. The cumulative result of capturing the eye movements of an observer throughout an activity provides quantitative data of the user’s visual and attentional processes over certain stimuli or target regions in a study, better known as areas of interest (AOIs) \[9\]. Data such as total time spent fixating on an AOI, referred to as dwell time, and pursuit percentage, which is the normalized dwell time, can reveal information related to visual perception (Appendix A). This information translates into an improved understanding of an individual’s visual interest, which can be used to learn and make inferences about the decision making process. For instance, the time spent looking at a certain portion of food packaging by a shopper may indicate what factors play a role in their decision to purchase the product. This insight can help in many ways, such as understanding and improving the user experience when interacting with technologies, increasing sales of a product based on a shopper’s eye movement in the store, and determining key areas of growth and training within professional fields.

In order to demonstrate the utility of eye tracking data for proficiency testing, it is essential to first establish that experts consistently exhibit more efficient and effective gaze behavior than novices. Such studies have been conducted in the medical field, in which experienced surgeons spent more time fixating on task-relevant areas than novices, who split their attention between focusing on targets and incidental tools \[10,11\]. Results from these studies have been further explored in which gaze eye movement data has been used to assist training with virtual reality surgical simulators \[11–13\]. In fact, gaze training has been proven to not only expedite learning \[14\], but also to improve the transfer and retention of surgical skills \[12\].

Although the aforementioned studies indicate that experienced surgeons perform better due to longer fixations on task-relevant areas, the same cannot be said concerning forensic-related tasks. In fact, previous research has failed to consistently find significant differences in gaze behavior between expert and novice analysts. For example, Dyer et al. (2006) conducted a study comparing the abilities of forensic document examiners and lay people to analyze signatures for authenticity \[15\]. Although both groups had similar mean fixations on different areas of interest (AOIs) within the signature, experts demonstrated higher performance \[15\]. The authors concluded that performance of a forensic document analyst is not predicted by the time spent examining a particular feature, but could be dependent
on the expert’s cognitive understanding of the “rarity” and “weight” of certain features present in the signatures [15]. In other words, detailed observation alone does not make one an effective examiner, but careful consideration in ascertaining the importance of a signature’s features is necessary. On the other hand, Busey et al. (2011) found that without time limits, latent print experts had greater variability in visual search patterns and spent more time examining the fingerprints compared to novices [16]. However, in a time-controlled situation, experts were more consistent in fixation patterns as a group, unlike novices [16].

Despite the previous examples, eye tracking has been largely underutilized in forensic settings, and when applied to forensic problems, limited to laboratory applications (such as questioned document and latent fingerprint comparisons) [15][16]. However, technological advancements have resulted in eye trackers that are smaller and more portable, extending their use to field research. Taking advantage of this increased mobility and ease of application, this study explored the utility of eye trackers as a means to study differences between crime scene investigators, including downstream implications regarding the proficiency testing of scene analysts. More specifically, the aim of this research was to evaluate differences in the ability of crime scene examiners to recognize evidence at a scene and understand its significance given variations in educational background and practical experience.
2. Method

2.1 Equipment

The eye tracker used for this study was an UltraFlex DB9-HG-3 Adult Headgear unit with a Mobile Eye Tracking Laboratory (METL) interface from Positive Science, LLC [17]. The unit, as illustrated in Figure 2.1(a), has three main components: (i) the forward facing camera, (ii) the eye camera, and (iii) an infrared LED. The forward facing camera records the scene, whereas the eye camera records the observer’s gaze. Recording of the eye allows the unit to track the pupil location and the reflection of the infrared LED from the user’s eye (Figure 2.1(b)), thereby enabling automatic (but supervised) determination of the observer’s gaze location. The observer’s gaze position can then be merged with the scene video to reveal the user’s continuous point of regard (Figure 2.2) during a specified activity (as documented by a crosshair that is overlaid within the scene video).

Figure 2.1: (a) Close up view of the eye tracker, including the two cameras and LED light. (b) Eye image recorded by camera, highlighting the iris, pupil and LED reflection.
2.2 Crime Scene

This study was conducted at West Virginia University’s Crime Scene Complex, a training facility owned and operated by the University’s Department of Forensic and Investigative Science. A bedroom in one of the houses in the complex was set up to simulate a shooting incident (Figure 2.3).

To assess the examiners’ understanding of the scene and the significance of evidence present, several relevant (Table 2.1) and “noise” items (Table 2.2) were included in the scene. “Noise” items are defined as “forensic-like” objects that may appear to be evidence but carry little to no significance in the case. Additionally, the scene was cluttered with clothes and books to appear lived-in. All items that could either be irrelevant or appear natural were included to make the mock scene as realistic as possible.
Figure 2.3: Sketch of the crime scene.
Table 2.1: Relevant evidence and significance.

<table>
<thead>
<tr>
<th>Relevant Evidence</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cartridge case(s)</td>
<td>Number of shots fired</td>
</tr>
<tr>
<td></td>
<td>Orientation of the shooter when the gun was fired</td>
</tr>
<tr>
<td></td>
<td>Type of firearm</td>
</tr>
<tr>
<td>Blood spatter pattern and bullet lodged in wall</td>
<td>Position of the deceased during the shooting</td>
</tr>
<tr>
<td></td>
<td>Position of the shooter during the shooting</td>
</tr>
<tr>
<td>Bloody footwear (female heels) impression and female heels with blood on the outsole</td>
<td>Indicate that a female was present at the scene either during or after the shooting</td>
</tr>
<tr>
<td>Deceased’s footwear</td>
<td>Suggests that the victim took off his shoes upon entering the room</td>
</tr>
</tbody>
</table>

Table 2.2: “Noise” items and significance.

<table>
<thead>
<tr>
<th>Irrelevant Evidence</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dirty baseball bat</td>
<td>Stain at the tip of the bat that appears to look like dried blood</td>
</tr>
<tr>
<td>A few pairs of footwear</td>
<td>To appear as if the room was lived-in to assess the participants’ ability to distinguish the deceased’s footwear (discussed in Table 2.1 the pair close to the bedroom door) from all the other footwear in the room</td>
</tr>
</tbody>
</table>
The following information was provided to all participants before they were asked to process the scene:

### First Responder’s Notes

**Time dispatched:** 1716 hours  
**Arrival at scene:** 1731 hours  
**Location:** 383 Oakland Drive, Morgantown, WV 26505  
**Person(s) present:** Mr. & Mrs. Anderson

Mr. Anderson (the homeowner) called in the incident. When I arrived at the scene, Mr. Anderson was waiting in the living room. The victim was on the floor of the bedroom in a pool of blood near the door. I secured the crime scene (bedroom on the second floor) with crime scene tape and contacted the crime scene unit. According to Mr. Anderson, nothing was moved or changed at the scene while waiting for the police to arrive. When asked if there were any other eye witnesses, Mr. Anderson said he was home alone when the incident took place. Upon questioning, Mrs. Anderson claimed that she was not present during the incident and arrived home after Mr. Anderson called the police. Mrs. Anderson appeared to be shaken and terrified. Mr. Anderson was calm and collected. The scene was released to the crime scene investigation unit at 1749 hours.

### Homeowner’s statement of the incident

I was home alone asleep when I suddenly heard someone walking up the house stairs towards my bedroom. I glanced at the clock and was very positive that it was not my wife since I did not expect her home yet. It was only 5:00 p.m. and my wife gets out of work at 7:00 p.m.. I reached towards the gun that I always keep in the drawer of my bedside table. Suddenly, a man opened the bedroom door and seemed shocked to see me. The man started running towards me with a knife in hand. I was afraid he would kill me, so I fired a single shot at the man’s chest while I was still in bed. I saw him fall to the floor by the door. I immediately called 911 and reported the incident.

### 2.3 Participants

A total of 37 analysts with various educational and practical experience were recruited for this study (Appendix C). Based upon personal responses to a background survey, each
participant was categorized into one of two major classes (trained novice or expert), each with several subcategories (Table 2.3). Examiners who did not fit into these predefined groups were placed into a subcategory that was the closest fit in terms of their educational and practical backgrounds.

In total, 19 trained novices and 18 experts were recruited. However, due to technical difficulties, videos from 15 trained novices and 17 experts were retained for data analysis.

Table 2.3: Description of the different categories and the corresponding number of participants assigned to each group. The numbers in the bracket indicate the retained participants for each subcategory.

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Description</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trained Novice</td>
<td>Undergraduate Student</td>
<td>Individuals who are currently enrolled in a program offering a bachelor’s degree in forensic science and have had a relevant internship experience</td>
<td>5 (5)</td>
</tr>
<tr>
<td></td>
<td>Graduate Student I</td>
<td>Individuals who are currently enrolled in a program offering a master’s degree in forensic science and have not had any internship experience</td>
<td>5 (3)</td>
</tr>
<tr>
<td></td>
<td>Police Officer I</td>
<td>Individuals who are currently officers in a police department and have attended fewer than 50 crime scenes during their career</td>
<td>9 (7)</td>
</tr>
<tr>
<td>Expert</td>
<td>Graduate Student II</td>
<td>Individuals who are currently enrolled in a program offering a master’s degree in forensic science and have had a relevant internship experience</td>
<td>5 (5)</td>
</tr>
<tr>
<td></td>
<td>Police Officer II</td>
<td>Individuals who are currently officers in a police department and have attended more than 50 crime scenes during their career</td>
<td>7 (6)</td>
</tr>
<tr>
<td></td>
<td>Detectives</td>
<td>Individuals who are currently detectives in a police department and have either attended several crime scene investigation-based courses or received on-the-job training for processing crime scenes</td>
<td>6 (6)</td>
</tr>
</tbody>
</table>
2.4 Procedure

Data collection was completed according to the work flow illustrated in Figure 2.4. Upon agreeing to participate in the study, examiners were asked to complete a survey (Appendix B) to obtain information on their education and experience levels. To maintain the anonymity of the investigators, no identifying information was collected via the survey, and each examiner was assigned an alphanumeric code for identification moving forward.

![Figure 2.4: Steps in data collection.](image)

Participants were then briefed with instructions on scene processing, during which they were encouraged to tag any evidence that he or she would typically collect, and take photographs of any object for personal reference during the post-processing phase. Then, investigators were provided with scene information, including the first responder’s notes and the statement from the homeowner, as previously detailed. After reading the instructions and initial information, participants were asked to wear the eye tracker for pre-scene calibration.

Before processing the mock scene, the eye tracker was calibrated according to each individual’s eye behavior, in order to determine the gaze point (i.e., point on the scene where the examiner is looking) [18]. Since this can only be done when an actual gaze point is known [18], participants were asked to focus their eyes on a circular target, which was placed at eye level, while moving his or her head slowly with pauses in horizontal, vertical and diagonal directions [18]. This was done with the individual located at approximately 30 inches and 130 inches from the target to account for variations in possible viewing distances within the prepared crime scene. Following calibration, investigators were asked to process the mock scene while their gaze behavior was recorded using the eye tracker. No time limit was imposed, and each participant was provided with a basic crime scene kit containing flashlights, rulers, crime scene tags, a DSLR camera, notepad and pens.

Once the examiner finished processing the scene, a second calibration was performed to determine the degree to which the eye tracker had moved while the analyst was performing investigative activities. For the post-processing, participants were brought to a staging area where they were provided with several items from the scene and also clothing from both homeowners. These items could either be relevant (Table 2.4) or “noise.”

The “noise” items (irrelevant evidence) from the scene included wine glasses and a

12
Table 2.4: Relevant evidence made available to the participants after processing the mock crime scene.

<table>
<thead>
<tr>
<th>Relevant Evidence</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homeowner’s shirt</td>
<td>Blood spatter pattern reveals the general position (standing, kneeling, etc.) of the victim when he was shot</td>
</tr>
<tr>
<td>Fiber evidence on the homeowner’s wife’s sweater</td>
<td>Transfer evidence between the wife and the deceased</td>
</tr>
</tbody>
</table>
| Gun | Indicate the number of shots fired if magazine was checked  
Determine direction of shooting based on the ejection port of the gun |

dirty baseball bat. Investigators were given the option to assess the items presented to determine if any additional information could be obtained that might further confirm or refute their hypothesis about the scene.

Once the analyst completed the post-processing phase, they were asked to answer questions (Appendix D) regarding the reconstruction of the scene, using their personal notes and photographs. After answering the questionnaire, a final round of calibration was completed before the eye tracking recording was stopped.

2.5 Data Processing

Data processing for this study was completed in three stages. The first stage was establishing the analyst’s gaze location using Yarbus software version 2.4.3 from Positive Science, LLC [19]. More specifically, this application was used to map the examiner’s eye position to the scene video. Although the software automatically synchronizes the eye and scene videos, several calibration points were manually added to the scene video to ensure an accurate mapping. These points were chosen while the participant’s actual gaze point was known (i.e., when the analyst pauses on the target during the eye tracker calibration process). A minimum of 12 fixations, while the participant was 30 inches away from the target plane, were selected in this process to obtain the most accurate gaze point [20]. Once this step was completed, the gaze location of each participant was displayed as a crosshair on the scene. If the gaze locations were not correctly overlaid on the target during calibration from 130 inches away, additional points were added, as needed. The closer distance was calibrated first because the angle for error is greater when the object is closer. To determine the degree (if any) to which the eye tracker had moved during investigative activity, the location of the crosshair was reassessed during the second and third calibration processes as well. If the crosshair did not appear to be synchronized with the
location of the known gaze point during recalibration, more points were added to adjust for the deviation. In situations where the eye tracker had moved significantly such that adding calibration points could not ensure proper identification of gaze locations (which happened with only two participants in this study), the recorded video was divided at the point where the eye tracker moved, and two separate calibrations were completed for a single participant. The videos with the overlaid crosshair, indicating the area of fixation, were saved as Quicktime movies (H.264 file format) to be used for feature identification. Note that the calibration step was performed by a single experimenter.

The second stage of data processing involved labeling fixations as a function of AOIs using GazeTag software version 0.94 from Positive Science, LLC [21]. Fixations occurring longer than 100 ms in the Yarbus output videos were coded according to what was seen within and around the crosshair. For example, if an investigator fixated on a knife for 120 ms, then the point was labeled as “knife.” The coded video of each participant was then used to assess fixation frequency, dwell time and pursuit percentage on AOIs. Since this phase was more time-consuming, it was performed by two experimenters.

The last stage of data processing was computing the gaze accuracy for each participant using Yarbus software [19]. This was achieved by evaluating the visual angle (Equation 2.1) formed between the (x,y) coordinates of a known target, versus the post-calibration (x,y) coordinates of the scene video crosshairs. Since error is known to vary inversely with the distance between the observer and a known target, 10-15 offsets were computed under worst-case scenario conditions (for a known object situated close (approximately 30 inches) to the observer). The repeated measurements per participant were then averaged to estimate an overall accuracy for a given eye tracking video, and finally, for the entire project.

Error, \( \theta(\circ) = \tan^{-1}\left[ \frac{\text{offset distance}}{\text{distance from the target plane}} \right] \) (2.1)

The average gaze error for this study was found to be 3.77\( \circ \) (SD = 1.89 with a 95% confidence interval that the true value lies between 3.12 and 4.43). Note that this average visual error corresponds to a physical offset of 1.98 inches during close-up inspection of evidence.

2.6 Data Analysis

2.6.1 Reconstruction Accuracy

The answers provided by participants in the questionnaire were used to evaluate the reasoning processes of each participant and estimate a reconstruction score. A point-based system (Appendix E) was implemented such that for every accurate chain of events, one (1) point was awarded, whereas for every inaccurate causal sequence, no points were awarded. For example, if an examiner provided the correct directionality of the shooting based on the location of cartridge cases and the gun’s ejection port, one point was added to his
or her reconstruction score. The total points for each participant were summed to determine their reconstruction accuracy; the higher the points earned, the more accurate the reconstruction. For this study, 22 was the maximum number of points a participant could earn.

2.6.2 Pursuit Percentage on AOIs

By recording the gaze behavior of crime scene investigators on scene, dwell time, defined as the time spent looking at a certain AOI or performing an activity, was available for analysis. Since no time limit was imposed on the investigators during this experiment (where experts on average spent about 26.9 minutes on scene and novices spent 21.0 minutes on scene), dwell times were also normalized by the total time spent performing an activity (Equation (2.2)) to allow for an additional normalized comparison (besides comparing dwell times) between participants. The normalized dwell time will be referred to as pursuit percentage moving forward.

\[
\text{Pursuit percentage (\%)} = \frac{\text{Dwell time}}{\text{Overall time performing an activity}} \times 100
\]  

(2.2)

For this study, pursuit percentage was computed not only for AOIs, but extended to determine the percentage of time spent conducting certain crime scene investigation-related activities (e.g., searching for evidence, taking notes, photographing items of interest, etc.), allowing comparison of the time spent doing different tasks between participants.

2.6.3 A Priori Classification as a Function of Participant Attributes

The participants for this study were placed into intuitive expertise groups based on their responses to the background survey. This grouping was evaluated using Canonical Discriminant Analysis (CDA). The purpose of CDA is to maximize separation between-groups while minimizing within-group variances. This is completed by deriving new variables from a linear combination of the original pursuit percentages on AOIs [22]. First, the total sample matrix and within-class matrix sums of squares were computed using Equation (2.3) and Equation (2.4), respectively, where \( \bar{x} \) is mean of all observations, \( \bar{x}_j \) is mean of the \( j \)th class, \( x_{ij} \) is the \( i \)th observation of the \( j \)th class, \( n_j \) is the size of the \( j \)th class [22] and \( m \) is the total number of classes.

\[
T = \sum_{j=1}^{m} \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2
\]  

(2.3)

\[
W = \sum_{j=1}^{m} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2
\]  

(2.4)

From these, the between-class sum of squares matrix was calculated as shown in Equation (2.5) [22].

\[
B = T - W
\]  

(2.5)
Next, eigendecomposition was performed on $W^{-1}B$ to obtain eigenvalues ($\lambda$) and eigenvectors ($a_1, a_2, ..., a_p$) [22]. The eigenvectors form the canonical discriminant functions, or the coefficients used to define the linear combination of the input (pursuit percentages) according to Equation (2.6), where $X$ corresponds to the pursuit percentages and $p$ is the total number of AOIs of interest [22].

$$CD_1 = a_{11}X_1 + a_{21}X_2 + ... + a_{p1}X_p$$

$$CD_2 = a_{12}X_1 + a_{22}X_2 + ... + a_{p2}X_p$$

(2.6)

The new variables are ranked in descending order such that $CD_1 = a_1^T X$ captures more group differences than $CD_2$, and so on, as a function of the rate at which the eigenvalues decrease [22].

To estimate the classification accuracy associated with using the newly formed canonical discriminant variables, the square of the Mahalanobis distance (Equation (2.7); also referred to as Linear Discriminant Analysis or LDA) between each data point ($x_q$) and both group centroids ($m_i$) was computed, using leave-one-out-cross-validation (LOOCV), and a pooled class covariance matrix (Equation (2.8)), where $n_i$ is the total number of samples in the $i^{th}$ class, $N$ is the total number of samples in all classes, $m$ is the total number of classes, and $W_i$ is the individual class covariance matrix.

$$LDA = (x_q - m_i)S_{\text{pooled}}^{-1}(x_q - m_i)$$

(2.7)

$$S_{\text{pooled}} = \frac{\sum_{j=1}^{m}(n_i - 1)W_i}{N - m}$$

(2.8)

For each computation, the questioned vector was assigned to the class with the smallest LDA value, and a confusion matrix was created to quantify the total accuracy of the classifier using the first two canonical discriminant variables.

### 2.6.4 A Posteriori Classification as a Function of Scene Activity

Besides using a priori information for classification, participants’ pursuit percentage in AOIs were examined on an effort to seek out latent clusters. Several clustering algorithms were employed to assess this.

The first technique used was the agglomerative hierarchical clustering (bottom-up approach), which attempts to find natural clusters in a dataset by successively merging the most “similar” participants (data points) into a single cluster [23]. To compute the similarity between each participant, Euclidean distance (Equation (2.9)) was employed, where $P$ and $Q$ are participants with pursuit percentage on various AOIs represented as $p_1, p_2, ..., p_n$ and $q_1, q_2, ..., q_n$, respectively [23].

$$d(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (p_n - q_n)^2}$$

(2.9)
Since this study had 32 participants, all pairwise distances were calculated for all participants, resulting in a $32 \times 32$ distance matrix, $M_D$.

The distance matrix was then used as input for four different agglomerative techniques: (i) single linkage, (ii) complete linkage, (iii) average linkage and (iv) Ward’s method [23]. In all four methods, the two groups with the smallest distance between them are merged to form a new cluster; the algorithms are differentiated by the criteria used to update the distance matrix [23]. In the single linkage method, also commonly known as nearest neighbor, the distance between two clusters is defined as the minimum distance between any two points in each of the two clusters [23]. On the other hand, with the complete linkage method, commonly known as farthest neighbor, the maximum distance between any two members from different clusters is used as the new distance [23]. The average linkage method approximates the distance between two merged groups as the average of the distance between all points in any two clusters [23]. Lastly, Ward’s method merges groups that minimize the within-cluster (squared) distances [23].

An example of a distance matrix is shown below, where the matrix is symmetric and the diagonals have a value of zero.

\[
\begin{array}{cccc}
A & B & C & D \\
A & 0 & 5 & 9 & 8 \\
B & 5 & 0 & 4 & 5 \\
C & 9 & 4 & 0 & 3 \\
D & 8 & 5 & 3 & 0 \\
\end{array}
\]

To illustrate the nearest neighbor method, first, the closest clusters in the matrix (C and D in this case) are merged to form a single cluster, $C_1$. Subsequently, a new distance matrix is created, where the distance for $C_1$ is represented by the smallest distance between clusters C and D, and the size of the matrix is reduced by one.

\[
\begin{array}{ccc}
A & B & C_1 \\
A & 0 & 5 & 8 \\
B & 5 & 0 & 4 \\
C_1 & 8 & 4 & 0 \\
\end{array}
\]

These steps are repeated until the whole data set is combined into a single cluster [23]. The groups formed in all four cluster analysis methods were compared to the intuitive grouping performed by the experimenter, and independently evaluated in an effort to find latent clusters.

Additionally, a chi-square test of independence ($\chi^2$ significance test with $\alpha = 0.05$) was utilized to determine if the observed frequency of individuals (per hierarchical cluster) was dependent on any background attribute, wherein the expected frequency per hierarchical cluster was computed according to Equation (2.10) where $M_R$ (background attribute) is the row total and $M_C$ (resulting cluster) is the column total for a specific combination of categorical variables, and $n$ is the total sample size [24].

\[
\text{Expected, } E = \frac{M_R \times M_C}{n} \quad (2.10)
\]
Once the expected values have been calculated, the $\chi^2$ can be obtained using Equation (2.11) where $O$ is the observed frequency in cell $i$, $E$ is the expected frequency in cell $i$, and $m$ is the total number of cells [24].

$$\chi^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i} \quad (2.11)$$

Note that the Yates correction [25] was utilized in any scenario where the background attribute and clustering combination created a $2 \times 2$ contingency table (Equation (2.12)).

$$\chi^2_{\text{Yates}} = \sum_{i=1}^{m} \frac{(|O_i - E_i| - 0.5)^2}{E_i} \quad (2.12)$$

The chi-square statistic was computed using the following background categorical variables: (i) expertise level (expert or trained novice), (ii) CSI education (yes or no), (iii) professional training (yes or no), (iv) current academic or employment position (undergraduate student, graduate student I, graduate student II, police officer I, police officer II or detective), (v) number of crime scenes processed (0, 1 to 20, 21 to 50, 51 to 99, 100 to 199, 200 to 299, and 300+), (vi) number of mock scenes processed (0 to 9, 10+), (vii) number of mock scenes created (0 to 9, 10+), and (viii) highest level of education (high school diploma, associate’s degree, bachelor’s degree or master’s degree). Note that the null hypothesis was that the observed and expected groups were not significantly different from each other.

In addition to the agglomerative hierarchical clustering algorithm, a partitioning approach (the $k$-means method) was also used to cluster the data. One advantage of using this method is that it allows participants to move between clusters, if doing so is beneficial to finding latent groups, which is not an option in hierarchical methods [23]. This algorithm was initiated by selecting $k$ centroids from the Ward’s hierarchical method (to serve as “seeds”) [23]. Using the centroids as seeds, participants were assigned to the closest seed (using Euclidean distance as described in Equation (2.9)) [23]. Upon addition of a member to a cluster, the centroid of each new “seed” was recalculated as the average of all data points in that group [23]. After all the participants were assigned to one of the $k$ groups, the members of the groups were re-assessed to determine if they should be moved to another group. Each time a new member is added or removed from a cluster, the group centroid is recomputed. These steps were repeated until no further reassignments occurred [23]. The group assessments were then transformed using canonical discriminant analysis (CDA), allowing for a visual interpretation of the results.

### 2.6.5 Gaze Similarities within Group

To further evaluate the similarity of the examiners, Earth Mover’s Distance (EMD), which is a measure of distance between two distributions in some space [26] was computed. This
metric was selected as it considered both individual pursuit percentage and location of the AOIs when assessing within- and between-group similarity.

The Earth Mover’s Distance (EMD) metric by Rubner (1999) is based on a proposed solution for a transportation problem \[26\]. The computation involves finding an optimal flow between two distributions (suppliers and consumers) to satisfy all of the customer’s requests with the least expensive amount of work.

More specifically, let \( P \) be the first distribution, or suppliers, with an amount of goods of weight \( w_i \) at locations \( p_i \) where \( i = 1, \ldots, m \) and \( Q \) be the second distribution (the consumer’s demands) with a need for an amount of goods \( w_j \) at locations \( q_j \), where \( j = 1, \ldots, n \) \[26\].

\[
P = \{(p_1, w_{p_1}), (p_2, w_{p_2}), \ldots, (p_m, w_{p_m})\}
\]
\[
Q = \{(q_1, w_{q_1}), (q_2, w_{q_2}), \ldots, (q_n, w_{q_n})\}
\]

The minimum work needed to fulfill the consumers’ demands given a number of suppliers can be calculated using Equation (2.13), where \( F \) is the amount of goods transported from \( p_i \) to \( q_j \) called the flow between \( p_i \) and \( q_j \) and denoted as \( f_{ij} \), while \( d(p_i, q_j) \) is the ground distance between distributions \( p_i \) and \( q_j \) \[26\]; in most scenarios, Euclidean distance is valid \[27\].

\[
\text{WORK}(P, Q, F) = \sum_{i=1}^{m} \sum_{j=1}^{n} d(p_i, q_j)f_{ij} \tag{2.13}
\]

\[
F = [f_{ij}]
\]
\[
D = [d_{ij}] = d(p_i, q_j)
\]

Equation (2.13) is subject to four constraints. Constraint 1 limits movements from distributions \( P \) (suppliers) to \( Q \) (consumers) and not vice versa \[26\]. Constraints 2 and 3 restrict the amount of goods transported by distribution \( P \) to its weights and the amount of supplies received by distribution \( Q \) to its weights, respectively \[26\]. Finally, constraint 4 controls the maximum amount of goods that can be transported to consumers, which is also known as the total flow \[26\]. The EMD can then be modified as the ratio of the resulting work, normalized by the total flow, as shown in Equation (2.14) \[26\]. Hence, the larger the EMD, the more work needed to attain the optimal flow.

\[
\text{EMD}(P, Q) = \frac{\sum_{j=1}^{n} \sum_{i=1}^{m} d(p_i, q_j)f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} \tag{2.14}
\]

The applicability of EMD can be extended to any question that mirrors the transportation problem. This study adapted the use of EMD to assess dissimilarities between gaze behaviors by using one individual’s fixation distribution as the suppliers and another individual’s fixation distribution as the consumers \[16\] \[27\]. For this research, EMD was assessed in the feature domain, where pursuit percentage on AOIs were used as the weights, and coordinates of AOIs were used as locations for the fixation distribution. EMD was computed as shown in Equation (2.14) for all pairwise comparisons between

\[19\]
the 32 participants. Intuitively, the smaller the EMD, the more similar the two fixation distributions being compared.

2.6.6 Search Sequence Similarities within Group

Despite fixations being fundamentally sequential, where one fixation is followed by another, the EMD metric only considers the locations and duration of these fixations. In order to evaluate the similarities in search sequences of participants in this study, a robust algorithm that accounted for temporal and spatial features, as well as sequential information, was required. For this purpose, the Needleman-Wunsch (N-W) algorithm, was implemented, borrowing from the field of bioinformatics where it is typically used to analyze the similarity of biological DNA sequences [28].

To utilize this sequence alignment algorithm, the fixation sequence, locations and pursuit percentages must be converted to a character string, thus mimicking the structure of a DNA sequence. To accomplish this, letters were assigned to the AOIs and pursuit percentages were binned, such that the length of a repeated string directly corresponded to a fixation duration [28]. For example, a pursuit percentage on location “A” for 4.2% was denoted as a character string of length four or “AAAA”. By concatenating AOI strings, the eye movement of participants on scene was transformed into a unique character string, much like a DNA sequence. For example, Figure 2.5(a) can be seen as a search sequence with A, B and C representing the locations of AOIs and the diameter of the circles proportional to the duration of each fixation. If the duration on A, B and C were 4%, 1% and 2%, respectively, then the first participant’s sequence can be represented as “AAAABCC” and the second participant’s sequence as “ACCCCBB” (given durations of 1%, 2% and 4%, respectively, on A, B and C).

![Figure 2.5: Search sequences of two participants, where the letters represent AOI location, arrows as the sequential information and the diameter of the circles as the duration of fixations on AOIs.](image)

The two sequences can be represented in a two-dimensional array, where one sequence is the row and the other is the column. The similarity of two fixation sequences are then
scored using the Needleman-Wunsch algorithm, which seeks the optimal alignment by tracing the array from the top left corner to the outermost column or row, as a function of: (i) a substitution matrix and (ii) a penalty gap. The substitution matrix provides a score for aligning two letters. For this study, the inverse of Euclidean distance between AOI locations constituted the substitution matrix. Therefore, aligning “A” and “A” gives a score of 1, whereas AOIs that were spatially furthest apart were given the lowest score (e.g., based on the sequences illustrated in Figure 2.5, aligning “A” and “C” will give a score that is bigger than aligning “A” and “B”). On the other hand, the gap penalty provides a score for aligning any letter to a gap in the sequence, wherein a negative penalty was implemented to discourage gaps. The total alignment score for the comparison is based on the optimal alignment of two strings, where all substitution scores and gap penalty scores are summed. An example of computing scores for a simple alignment is illustrated below, where the first row and first column are the sequences, and each cell of the matrix reflects the substitution score associated with aligning the row and column index. The final score is the optimal flow, which translates into the maximum score. For this example, the maximum score is 6.1, which happens to be the primary diagonal of this matrix.

\[
\begin{array}{ccccccc}
A & 1 & 1 & 1 & 1 & 0.2 & 0.8 & 0.8 \\
C & 0.8 & 0.8 & 0.8 & 0.8 & 0.9 & 1 & 1 \\
C & 0.8 & 0.8 & 0.8 & 0.8 & 0.9 & 1 & 1 \\
C & 0.8 & 0.8 & 0.8 & 0.8 & 0.9 & 1 & 1 \\
C & 0.8 & 0.8 & 0.8 & 0.8 & 0.9 & 1 & 1 \\
B & 0.2 & 0.2 & 0.2 & 0.2 & 1 & 0.9 & 0.9 \\
B & 0.2 & 0.2 & 0.2 & 0.2 & 1 & 0.9 & 0.9 \\
\end{array}
\]

\[(1 + 0.8 + 0.8 + 0.8 + 0.9 + 0.9 + 0.9) = 6.1\]

Since the algorithm is dependent on the length of the sequences, the scores were normalized using Equation (2.15).

\[
\text{Normalized score} = \frac{\text{alignment score}}{\max (\text{substitution matrix}) \times \text{length of the longest sequence}} \quad (2.15)
\]

After normalization, two strings that are exactly alike will be given a score of 1, whereas sequences that are different will be less than 1.

### 2.6.7 Differences between Groups

The Mann-Whitney U test, with $\alpha = 0.05$, was utilized to determine whether or not significant differences existed between groups based on intuitive clustering. This was evaluated as a function of the following: the medians of the (i) overall time taken to
complete processing the scene, (ii) dwell time and pursuit percentage on each AOI, (iii) dwell time and pursuit percentage on all AOIs, (iv) dwell time and pursuit percentage on all non-AOIs, (v) percentage of time spent documenting (i.e., taking notes, photographs, etc.,) on scene (vi) percentage of time spent searching for evidence, (vii) EMD scores and (viii) N-W scores. Note that the null hypothesis was that the groups were not significantly different from each other.
3. Results & Discussion

3.1 Reconstruction Accuracy

Based on the answers provided by each participant in the questionnaire, a reconstruction score was estimated. The scores ranged from 1 to 16, out of 22 possible points for each individual. For the trained novices (based on the intuitive grouping), the lowest score was 1 and the highest was 11 with a mean score of $7 \pm 3$ (median = 7.0) and a 95% confidence interval between 5 and 8 (Figure 3.1). On the other hand, the experts' scores ranged from 1 to 16 with a mean score of $10 \pm 4$ (median = 10) and a 95% confidence interval between 8 and 12 (Figure 3.1). As expected, there were experts who performed poorly and trained novices who scored higher than anticipated. The Mann-Whitney U test on the scores gave a p-value of 0.03 ($< \alpha = 0.05$), which indicated there was a significant difference in the group medians. Therefore, it can be said that experts, as a group, (on average) performed significantly better than their counterparts with less experience.

Figure 3.1: Boxplot illustrating the distributions of the reconstruction scores for each group.
Although it was hypothesized that experience would directly relate to ability, reconstruction accuracy was not linearly correlated ($R^2 = 0.08$) to the total number of scenes an examiner processed (Figure 3.2). However, it is acknowledged that other relationships (such as a logarithmic increase) were not explored. Moreover, a metric like the total number of cases analyzed has no regard for “case-type” (such as shooting reconstruction), which is likely more important than total number of scenes attended to.

Figure 3.2: Scatter plot illustrating the distributions of the reconstruction scores of participants and the number of scenes they have processed.

When processing a crime scene, there are two aspects that allow for better understanding of the chain of events that took place: (i) detection of evidence and (ii) appreciation of its significance. Although detection of evidence plays a huge role in the reconstruction ability of an analyst, detection must be followed by comprehension of the implications associated with the evidence being scrutinized. For instance, almost every participant in this study located at least one of the two cartridge cases on the scene; however, only 33% of trained novices (officers who have experience carrying and handling firearms) provided the correct direction of shooting, compared to 77% of experts. Naturally, this indicated that finding evidence does not necessarily lead to an accurate deduction of the significance of the evidence.

3.2 Agglomerative Hierarchical Clustering

Hierarchical clustering for this study was based on pursuit percentage on AOIs, such that participants who were most “similar” in the amount of time they fixated on AOIs were clustered together. Since this study divided the participants into two main categories: experts and trained novices ($k = 2$) with 3 subcategories each ($k = 6$), the dendrograms
produced by the algorithm were cut into either 2, 3, 4, 5 or 6 groups depending on the optimal clustering output per method. It was expected that the clusters would not mirror the intuitive classifications exactly since this latter grouping was based on a priori information (background of participants, but lacking clear thresholds) and not their behavior while processing the scene; however, the resulting dendrograms were inspected to see if any trends in participants’ background information prevailed.

Figures 3.3 and 3.4 illustrate the dendrograms produced using single linkage with $k = 2$ and $k = 6$, respectively (the labels on the ends of the dendrograms are the participants’ code, whereas the numerical value in parenthesis is their individual reconstruction score). The single linkage method performed sub-optimally as it was only able to differentiate a single participant, out of the 32, to form two clusters. Besides that, increasing the number of clusters (instead of $k = 2$) for this method only allowed for a successive separation of a single participant into a new cluster of his or her own. Moreover, there were no clear trends found in the clustering method based on the individual’s background (e.g., number of scene processed, highest level of education, etc.). Therefore, this method was deemed to be ineffective for the purposes of assessing possible group differences.

Figure 3.3: Dendrogram illustrating the groupings formed using the single linkage hierarchical clustering algorithm ($k = 2$).
Figure 3.4: Dendrogram illustrating the groupings formed using the single linkage hierarchical clustering algorithm ($k = 6$).

On the other hand, the complete linkage (Figure 3.5) and average linkage (Figure 3.6) methods were very similar in the resulting groupings, with the exception of a single participant (1att9). Increasing the number of groups ($k = 6$) for the average linkage method further separates the existing group into smaller clusters and extracts participant 1att9 out of the bigger cluster (Figure 3.7). All of the participants who were placed into this smaller group had some form of crime scene investigation training (either on-the-job or from coursework) except for participant 1att9. Once again, no absolute trends as a function of participants’ background were identified.

Since the most obvious evidence at the scene was the blood spatter pattern, both of the clustering algorithms were strongly impacted by the pursuit percentage on this particular AOI. More specifically, the participants in the smaller groups were placed together due to the high pursuit percentage on the blood spatter pattern on the floor near the bed, as compared to the investigators in the other group. Having a high pursuit percentage on a particular AOI directly increases the pursuit percentage on total AOIs, which was also one of the reasons the groups were clustered together. Participants in the smaller group spent at least 5% of their time on the scene looking at various relevant evidence, including the blood spatter pattern, while members of the other cluster spent less than 5% of their time looking at the AOIs. Although this may imply that the participants who were clustered together behaved very similarly at the scene, the reconstruction score showed that fixation on individual AOIs is not a good indicator of the analysts’ proficiency in interpreting evidence. Nonetheless, this further supports that detection and observation of
evidence alone do not ensure a more successful reconstruction; instead, it is the expertise of an investigator and their ability to discern the meaning of evidence that drives an investigation.

Figure 3.5: Dendrogram illustrating the groupings formed using the complete linkage hierarchical clustering algorithm \( (k = 2) \).

Increasing the number of groups \( (k = 6) \) for the complete linkage method further separates the smaller group, but also managed to sort the bigger cluster into two smaller groups (Figure 3.8). Once more, no distinct trend in the group members’ background was found, but the seven participants in this (green) cluster spent about 3.5% - 4.8% of their time on the scene looking at various evidence. Members of this group also had varying reconstruction scores, providing additional support to the previous statement that detection and observation of evidence alone do not ensure a more successful reconstruction.

The last agglomerative hierarchical clustering technique used in this study (Ward’s method) produced the dendrograms showed in Figures 3.9 and 3.10. The method was the most successful in producing an almost even grouping when \( k = 2 \) (Figure 3.9). The results of cutting the tree into six groups (Figure 3.10) were quite similar to the complete linkage method (Figure 3.8), with minor changes in the members of the groups. Nevertheless, even with these changes, a significant trend was not apparent; the only similarity among members of the green group is that all of them had some form of crime scene investigation...
Figure 3.6: Dendrogram illustrating the groupings formed using the average linkage hierarchical clustering algorithm ($k = 2$).

Figure 3.7: Dendrogram illustrating the groupings formed using the average linkage hierarchical clustering algorithm ($k = 6$).
training. However, almost all participants recruited in this study had different levels of training ranging from one-day classes to three years of coursework related to crime scene investigation (along with practical training). Therefore, it was not possible to narrow down the groupings based on training alone.

Since the intuitive classification was performed using the top-down approach (expertise background), as opposed to the agglomerative clustering method that used the bottom-up approach (eye tracking data), the difference in grouping was anticipated. Nevertheless, some form of trend (i.e., participants with similar experiences) among the participants was hoped for within clusters.
Figure 3.9: Dendrogram illustrating the groupings formed using the Ward’s hierarchical clustering algorithm ($k = 2$).

Figure 3.10: Dendrogram illustrating the groupings formed using the Ward’s hierarchical clustering algorithm ($k = 6$).
3.2.1 Comparison of Observed and Expected Groups

The chi-square test of independence was performed to determine if background attributes (Table 3.1) could explain hierarchical clustering using Ward’s and the complete linkage method with \(k = 2\).

Table 3.1: Summary of the p-values obtained for each of the comparison of the frequencies obtained between the observed and expected groups.

<table>
<thead>
<tr>
<th>background attribute</th>
<th>p-value using complete linkage method</th>
<th>p-value using Ward’s method</th>
</tr>
</thead>
<tbody>
<tr>
<td>expertise level</td>
<td>0.50</td>
<td>0.62</td>
</tr>
<tr>
<td>CSI education</td>
<td>0.93</td>
<td>0.82</td>
</tr>
<tr>
<td>professional training</td>
<td>0.80</td>
<td>0.51</td>
</tr>
<tr>
<td>current academic or employment position</td>
<td>0.45</td>
<td>0.93</td>
</tr>
<tr>
<td>number of crime scenes processed</td>
<td>0.35</td>
<td>0.90</td>
</tr>
<tr>
<td>number of mock scenes processed</td>
<td>0.93</td>
<td>0.17</td>
</tr>
<tr>
<td>number of mock scenes created</td>
<td>0.81</td>
<td>0.54</td>
</tr>
<tr>
<td>highest level of education</td>
<td>0.17</td>
<td>0.54</td>
</tr>
</tbody>
</table>

The p-values obtained do not show statistical dependence between participant background attributes (as measured according to Table 3.1) and the resulting hierarchical clusters based on AOIs.

3.3 \(K\)-Means Partitioning

The \(k\)-means method (Figure 3.11) produced results that were very similar to the complete linkage method (Figure 3.5), with the addition of a single participant. All of the participants who were clustered together in group one, once again, had some form of crime scene investigation training (either on-the-job or coursework) except for participant 1att9. On the other hand, using \(k = 4\) allowed for a different grouping, where the 4th cluster consisted of mostly detectives (three out of four members), as illustrated in Figure 3.12. Interestingly, the second cluster contained participant 1att9 (located at -2.0, -3.5), rather than grouping this individual with cluster three. Although the groups clustered better, even with this method, there were no notable trends found in terms of the participants’ background, just as discussed in the hierarchical clustering results.
Figure 3.11: Clusters formed using Ward’s $k$-means algorithm ($k = 2$).

Figure 3.12: Clusters formed using Ward’s $k$-means algorithm ($k = 4$).
3.3.1 Canonical Discriminant Analysis (CDA)

An eigendecomposition (assuming symmetric is true) on $W^{-1}B$ of the AOI data for all participants produced eigenvalues $\lambda_1 = 1.26$ and $\lambda_2 = 3.52e-16$, which indicated that the first eigenvector represented more than 99.9% of the differentiability of the two groups. Inspection of the first eigenvector indicated that detection of the blood swipe pattern underneath the deceased, the blood spatter pattern on the bedside table, the bullet in the wall and the blood spatter pattern on the wall near the bed played a large role in distinguishing participants based on their assigned expertise, as illustrated in Figure 3.13.

![First Eigenvector](image)

Figure 3.13: Breakdown of the first eigenvector, where cartridge 1 = cartridge casing near the book rack, cartridge 2 = cartridge near the bed underneath the comforter, floor = blood spatter pattern on the floor, bedsheets = blood spatter pattern on the bed sheets, book = blood spatter pattern on book found on the bedside table, pot = blood spatter pattern on flowerpot on top of bedside table, table = blood spatter pattern on the bedside table, pillow = blood spatter pattern on pillow, wall = blood spatter pattern on wall near the bed, bullet = bullet lodged in wall near bed, GSW leg 1 = bullet entrance wound on deceased’s leg, GSW leg 2 = bullet exit wound on deceased’s leg, GSW chest = bullet entrance wound on deceased’s chest and blood swipe = blood swipe pattern underneath the deceased’s body.
The first two canonical discriminant functions, $CD_1$ and $CD_2$, were used to create new variables that were plotted, as shown in (Figure 3.14). Since CDA attempts to find separation by maximizing the between-sample variance and minimizing the within-sample variance, the algorithm was able to largely separate the groups based on expertise as assigned through intuitive classification. Most experts (group one in Figure 3.14) detected the blood spatter pattern on the wall (88.2%) and additional evidence in this vicinity (e.g., bullet in the wall (58.8%)). Conversely, 66.7% and 40.0% of the novice counterparts detected the blood spatter and bullet on/in the wall, respectively.

**Figure 3.14: Plot of $CD_1$ vs $CD_2$. The group numbers indicate the intuitive expertise groups that the participants belong to, where group one are experts and group two are trained novices.**

Moreover, CDA was repeated using additional background attributes as demonstrated in Appendix C. Figures 3.15 and 3.16 illustrate the canonical discriminant plots based on CSI education (internship) and professional training. Both plots show good separability with the exception of a few participants who appear further away from their assessed groups. Leave-one-out-cross-validation using LDA as a function of CD1 and CD2 provided classification accuracies of 81.3% and 53.1% as a function of CSI education and professional training, respectively. This suggests that primary education may play a larger role than professional development in terms of class separation, but this claim is limited because (i) the cohort is not random (all experts were sampled from a fixed (and limited) geographical location and a large percentage of the trained novices are the product of a single curriculum (university) of study), and (ii) the majority of individuals that have acquired professional development also have primary education in forensic science.
Figure 3.15: Plot of $CD_1$ vs $CD_2$. Group one has no formal university-based CSI education, while group two has some form of CSI education (including university coursework and internship).

Figure 3.16: Plot of $CD_1$ vs $CD_2$. Group one has no formal CSI professional development courses or on-the-job training, while group two has some form of professional development/training.
An eigendecomposition (assuming symmetric is true) on $W^{-1}B$ of the reconstruction scoring checklist (Appendix E), using intuitive expertise for classification, produced eigenvalues $\lambda_1 = 3.11$ and $\lambda_2 = 2.72e^{-15}$, which indicated that the first eigenvector represented more than 99.9% of the differentiability of the two groups. Inspection of the first eigenvector indicated that ability of an investigator to determine the shooting direction (coefficient #22) drove the discrimination of the two groups, as illustrated in Figure 3.17. The first two canonical discriminant functions, $CD_1$ and $CD_2$, were used to create new variables that were plotted, as shown in (Figure 3.18).

![First Eigenvector](image)

**Figure 3.17**: Breakdown of the first eigenvector, where the numbers indicate the reconstruction scoring checklist as listed in Appendix E. Note that coefficient #22 describes the investigator’s ability to correctly localize the shooter and his direction of fire.

Lastly, CDA was also performed on the reconstruction scoring checklist (Appendix E), but using additional background attributes for class membership. Although 76.5% of experts and 40.0% of trained novices were able to accurately provide the shooting direction, it is also worth noting that the separability of the groups was influenced by the investigators’ ability to detect additional items of evidence on scene (e.g., bullet in wall (#15), 58.8% for experts and 20.0% for novices) and realizing that the gun ejected casings toward the right (question (#17), 41.0% for experts and 20.0% for novices). High group separability was also achieved using CSI education (internship) and professional training, as illustrated in Figures 3.19 and 3.20, respectively. Leave-one-out-cross-validation using LDA as a function of CD1 and CD2 provided classification accuracies of 96.9% and 79.1% as a function of CSI education and professional training, respectively.
Figure 3.18: Plot of $CD_1$ vs $CD_2$. The group numbers indicate the intuitive expertise groups that the participants belong to, where group one are experts and group two are trained novices.

Figure 3.19: Plot of $CD_1$ vs $CD_2$. Group one has no formal university-based CSI education, while group two has some form of CSI education (including university coursework). Classification as a function of internship experience mirrors this distribution of the data.
Figure 3.20: Plot of $CD_1$ vs $CD_2$. Group one has no formal CSI professional development courses or on-the-job training, while group two has some form of professional development/training.

Interestingly, the reconstruction scoring checklist was also able to separate the groups into subcategories as described in Table 2.3 based on their current academic and employment position (Figure 3.21).

Figure 3.21: Plot of $CD_1$ vs $CD_2$. Group one = detectives, group two = graduate students I, group three = graduate students II, group four = police officers I, group five = police officers II and group six = undergraduate students.
3.4 Direct Comparison of AOIs and Dwell Times

The Mann-Whitney U test, with $\alpha = 0.05$, was performed on pursuit percentage and dwell time for each AOI, as well as crime scene-related activities assuming an intuitive classification scheme based on a priori participant attributes (Tables 3.2 and 3.3).

Table 3.2: Summary of the p-values obtained for each of the pursuit percentages (%) on AOIs based on intuitive groupings. Asterisk (*) indicates the AOIs that had a significant difference.

<table>
<thead>
<tr>
<th>AOIs</th>
<th>mean of experts (%)</th>
<th>median of experts (%)</th>
<th>mean of novices (%)</th>
<th>median of novices (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>cartridge casing near book rack</td>
<td>0.57</td>
<td>0.49</td>
<td>0.47</td>
<td>0.25</td>
<td>0.24</td>
</tr>
<tr>
<td>cartridge casing near bed underneath comforter</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.05*</td>
</tr>
<tr>
<td>bloodspatter pattern on floor</td>
<td>1.04</td>
<td>0.54</td>
<td>0.94</td>
<td>0.60</td>
<td>0.31</td>
</tr>
<tr>
<td>bloodspatter pattern on sheets</td>
<td>0.56</td>
<td>0.56</td>
<td>0.59</td>
<td>0.58</td>
<td>0.56</td>
</tr>
<tr>
<td>bloodspatter pattern on book</td>
<td>0.20</td>
<td>0.13</td>
<td>0.14</td>
<td>0.14</td>
<td>0.65</td>
</tr>
<tr>
<td>bloodspatter pattern on pillow</td>
<td>0.33</td>
<td>0.20</td>
<td>0.19</td>
<td>0.09</td>
<td>0.27</td>
</tr>
<tr>
<td>bloodspatter pattern on bedside table</td>
<td>0.13</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00*</td>
</tr>
<tr>
<td>bloodspatter pattern on wall near bed</td>
<td>0.34</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00*</td>
</tr>
<tr>
<td>entrance GSW on leg</td>
<td>0.41</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.41</td>
</tr>
<tr>
<td>bullet GSW on leg</td>
<td>0.19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.14</td>
</tr>
<tr>
<td>entrance GSW on chest</td>
<td>0.09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.14</td>
</tr>
<tr>
<td>bullet shot underneath decreased</td>
<td>0.30</td>
<td>0.11</td>
<td>0.12</td>
<td>0.10</td>
<td>0.04</td>
</tr>
<tr>
<td>all AOIs</td>
<td>4.78</td>
<td>4.76</td>
<td>3.90</td>
<td>3.94</td>
<td>0.06*</td>
</tr>
<tr>
<td>all non-AOIs</td>
<td>87.7</td>
<td>86.4</td>
<td>86.2</td>
<td>87.6</td>
<td>0.80</td>
</tr>
<tr>
<td>documenting (taking notes and photographs)</td>
<td>7.52</td>
<td>6.21</td>
<td>10.3</td>
<td>9.50</td>
<td>0.18</td>
</tr>
<tr>
<td>searching for evidence</td>
<td>99.5</td>
<td>91.3</td>
<td>30.7</td>
<td>30.0</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Although experts on average spent about 1.47 minutes (median = 1.17 minutes) compared to novices 0.79 minutes (median = 0.64 minutes) looking at AOIs on scene, significant differences were found for only two AOIs encountered during scene processing: the blood spatter pattern on the wall near the bed (average dwell time (in seconds) of 3.20 and 0.76 for experts and novices, respectively) and the second cartridge case underneath the comforter (average dwell time (in seconds) of 4.81 and 0.64 for experts and novices, respectively). In both comparisons, experienced analysts fixated significantly longer on
both AOIs (Figures 3.22 and 3.23), with p-values of 0.049 and 0.035, respectively, which was also seen in Busey et al. (2011) when no time limit was enforced.

Figure 3.22: Boxplot illustrating the pursuit percentages on the second cartridge case located near the bed hidden underneath the comforter. Note that experts had a significantly higher fixation median on the AOI than novices (p = 0.049). Outliers (o) are defined as any data point that falls outside of 1.5 times the interquartile range (IQR).

Figure 3.23: Boxplot illustrating the pursuit percentage on blood spatter pattern on wall near the bed. Note that experts had a significantly higher fixation median percentage on this AOI than novices (p = 0.035). Outliers (o) are defined as any data point that falls outside of 1.5 times the interquartile range (IQR).

Although there was a significant difference in pursuit percentage on the blood spatter pattern on the wall near the bed, the reconstruction data indicated that of the 17 experts
that fixated on this evidence, 11 commented on it, 7 of the 11 located the bullet in the wall, and a total of 10 gave an accurate reconstruction. Conversely, only 10 of the 15 novices fixated on this evidence, 5 commented on it, 1 of the 5 located the bullet (although 2 others located the bullet without fixation on the bloodstain), and 3 of the 15 gave an accurate reconstruction (which were the same 3 participants that located the bullet). Moreover, fixation on the second cartridge case, which was located near the bed beneath the comforter, also revealed a difference in detection between experts and trained novices. Approximately half of the participants did not find the cartridge casing; out of the 32 participants that took part in this study, 10 experts (58.5%) and three trained novices (20.0%) recovered the second casing and indicated that at least two shots were fired in the scene. Given that the difference in total dwell time per participant classification was not found to be significantly different via the Mann-Whitney U test, but experts had significantly higher reconstruction scores than their counterparts, there is evidence to indicate that experts perform more efficiently.

3.5 Earth Mover’s Distance

To further evaluate the dissimilarity of the examiners within the intuitive groups, Earth Mover’s Distance (EMD) was computed for all pairwise comparisons. Performing the Mann-Whitney U test showed that the group medians (6.10 and 4.42 for experts and novices, respectively) were significantly different from each other (p = 9.28e-06). Note that the mean EMD values were $17.4 \pm 16.1$ (95% confidence interval between 25.0 and 9.70) and $9.71 \pm 8.45$ (95% confidence interval between 14.0 and 5.43) for experts and trained novices, respectively. When groups were inter-compared, the EMD was found to be $13.0 \pm 13.2$ (with a 95% confidence interval that the true value lies between 6.28 and 19.7). However, this result could be due, in part, to the EMD inputs, namely the pursuit percentage on various AOIs. This study found that the majority of trained novices missed one or more items of interest (e.g., the second cartridge case near the bed). This phenomena could possibly decrease the overall “weights” (pursuit percentages on AOIs) that needed to be “transported”, resulting in a smaller EMD value overall.

In addition to using Euclidean distance to compute the similarity between each participant for the agglomerative hierarchical clustering method, EMD values were also used to assess possible natural clusters. Since the Ward’s method of clustering was the most successful grouping technique in this study, it was repeated with EMD to produce the dendrogram shown in Figure 3.24. Although there were obvious changes in the groupings formed (compared to Ward’s method using Euclidean distance), there were no notable trends found in terms of the members’ backgrounds. Moreover, members of each of the groupings also had varying reconstruction scores, suggesting that a similar searching and observing technique does not equate with a consistent score.
Figure 3.24: Dendrogram illustrating the groupings formed using Ward’s hierarchical clustering algorithm ($k = 4$).

### 3.6 Needleman-Wunsch (N-W) Algorithm

As previously discussed, the Needleman-Wunsch (N-W) pairwise comparison score was computed to allow for an additional similarity metric that incorporated search sequence, duration and fixation locations. The mean N-W pairwise comparison scores were $-0.55 \pm 0.44$ (95% confidence interval between -0.34 and -0.76) for experts and $-0.69 \pm 0.49$ (95% confidence interval between -0.44 and -0.94) for trained novices. Performing the Mann-Whitney U test showed that the group medians (experts = $-0.55$ and trained novices = $-0.68$) were significantly different from each other ($p = 0.02$), wherein experts were more consistent in their searching approach as compared to novices, which was contrary to the results found using EMD. Interestingly, one of the highest similarity scores computed in this study was between experts (police officers) and a graduate student. Moreover, this student had a substantial amount of practical experience (creating mock crime scenes, on top of actually performing crime scene investigations). This could mean that assisting and/or observing other students process a crime scene may increase performance, however, more than one data point is needed to further validate this interpretation.
4. Conclusion

4.1 Major Findings

This research represents a quantitative attempt to characterize the visual fixation duration, sequence and reconstruction ability of crime scene analysts with varying levels of educational and practical experience. The major conclusions are four fold. First, experts spent 92% of their time searching for evidence in this scene (26.9 minutes), 5% of their time looking at relevant evidence (1.44 minutes), and 8% documenting said evidence (2.34 minutes). Conversely, novices spent 90% of their time searching (21.0 minutes), an average of 3% looking at AOIs (0.70 minutes), and 10% documenting (2.33 minutes). Moreover, none of these comparisons were found to be significant in terms of either pursuit percentage or dwell time. However, a significant difference in reconstruction score was detected between groups, and although a direct causal link cannot be stated, the results suggest that expertise leads to efficiency and effectiveness.

Second, clustering results (both hierarchical and $k$-means) based on the coarse sampling of pursuit percentage on AOIs identifies natural clusters in the data; however, to date, the clusters do not appear to mirror a priori classification based on participant background. Therefore, latent variables not yet identified, may help explain the underlying group structures. Conversely, canonical discriminant analysis (CDA) on pursuit percentage of AOIs, when classified based on a priori group identity, and reconstruction scorings found reasonable categorization. Moreover, when CDA classification was subdivided using CSI-based formal education and professional training, natural clusters were likewise identified. Leave-one-out-cross-validation using linear discriminant analysis (LDA) indicated that classification success is maximized using CSI-based formal education (81.3% and 96.9% for AOIs and reconstruction scores, respectively) closely followed by professional training (53.1% and 79.1% for AOIs and reconstruction scores, respectively). Despite this observation, a causal relationship cannot be confirmed since it is possible that (i) latent variables not yet identified are actually driving the separation, and (ii) patent and/or latent variables may be interacting. For future study, the first possibility (latent variables) can be explored by using a more detailed and refined questionnaire to capture a better understanding of the variables that determines one’s expertise level. For instance, instead of just asking for the types of scenes that were processed by an individual in the past, it would be more informative to ask participants to divided this by scene type, since experi-
ence with a volume crime (such as a burglary) does not necessarily lend itself to increased experience in solving a homicide.

Interestingly, clustering was not apparent based on the binned number of cases processed (0-5, 6-10, 11-50, 51-100, 101-400 and 401+). Furthermore, when a linear relationship between cases processed versus reconstruction score was sought, the $R^2$ value was 0.08 (note: it is acknowledged that additional models (such as exponential or a polynomial fit) were not explored). Two possible explanations for this are as follows: (i) repetition without a foundation (formal education) or intervention (professional development) may not lead to increased performance, and (ii) repetition may lead to increased performance, but it must be scene-specific.

Third, when considering pursuit percentage on areas of interest in this scene, the Earth Movers Distance metric and the Mann-Whitney U test indicated that differences between groups could not be explained by random chance alone, and that smaller EMD values result when intra-comparing novices, meaning that less work is needed to relate the visual fixation pattern of novices. The corollary to this is that expert fixation patterns, although dissimilar to each other, can still arrive at higher reconstruction scores. This complements past laboratory results wherein experts, in time-unlimited experiments, had more variability in their gaze patterns compared to novices, while still exhibiting higher performance [16].

The last major conclusion of this study found that the search patterns (AOI, duration and fixation sequence) of experts, as assessed using the Needleman-Wunsch algorithm and the Mann-Whitney U test, were more similar to each other than those exhibited by novices. Again, although a causal relationship cannot be ascertained, there is some merit to suggest that this observation may reflect hypothesis-based evidence interpretation. However, since the search sequence being compared is scene-wide, greater evidence to support this is sought using shorter linkages with predicted causal relationships. For instance, since this study found that there were significant differences in the pursuit percentage on the blood spatter pattern on the wall, which could have played a role in finding the bullet in the wall, a few AOIs around this item of interest (e.g., blood spatter pattern on the floor, bedsheets and pillow) can be compared with the Needleman-Wunsch algorithm. This may cluster participants who were able to make connections between these AOIs and eventually find the bullet, possibly indicating a hypothesis-based sequential search technique, which is expected of experts.

### 4.2 Caveats

Participants in this study were sampled from a fixed geographical location, which limits the above conclusions to a single cohort, and therefore not necessarily representative of the entire population of crime scene examiners. Second, many of the trained novices are a product of a single academic curriculum, and therefore a single pedagogical approach to forensic education, again limiting the above conclusions. Third, a homicide/shooting
reconstruction scene was evaluated, and the results found here may not be transferable to all types of crimes.

4.3 Future Research

One of the major advantages of using this technology over the current methods is that it allows the test provider to monitor and easily identify evidence (commonly expected to be seen at a certain scene or any that appear in the visual field) that were not detected by the candidate. It is also beneficial to identify evidence that the participant fixated on, but did not seem to incorporate into their hypothesis or narrative for the crime. With this information, the post-scene video can be used for coaching/training purposes, wherein an expert guides a novice to attend to visual stimuli apparently overlooked, thereby allowing both the individual and the agency to determine key areas of improvement and individualized training. Implementing a requirement to wear an eye tracker while processing all scenes could also provide two major advances for the training and proficiency testing of investigators as compared to traditional methods. First, a supervisor can evaluate an individual at a later time thus eliminating the need to be physically present on the scene. In addition, if the tracker is worn on a regular basis, proficiency tests can be conducted in which analysts are unaware of being tested, thus eliminating a deviation from normal behavior.

Although this testing framework is currently difficult to implement in the field setting, it bears the potential for greater applicability in the academic setting, whereby eye tracking information can be used to assess students’ learned behaviors and this information can be used to structure crime scene investigation classes. It allows a single instructor to pay equal attention to all students and provide personal guidance as necessary, which is currently not possible when a large group of students are divided into teams to process several different mock scenes. By having the opportunity to be “present” (viewing the videos) to monitor every single student in the course, a clearer understanding of each student’s weaknesses can be assessed, and plans to rectify the mistakes made by a particular student can be devised. This can be done by having the student view their eye tracking video while the instructor addresses issues pertaining to actions taken or not taken by the student on scene. Moreover, eye tracking data from experts can be used to assist with training of students and professionals, where it has been shown that gaze training expedites learning [14], and improves the transfer and retention of expert-like skills [12] in other fields.

Overall, this research presented a quantitative method of proficiency testing in the forensic science field using eye tracking data to compare experts and trained novices. Such a study is novel and to our knowledge, has not been attempted with crime scene investigators. Preliminary results support the fact that eye tracking data (specifically duration on AOIs and search sequence) differ with expertise level. When this is data is combined with qualitative observations, such as reconstruction accuracy, an analyst’s proficiency may be quantifiable. However, to assert this, a blind study should be conducted wherein participant background is not known to the research analyst, and instead, only a
posteriori crime scene data is used to attempt to classify volunteers.
A. Glossary

Figure A.1: A fixation sequence in which fixations, saccades, AOIs, visits and revisits are shown.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area of interest (AOI)</td>
<td>An area of interest is a target region determined in a study (Figure A.1). In this study, the AOIs are relevant evidence found at the mock scene, as well as items shown to participants during post-processing. Defining AOIs allows easy extraction of metrics (e.g., fixation frequency, dwell time and pursuit percentage) for each region separately and for comparison between examiners.</td>
</tr>
<tr>
<td>Dwell</td>
<td>A single visit to an AOI from entry to exit.</td>
</tr>
<tr>
<td>Dwell time</td>
<td>Total time spent (usually in ms) in or on an AOI that includes all fixations and saccades during the AOI visit.</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Fixation</td>
<td>A point where the eye is looking and remains still over a certain period of time (Figure A.1). For this study, the minimum fixation time is 100ms.</td>
</tr>
<tr>
<td>Fixation frequency</td>
<td>The number of fixation occurrences on an AOI per investigation.</td>
</tr>
<tr>
<td>Fixation sequence</td>
<td>Order of fixation on AOIs and the dwell time on each AOI are used to demonstrate the fixation sequence of an individual. For this study, the term search sequence is more appropriate as a series of fixations will be used to display the gaze plot.</td>
</tr>
</tbody>
</table>
| Pursuit percentage   | Dwell time transformed into percentage. For this study, the dwell time is normalized by dividing it by the overall time spent observing the mock scene to obtain the percentage of time spent fixating on an AOI:  
\[
Pursuit\ \ Percentage = \frac{Dwell\ \ Time\ \ on\ \ AOI}{Overall\ \ Time} \times 100\%
\]   |
| Saccades             | Rapid movement from one fixation to another (Figure A.1). During saccades, we are “blind” in which any objects that appear in the field of view at this time are not “seen.”                                              |
| Time to first fixation| Time taken (usually in ms) from the start of recording until the first fixation within an AOI. For this study, the time to first fixation on an AOI will be measured from the very first fixation on the mock scene until first fixation within the AOI. |
| Visits               | Each time a subject fixates over an AOI (Figure A.1). If a subject fixates on the same AOI more than once, it will be called a revisit (Figure A.1).                                                             |
B. Survey: Eye Tracking as a Proficiency Test for CSI

*To be filled in by project staff member:*
Participant Code:

*To be filled in by participant:*

Demographic questions that will be used for classification purposes.

1. What is your gender?
   - □ Male
   - □ Female

2. Please provide your current professional title (*i.e.*, crime scene analyst, criminalist, undergraduate student, graduate student, criminalist I, II, III, IV, V, footwear examiner, etc.,) and the number of years you have held this position.

3. Have you held a position other than that listed above (but related to forensic science)? If so, please list the position(s) and the number of year(s) of service for each.

4. Please list any certifications you hold related to the forensic task(s) to be performed (*examples include the American Board of Criminalistics-Diplomate/Fellow, the International Association of Identification Certified Crime Scene Investigator/Analyst, etc.*)

5. Please itemize your educational history (most recent to oldest).

6. Have you taken any forensic science course prior to this?
   - □ Yes
   - □ No

   (a) If YES please list the course and credit hours.
7. Please describe any training you have received related to the forensic task(s) to be performed. This may include college/university courses, on-the-job training, workshops, successful completion of training/proficiency programs, etc.

8. Have you had any forensic-based internship
   □ Yes
   □ No
   (a) If YES please list and describe your major roles during the internship.

9. Have you processed an actual crime scene in the past?
   □ Yes
   □ No
   (a) If YES, please provide a rough estimate of the number of crime scenes you have processed.

10. Have you processed a mock or instructional crime scene in the past?
    □ Yes
    □ No
    (a) If YES, please provide a rough estimate of the number of mock scenes you have processed.

11. Have you created a mock or instructional crime scene in the past?
    □ Yes
    □ No
    (a) If YES, please provide a rough estimate of the number of instructional scenes you have created.

12. Please quantify (in days, months or years) the last scene (mock or otherwise) you processed/investigated.

13. Do you watch forensic or police television shows (fictional or otherwise, such as 48 Hours, Forensic Files, CSI, etc.)? If so, please list the shows and estimate the number of times a month you typically view this type of programming.
14. Which of the following best describes your personal rating of experience related to the forensic task(s) to be performed (check all that apply):

☐ Limited Practical Experience
☐ Limited Theoretical Experience
☐ Extensive Practical Experience
☐ Extensive Theoretical Experience
☐ Inexperienced
☐ Competent
☐ Above Average
☐ Expert
☐ Intermediate
☐ Supervisor
☐ Trainee
☐ Overqualified
C. Background Summary of Participants

Table C.1: Summary of trained novices’ background information. UG = Undergraduate Student, G1 = Graduate Student 1, P1 = Police Officer 1, HS = High School, AA = Associate of Art, BA = Bachelor of Art, BS = Bachelor of Science, MS = Master of Science, Y = Yes.

<table>
<thead>
<tr>
<th>Participant Code</th>
<th>0ggb9</th>
<th>1att9</th>
<th>3g14i</th>
<th>47vi</th>
<th>4x6r</th>
<th>5l8fr</th>
<th>77gb9</th>
<th>8w7b9</th>
<th>9hpvi</th>
<th>bzkt9</th>
<th>dobtt9</th>
<th>fgvi</th>
<th>gf1or</th>
<th>m5cdi</th>
<th>te29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Position</td>
<td>UG</td>
<td>P1</td>
<td>P1</td>
<td>G1</td>
<td>UG</td>
<td>P1</td>
<td>UG</td>
<td>P1</td>
<td>UG</td>
<td>P1</td>
<td>G1</td>
<td>P1</td>
<td>G1</td>
<td>UG</td>
<td>UG</td>
</tr>
<tr>
<td>Other Relevant Position</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certification(s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>HS</td>
<td>BA</td>
<td>HS</td>
<td>BA</td>
<td>HS</td>
<td>BA</td>
<td>HS</td>
<td>BA</td>
<td>MS</td>
<td>BA</td>
<td>BS</td>
<td>HS</td>
<td>HS</td>
<td>HS</td>
<td></td>
</tr>
<tr>
<td>Forensic Internship</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Actual scenes processed</td>
<td>20</td>
<td>5</td>
<td>15</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mock scene created</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSI-based education</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>CSI-based training</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>

Table C.2: Summary of experts’ background information. D = Detective, G2 = Graduate Student 2, P2 = Police Officer 2, HS = High School, BA = Bachelor of Art, BS = Bachelor of Science, MS = Master of Science, Y = Yes.

<table>
<thead>
<tr>
<th>Participant Code</th>
<th>1ra4i</th>
<th>1w0zfr</th>
<th>3q5mi</th>
<th>3z0k9</th>
<th>561or</th>
<th>6yldi</th>
<th>b8d7vi</th>
<th>bbj4i</th>
<th>dj9k9</th>
<th>dzpvi</th>
<th>huxr</th>
<th>o8yqfr</th>
<th>oflxr</th>
<th>os9k9</th>
<th>pui4d</th>
<th>sc3di</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Position</td>
<td>D</td>
<td>P2</td>
<td>P2</td>
<td>D</td>
<td>P2</td>
<td>D</td>
<td>G2</td>
<td>P2</td>
<td>D</td>
<td>G2</td>
<td>D</td>
<td>G2</td>
<td>D</td>
<td>G2</td>
<td>D</td>
<td>G2</td>
</tr>
<tr>
<td>Other Relevant Position</td>
<td>Y</td>
<td></td>
<td></td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Certification(s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>BA</td>
<td>BS</td>
<td>BS</td>
<td>BA</td>
<td>HS</td>
<td>BA</td>
<td>AA</td>
<td>MS</td>
<td>HS</td>
<td>HS</td>
<td>HS</td>
<td>BA</td>
<td>BS</td>
<td>HS</td>
<td>HS</td>
<td></td>
</tr>
<tr>
<td>Forensic Internship</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Actual scenes processed</td>
<td>100</td>
<td>400</td>
<td>400</td>
<td>200</td>
<td>40</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100+</td>
<td>100+</td>
<td>100+</td>
<td>100+</td>
<td></td>
</tr>
<tr>
<td>Mock scene created</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSI-based education</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>CSI-based training</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>
D. Post-processing Questionnaire for Participants

1. Was the homeowner truthful in his statement to the first responder?
   □ Yes
   □ No

2. Please indicate the location of the:
   (i) Homeowner during the shooting (draw a "O" to show his location and an arrow to show the direction the weapon was facing)
   (ii) Deceased during the shooting (draw an "X" to show his location)
3. Describe the position of the homeowner (*i.e.*, standing, kneeling, laying, etc.) while shooting.

4. Describe the position of the victim (*i.e.*, standing, kneeling, laying, etc.) while being shot.

5. Provide an overall reconstruction of what happened at this crime scene.

6. What evidence allowed you to develop and support this hypothesis?
E. Reconstruction Accuracy Scoring

Table E.1: Breakdown of reconstruction score per participant. In the checklist, position is an activity that the individual was doing (e.g., standing, sitting, kneeling, etc.), whereas location is the point in physical space.

<table>
<thead>
<tr>
<th>Participant Code</th>
<th>Homeowner truthful?</th>
<th>Position of homeowner</th>
<th>Position of deceased</th>
<th>Afflictive</th>
<th>Blood/chips on wife’s sweater</th>
<th>Shoe of deceased</th>
<th>Unipped joints</th>
<th>Staged</th>
<th>Bully was missed</th>
<th>No blood on knife</th>
<th>Blind spot pattern</th>
<th>2 shots fired</th>
<th>2 cartridge cases</th>
<th>Inlet in wall</th>
<th>Hemed on high</th>
<th>Gun ejects right</th>
<th>Magazine showed 2 spent cartridges</th>
<th>Wet was present</th>
<th>Location of deceased</th>
<th>Location of homeowner</th>
<th>Shooting direction</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Total score</td>
<td>36</td>
<td>16</td>
<td>14</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>55</td>
</tr>
</tbody>
</table>
## F. Pursuit Percentages

Table F.1: Computed pursuit percentage on AOIs and activities for all participants.

<table>
<thead>
<tr>
<th>Participant</th>
<th>cartridge 1</th>
<th>cartridge 2</th>
<th>floor</th>
<th>sheets</th>
<th>book</th>
<th>flowerpot</th>
<th>table</th>
<th>pillow</th>
<th>wall</th>
</tr>
</thead>
<tbody>
<tr>
<td>0gggb9</td>
<td>0.5</td>
<td>0.2</td>
<td>1.1</td>
<td>0.7</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1at9t</td>
<td>2.1</td>
<td>0</td>
<td>1</td>
<td>1.2</td>
<td>0.7</td>
<td>0.2</td>
<td>0</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>1ra4i</td>
<td>0.7</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>1w0zfr</td>
<td>0.4</td>
<td>0.3</td>
<td>0.5</td>
<td>0.7</td>
<td>0.4</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3g14i</td>
<td>0.6</td>
<td>0</td>
<td>1.8</td>
<td>0.8</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>3q5mi</td>
<td>0.5</td>
<td>0</td>
<td>0.4</td>
<td>0.7</td>
<td>0.5</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>3z0k9</td>
<td>1.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.2</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>47vi</td>
<td>0.7</td>
<td>0</td>
<td>1.5</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4x6r</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>561or</td>
<td>0.1</td>
<td>0</td>
<td>0.7</td>
<td>0.7</td>
<td>0.8</td>
<td>0.2</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5l8fr</td>
<td>0.1</td>
<td>0.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.6</td>
<td>0.2</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6yldl</td>
<td>0.4</td>
<td>0</td>
<td>3.0</td>
<td>1.9</td>
<td>1.2</td>
<td>0.6</td>
<td>0.5</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>77gb9</td>
<td>0.7</td>
<td>0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.5</td>
<td>0.1</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8w7b9</td>
<td>0.3</td>
<td>0</td>
<td>0.6</td>
<td>0.8</td>
<td>0.6</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9hpvi</td>
<td>0.1</td>
<td>0</td>
<td>0.6</td>
<td>0.7</td>
<td>0.4</td>
<td>0.1</td>
<td>0.2</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>b8d7vi</td>
<td>0.3</td>
<td>0.5</td>
<td>1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>bbj4i</td>
<td>0.5</td>
<td>0</td>
<td>3.0</td>
<td>0.2</td>
<td>0.7</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>1.1</td>
</tr>
<tr>
<td>bzkkt9</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.7</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dj9k9</td>
<td>0.7</td>
<td>0</td>
<td>1.3</td>
<td>1.3</td>
<td>1</td>
<td>0.6</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>dobnt9</td>
<td>0.2</td>
<td>0</td>
<td>0.6</td>
<td>0.3</td>
<td>0.5</td>
<td>0.4</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>dzpvi</td>
<td>0.9</td>
<td>0</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>fgvi</td>
<td>0.5</td>
<td>0</td>
<td>1.7</td>
<td>1.1</td>
<td>1.4</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>gflor</td>
<td>0.2</td>
<td>0</td>
<td>0.3</td>
<td>0.4</td>
<td>0.4</td>
<td>0</td>
<td>0.2</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>huxr</td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>m5cni</td>
<td>0.7</td>
<td>0.2</td>
<td>0.7</td>
<td>0.6</td>
<td>1.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>o8yqfr</td>
<td>1.4</td>
<td>0.7</td>
<td>2.7</td>
<td>1.6</td>
<td>1.2</td>
<td>1.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>offxr</td>
<td>0.5</td>
<td>0</td>
<td>3.0</td>
<td>0.3</td>
<td>0.5</td>
<td>0.4</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>os9k9</td>
<td>0.9</td>
<td>0.4</td>
<td>0.4</td>
<td>0.8</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>piudi</td>
<td>0.2</td>
<td>0</td>
<td>0.6</td>
<td>1.5</td>
<td>0.1</td>
<td>0.4</td>
<td>0</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>sc3di</td>
<td>0.8</td>
<td>0.4</td>
<td>0.5</td>
<td>0.2</td>
<td>1.7</td>
<td>0.4</td>
<td>0.1</td>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td>te29</td>
<td>0.1</td>
<td>0</td>
<td>0.5</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>x1or</td>
<td>0.1</td>
<td>0</td>
<td>1</td>
<td>0.7</td>
<td>0.3</td>
<td>0.3</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Participant</td>
<td>bullet</td>
<td>GSW leg enter</td>
<td>GSW leg exit</td>
<td>GSW chest</td>
<td>swipe</td>
<td>AOI</td>
<td>non-AOI</td>
<td>documenting</td>
<td>searching</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>---------------</td>
<td>--------------</td>
<td>-----------</td>
<td>-------</td>
<td>-----</td>
<td>---------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>0ggb9</td>
<td>0.6</td>
<td>0</td>
<td>0.2</td>
<td>0.2</td>
<td>3.9</td>
<td>87</td>
<td>9.1</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>1att9</td>
<td>0.4</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>6.2</td>
<td>76</td>
<td>17.8</td>
<td>82</td>
<td>82</td>
</tr>
<tr>
<td>1ra4i</td>
<td>0.4</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>3.7</td>
<td>95</td>
<td>1.4</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>1w0zfr</td>
<td>0.4</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>4.5</td>
<td>75</td>
<td>21</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>3g14i</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>3.5</td>
<td>90</td>
<td>6.6</td>
<td>93</td>
<td>93</td>
</tr>
<tr>
<td>3q5mi</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>2.7</td>
<td>96</td>
<td>1.0</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>3l0k9</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0.1</td>
<td>2.9</td>
<td>84</td>
<td>14</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td>47vi</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>3.5</td>
<td>90</td>
<td>6.6</td>
<td>93</td>
<td>93</td>
</tr>
<tr>
<td>4xfr</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>1.2</td>
<td>96</td>
<td>2.7</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>561or</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>2.9</td>
<td>85</td>
<td>12</td>
<td>88</td>
<td>88</td>
</tr>
<tr>
<td>5isfr</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>3</td>
<td>88</td>
<td>9.2</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>6yldi</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
<td>7.9</td>
<td>84</td>
<td>7.9</td>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>75gb9</td>
<td>0</td>
<td>0</td>
<td>0.4</td>
<td>0.1</td>
<td>3.5</td>
<td>83</td>
<td>14</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>8w7b9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>2.5</td>
<td>85</td>
<td>13</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td>9hpvi</td>
<td>0.1</td>
<td>1.4</td>
<td>0</td>
<td>0.1</td>
<td>3.9</td>
<td>87</td>
<td>8.9</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>8667vi</td>
<td>0.8</td>
<td>0.1</td>
<td>0</td>
<td>0.2</td>
<td>4.7</td>
<td>89</td>
<td>6.0</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>9bj4i</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>5.3</td>
<td>71</td>
<td>24.8</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>9bzk4i</td>
<td>0.2</td>
<td>0</td>
<td>0.3</td>
<td>0.2</td>
<td>2.1</td>
<td>97</td>
<td>0.8</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>9g9k9</td>
<td>0.6</td>
<td>0.4</td>
<td>0</td>
<td>0.2</td>
<td>6.7</td>
<td>89</td>
<td>4.2</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>9dob9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.4</td>
<td>2.8</td>
<td>85</td>
<td>12</td>
<td>88</td>
<td>88</td>
</tr>
<tr>
<td>9dz4pi</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>2.6</td>
<td>97</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9gvi</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.7</td>
<td>85</td>
<td>9.5</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>9g1or</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>2</td>
<td>98</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9huxr</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>3.0</td>
<td>88</td>
<td>9.3</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>9m3di</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>4.6</td>
<td>80</td>
<td>15</td>
<td>85</td>
<td>85</td>
</tr>
<tr>
<td>9s8yfr</td>
<td>0.5</td>
<td>0.4</td>
<td>0.6</td>
<td>0.1</td>
<td>12</td>
<td>79</td>
<td>9.3</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>9ofrr</td>
<td>1.7</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>6.2</td>
<td>88</td>
<td>6.2</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>9as9k9</td>
<td>0</td>
<td>0.6</td>
<td>0</td>
<td>0.2</td>
<td>4.4</td>
<td>93</td>
<td>2.5</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>9pi4di</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>0.2</td>
<td>3.9</td>
<td>96</td>
<td>0.1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9sc3di</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>4.8</td>
<td>86</td>
<td>9.6</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>9te9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>1.4</td>
<td>88</td>
<td>11</td>
<td>89</td>
<td>89</td>
</tr>
<tr>
<td>9x1or</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.1</td>
<td>96</td>
<td>0.1</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Bibliography


