The Differences in Gaze Behaviour Over Time and the Effects of Experience in Crime Scene Investigation

I.S.A. van den Heuvel
Student Number: 554801
554801ih@eur.nl
Master thesis Psychology
Erasmus School of Social and Behavioural Sciences
Erasmus University Rotterdam
Thesis advisor: Christine van Nooijen
Independent reviewer: Reshmi Marhe
8 April 2026
Word count: 8581

Abstract

Crime scene investigation is a critical component of the justice system, and errors in evidence collection can lead to miscarriages of justice. This exploratory study examined gaze behaviours of trainee forensic investigators during a 2-week burglary investigation training at the Dutch Forensic Investigation Academy. Four participants, varying in prior experience, investigated multiple mock crime scenes while wearing eye-tracking glasses and provided feedback through semi-structured interviews. A repeated-measures design compared gaze behaviours at the beginning (T1) and end (T2) of training, measuring fixation duration, saccade duration, and saccade amplitude, with prior experience as a potential moderator. Due to the very small sample size, the study was substantially underpowered, and results should be interpreted with caution. Paired-samples t-tests revealed no significant differences in gaze metrics between T1 and T2, although moderate-to-large effect sizes suggest potential trends that may not have been detectable. Mixed-design ANOVA indicated that experienced participants showed smaller changes in gaze behaviours over time, with a significant interaction for saccade duration. Qualitative findings highlighted that participants found reviewing their eye movements informative but emphasized the need for feedback from experienced trainers to interpret data effectively. Participants also reported that expectations about the crime type and scene familiarity influenced their visual strategies. These findings suggest that while training may influence gaze patterns, prior experience, task expectations, and scene familiarity likely shape observed behaviours as well. The study highlights the potential value of eye-tracking feedback in forensic training while emphasizing the need for cautious interpretation of results due to limited statistical power.

Crime has far-reaching societal consequences, affecting not only those directly involved but society as a whole. In the Netherlands, around 30,000 people were sentenced to prison in 2022, approximately 125 people were murdered in 2023, and roughly 22,000 residential burglaries were recorded in 2024, corresponding to an average of more than 60 burglaries per day (Centraal Bureau voor de Statistiek, 2024, 2025; De Rechtspraak, 2023). Since crime affects so many people, it is essential that the investigation is done thoroughly (Chang & Tsai, 2022). A large part of that investigation is forensic evidence; " …the contribution of forensic science is significantly greater than previously reported. DNA is utilised in greater than 95% of homicide enquiries. Fingerprints are utilised in 72% of murder enquiry is and footwear evidence in around 35% of all enquiries." (Green, 2019). Crime scene investigation is an essential part of this process (Chang & Tsai, 2022). A thorough and accurate investigation should lead to credible evidence and ultimately the conviction of the guilty or exoneration of the innocent.

Mistakes made during crime scene investigations can have severe consequences. Sadly, miscarriages of justice still happen in the Netherlands. Take for example the 14 cases that were added to the EU-database for miscarriages of justice in January 2024 (NU.nl, 2024). Research from the European Registry of Exonerations indicates that wrongfully convicted individuals in Europe spend an average of seven years incarcerated before being exonerated (Van Putten, 2024). Incomplete evidence collection at crime scenes can contribute to this (Edmond, 2002). Investigators still tend to make mistakes or use wide-ranging differences in strategies, even though investigating the crime scene is standard procedure when crimes have happened (Chowdhury, 2021; De Gruijter et al., 2016; Lidén & Almazrouei, 2023).

For example, investigators often vary in their evidence selection at crime scenes, even when decisions are made under identical circumstances (Lidén & Almazrouei, 2023). Under the exact same conditions, one participant collected a total of 15 samples of evidence while the other collected a total of 38 samples. The country the investigator operated in did not significantly affect their impressions. Furthermore, when investigators were faced with limited resources for an investigation, it appears that investigators changed their investigation strategy, instead of doing a limited amount of the investigating strategy they would apply with unlimited resources (Lidén & Almazrouei, 2023). This implies that under different resource constraints, investigators approach a crime differently, potentially affecting their findings.

Additionally, strategies used to collect evidence can change depending on the contextual information investigators receive. When investigators saw the same crime scene photos, but received contextual information that suggested a suicide, they collected samples from different locations than those who received information suggesting a homicide (Lidén & Almazrouei, 2023). This last point has also been found in other research where investigators indeed investigated and interpreted a crime scene differently depending on the information they received beforehand (Van Den Eeden et al., 2016). This is also defined as context bias, which is the tendency for a consideration to be influenced by background information.

Finally, crime scene investigators were less concerned with reconstructing the crime and testing multiple hypotheses, and more concerned with potential traces of the perpetrator, such as DNA, blood, or fingerprints (De Gruijter et al., 2016). To test multiple hypothesis, traces should be examined to determine whether they are from the suspect, the victim or another person. When investigators only consider traces significant if they are thought to be from the offender, they ignore the significance of other traces for the reconstruction of the incident. Relevant traces could be discovered with such a focus, but they might be dismissed as unimportant if they are not assumed to be those of the perpetrator (De Gruijter et al., 2016). 62% of investigators in a crime scene investigation study operating in several different countries, stated that the investigators' critical thinking and personal judgment should determine what constitutes relevant traces of evidence, as there are no guidelines or recommendations on how to gather evidence (Lidén & Almazrouei, 2023).

Research thus shows many differences in approaches of crime scene investigation, which could lead to human errors in the process of investigating. A potential cause of this could be heuristics or biases influencing the investigation. Psychological heuristics are a normal part of human cognition, and can be seen as shortcuts the mind makes to navigate the world more easily (Meterko & Cooper, 2022). However, when these shortcuts are applied during investigation it could influence the thoroughness of the search as well as the selected evidence. Research shows that professional investigators are equally susceptible for confirmation bias as lay people, with several different factors influencing this risk (Meterko & Cooper, 2022). Confirmation bias is defined as the tendency to interpret new evidence as confirmation of one's existing beliefs or theories. Applying this to evidence collection in an investigation could thus influence the outcome. Also, research shows that irrelevant non-medical information such as ethnicity of the victim influences the decisions investigators make (Dror et al., 2021). This most likely is due to base-rate expectations. This also applies when investigators have contact with people relevant to the case, and their beliefs around them (Dror et al., 2021). Other research also shows a susceptibility for confirmation bias, but also context biases, exemplar biases as well as consistency biases (Cooper & Meterko, 2019). Exemplar bias being when investigators are influenced by the presence of an exemplar, such as a fingerprint or handwriting comparison, and consistency bias being influences of previous investigation outcomes.

The research mentioned above indicates different strategies and influencing factors when investigating a crime scene between investigators. An explanation for this may be the expertise of the investigator. Research shows a significant difference between experts and novices within the forensic field. It is shown that novices in the field can often vary on their impression of a scene, while under the exact same circumstances as an expert (Lidén & Almazrouei, 2023). They were less likely to consider other scenarios than experts were, which indicates that novices are more likely to jump to conclusions. This leaves room for the question regarding the differences between experts and novices within their investigation methods, and how this expertise can be developed (Lidén & Almazrouei, 2023). Research on this shows that experts have more complex scan paths and longer average fixation duration than novices, indicating a more thorough look through (Durugbo, 2021). This is also seen in a preliminary evaluation, resulting in experts identifying almost all evidence at the scene, in contrast to novices identifying only about half of all evidence, impacting their knowledge of the crime (Durugbo, 2021). This trend is also seen when looking at strategy. Research shows that, while both experts and novices within the field look for a likely way of operating of the perpetrator, they do this with different methods (Baber & Butler, 2012). While both groups focused on the likely method of the offender, the experts focused more on objects with evidential value, while the novices focused more on individual objects. Novices examine the scene by looking at the items it contains, while experts think about the evidence analysis that can be done as a result of the examination. This means that, while the experts are working on the likely actions that can be taken as a result of the examination, the novices are working on describing the scene in terms of its features (Baber & Butler, 2012).

A type of crime that occurs daily and is often affected by differences in investigation is residential burglary. Residential burglaries often go unsolved (Goedhart et al., 2025). This is due to trouble collecting fingerprints, DNA and other evidence that belongs to the perpetrator instead of the resident. Therefore, it is important to know the target locations for evidence for investigating burglaries. Research shows that different burglary methods lead to distinct contact patterns (Goedhart et al., 2025). However, concrete information regarding target locations in different types of burglary is often unavailable to investigators. This leads to investigators having to rely primarily on their personal experience and expertise (Goedhart et al., 2025). This illustrates the importance of processing previously learned information and developing expertise in the field.

Cognitive processes such as memory and attention play a role in crime scene investigating and developing expertise. Approximately 80% of all information people register is gathered visually (Chang & Tsai, 2022). Investigators' memory can determine what evidence is used or what goes lost (Shirkey, 2010). This due to what an investigator remembers of what they have seen at the crime scene, how they remember it and if they then act upon these memories. Visual attention is influenced by many things, such as familiarity with the surroundings. Canham & Hegarty (2010) show that familiarity with the investigated type of crime influenced where investigators looked, and also how they interpreted the evidence they saw. This is also seen in a study by Chang & Tsai, (2022), who found that successful decision-makers in crime scenes often detected evidence more quickly and were more attentive to evidence in surrounding areas.

Gaze behaviours are indicative of visual attention and thus cognitive processes. Gaze behaviour can be divided in two components, namely fixations and saccades (Rayner, 2009). Fixations are when the eyes are focused on one location. This indicates the processing of information and thus registering what is seen to the memory. When participants look at certain Areas of Interest (AOIs), longer fixations are indicative of a deeper level of cognitive processing of the information that is taken in (Lee et al., 2019). Saccades are relatively opposite to fixations, as they are rapid movements of the eyes with little processing of the information that is seen (Rayner, 2009). A way to measure these gaze behaviours is through eye-tracking (Mele & Federici, 2012). This technology allows for real-time recording of eye movements and gives insight into the gaze behaviours and thus a valuable indication of underlying cognitive processes, which was not possible to observe before (Jacob & Karn, 2003; Mele & Federici, 2012). Eye-tracking technology can produce recordings that show how a person's eye gaze move as they complete a task (Emhardt et al., 2023).

Research shows that people are often unaware of their gaze behaviour, which can be problematic when learning complex visual tasks such as investigating (Kok et al., 2017). Showing people their previous fixations in these visual tasks only improves memory slightly at best, which means other learning possibilities must be applied (Kok et al., 2017).

Research has proven that feedback has positive effects on the implementation of search strategies, compared to regular teaching methods (Carroll et al., 2013). Research also shows that with a variation in research designs, usage in different domains, and variation in implementation of different forms, eye-tracking still leads to positive results on learning and performance (Emhardt et al., 2023).

This technology has been tested on the decision-making of investigators looking at crime scene photos (Chang & Tsai, 2022) and has revealed that investigators making the correct decision on what crime has occurred identify critical information faster, as well as paying more attention to information linking to the answer options, than investigators who made the wrong decision. Seeing these patterns may offer insight into what training is needed for investigating crime scenes and provide more teaching methods (Chang & Tsai, 2022).

When errors are made during the crime scene investigation, this can lead to a fundamental misunderstanding of the occurred crime, which can be the foundation for further errors in the subsequent investigation. Above, it is illustrated how certain errors within these investigations occur but are difficult to combat. However, research did indicate differences between experts and novices, with novices seemly making more errors. This would indicate that the improvement of training and learning within the forensic field could improve the investigations. Research also indicated eye tracking technology has had positive effects when used during training and learning environments, with also having the possibility of making people's behaviours at an investigation concretely visible for themselves and others.

The current study has been performed at the Forensic Investigation Academy (Forensische Opsporing Academie / FOA), which provides training regarding the investigations of crime scenes. During this training, experts with years of experience within the field will teach the trainees investigative skills for burglary crime scenes. Trainees will be given some prior information and then asked to investigate mock crime scenes. Given the previous research mentioned, indicating the differences seen between experts and novices, this way of teaching can be valuable as experts can transfer their knowledge and strategies to the new generation. To illustrate gaze behaviours, eye-tracking technology will be used. Afterward, self-report questionnaires will be completed, as well as semi-structured interviews with the trainees to gain insight in their experience and review their own behaviour.

Based on previous research, the following questions were examined in this study: Does the gaze behaviour of trainee crime investigators change over time when in training? Does prior experience (expertise) influence the difference in gaze behaviours over time? Can eye tracking technology enhance trainee crime scene investigators' learning experience?

These questions were examined through quantitative data collected eye-tracking technology, as well as qualitative data collected through semi-structured interviews.

Based on this, three hypotheses have been formulated:

1. Participants will show a difference in gaze behaviours between the beginning and the end of the course, indicating a transition from novice-like to expert-like gaze patterns.

2. Prior experience will negatively influence the difference in gaze behaviours over time.
3. Eye tracking technology will positively enhance participants' learning experience during the course.

   Hypotheses one and two will be analysed quantitively, where hypothesis three will be analysed through qualitative data.

If this study shows significant results and shows positive effects of eye-tracking technology within the learning experience, this could be a first step to implementing modern techniques such as eye-tracking more in practice. Eye-tracking feedback could make the concept of investigating a scene less abstract and provide more concrete tools during training of investigators, both for trainees and teachers. If this would improve the teaching of trainee investigators it could lead to more concise and thorough investigating with less influence of biases, and thus hopefully less human error in the investigation process. This would then in turn cause a better working of the justice system.

Method

In this study, two rounds of data collection were conducted during separate training courses offered by the Dutch Forensic Investigation Academy. The first round, conducted in February 2025, focused on a course about unnatural deaths and served as a pilot study to test materials and determine which types of data could be usefully collected. This pilot employed multiple data collection methods during different parts of the course, but the data were not included in the analysis.

The second round took place in May 2025 during a basic course on burglaries. Although the sample size was small, participants were followed over time, allowing for repeated measurements and additional analyses. Only the second round of data collection was included in the present study.

Pilot study

The pilot study occurred in February 2025 during the training "Investigating Unnatural Death" provided by the Dutch Forensic Investigation Academy. Participants were active police officers training to become forensic investigators. The pilot aimed to test procedures and materials during both theoretical and practical components of the training.

First, the study's purpose and procedures were explained, and informed consent was obtained. Four participants took part, and data collection occurred in pairs.

Second, participants received theoretical instruction on unnatural deaths and crime scene investigation. During this session, participants viewed images and videos of unnatural deaths while wearing eye-tracking glasses to record gaze behaviour. They were also asked to rate the images based on emotional impact. However, the collected data were unsuitable for analysis due to mismatched questions and images across participants, as well as low-quality visual recordings that did not clearly capture attention.

Third, during the practical portion, participants investigated a mock crime scene set up in vacation homes. The scenario depicted an alleged suicide with several pieces of evidence placed within the scene. Participants wore eye-tracking glasses while investigating and were later asked to review their recordings and complete a questionnaire on their strategies. Data collection was inconsistent due to variations in the information provided by trainers, the skills tested, and the investigation time allotted per duo. Consequently, these data were excluded from the analysis.

The pilot study informed improvements to procedures and interview questions for the main study. Only the second round of data collection, which followed these refinements, was analyzed.

Participants

Participants were recruited from the burglary investigation training offered by the Dutch Forensic Investigation Academy, which forms the final part of the 18-week Basic Police Training (BPT). The training was held on multiple contact days in May 2025. Participation was voluntary, and no reward was offered.

Due to the limited availability of the course and willingness to participate, only four participants consented to take part in the study. Therefore, this study should be considered exploratory.

Demographic information was not collected due to privacy considerations and the professional nature of the participants. All participants were active police officers in Police East Netherlands, meeting the eligibility criteria for the training and studying to become forensic investigators. Participants were excluded if they did not complete study requirements, did not provide informed consent, or could not wear eye-tracking glasses.

Materials

Learning materials

All training materials were prepared by the Forensic Investigation Academy. Prior to the training days, participants received self-study resources, including a textbook and legal articles on crime scene investigation. Before each mock crime scene, participants received standardized case information. Participants then investigated nine mock crime scenes, staged by trainers or participants themselves. Each scenario depicted a burglary, with varying evidence placement and visibility to simulate realistic investigative circumstances.

Apparatus

While participants were investigating the mock crime scenes, they were wearing the Pupil Labs Neon eye-tracking glasses, linked to the Neon Companion app on a phone the participants carried in their pocket (Pupil Labs, 2025). The visual data was uploaded to Pupil Cloud, a software that is able to download and display the recorded data that is collected (Pupil Labs, 2025). This data consists of, among other things, number of fixations, fixation durations (in milliseconds), saccade durations (in milliseconds) and saccade amplitudes (in amount of pixels). The visual data collected will then be analysed through Pupil Labs (Pupil Labs, 2025).

Interviews

A semi-structured interview was used to inquire on the experience of participants with the study as well as their experience within the field. This was used to assess how participants experienced the use of eye-tracking for their training, whether they believed it had a positive effect on their learning experience, and whether prior experience influenced their investigative strategies. Additionally, the semi-structured interviews were used to assess participants previous experiences in the field and level of expertise, as well as discuss their gaze behaviours. The semi-structured interviews were recorded using the voice recording app on the phone of one of the researchers to process the data later. Information from the interviews were used as direct citations as well as analysed thematically to be able to describe the overall findings. Due to the small sample size this thematic analysis was done by hand.

Procedure

Information about the study, privacy precautions taken, and voluntary participation has been provided to all participants on the first day, the introduction day of the training. Participants were asked to provide informed consent through a form brought by the researchers. Every participant received a participant number, to keep all the data collected from one participant together, without risking their privacy.

Before starting the actual data collection, the materials were demonstrated to the participants, to increase their understanding of the material as well as increase their feeling of comfort while using the glasses and being observed. Before use of the eye-tracking glasses an offset correction was done to collect the most accurate data. Gaze samples with less than 90% accuracy were not included. This was only seen once, during the pilot study, all other data was fit for analysis.

The training took place over several days. Several mock crime scenes were used at once by several groups of participants, spread over several vacation homes, cars and containers on the property of the Forensic Investigation Academy. Every participant wore the glasses during most mock crime scenes they investigated, which created a repeated measures design for the study. After participants finished their investigation, they received feedback from the trainer on their performance as they would normally, however trainers did not review their recorded gaze behaviour. Afterwards, all participants were sent the video recordings of their investigations with overlaid eye movements, so they were able to review the data themselves when they returned home.

After the training exam, participants were interviewed regarding their experience with the technology and its feedback, their level of expertise, their usage of strategies as well as follow up questions regarding their answers on the questionnaire.

Data preprocessing

All data was organised and made anonymous through labels. Participants received a participant number and mock crime scenes were dated and numbered. To prepare the data for analyses, two fragments were chosen per participant, one on the first day of training and one on the last day of training to measure the largest time difference possible. These fragments were selected from a point where the participant mentioned starting their orientation/investigation of the scene and ended when the participant mentioned having finished investigating or continuing to the next part of the process. All fragments selected took place in a similar setting for all participants and were of similar lengths for the first fragments as well as the last fragments. A fragment of all participants on day one as well as day six were analysed through fixation durations, saccade duration and saccade amplitude. Fixation duration is defined as the length of time the eyes remain stationary on a specific location, typically reflecting cognitive processing of visual information. Saccade duration is defined as the time taken for the eyes to move rapidly from one fixation point to another. Saccade amplitude is defined as the angular distance covered by the eyes during a saccadic movement, which in this study will be measured in number of pixels that was covered during the saccade. Since all analysis were done on the same screen, this is the most reliable method, as all saccades have taken place in different locations of the scene. All data on number of fixations, fixation durations, saccade durations and saccade amplitudes were downloaded of the fragments. This data was then all compiled into averages per participant per fragment. This was then used for the statistical analysis.

Participants experience with eye-tracking technology was also analysed. This was done in a qualitative manner through semi-structured interviews. This data is then analysed through thematic analysis.

Analyses

To analyse the eye-tracking data that was recorded, two different types of statistical analyses were used. For the first hypothesis, a paired samples T-test was performed. Within this analysis, time was the independent variable. The analysis was performed 3 separate times with fixation duration, saccade duration and saccade amplitude as dependent variables.

For the second hypothesis, qualitative data was used to analyse the amount of experience a participant had, to divide participants into two equal groups. Afterwards, a mixed-methods ANOVA was used with time as a within-subject factor, and experience as a between-subject factor. These were the independent variables within the analysis. Dependent variables were fixation duration, saccade duration and saccade amplitude. The effect of both time and experience was analysed, as well as the interaction between the two.

For the third hypothesis, the qualitative data from the semi-structured interviews were used. Both thematic analysis and direct quotes were used to illustrate the findings.

Results

Analysed data

The amount of data analysed per participant for T1 and T2 is shown in table 1. In the chosen fragments there is the same number of fixations as saccades, since after every fixation there is a saccade. Additionally, an alternative T2 has been chosen for one participant. This is done due to this participant reporting a familiarity with the crime scenes over time because the locations were often reused. However, on the second to last day, participants had to investigate a crime scene at a location they had never seen prior to that day. This participant reported that he had the feeling this had a great effect on him, and he felt he changed his gaze behaviours to be more thorough. To investigate whether this conclusion of the participant can be scientifically proven, the fragment of this alternative crime scene was also analysed to see the effect on the overall results. The results of this are discussed in "*Alternative location"*. In table 2, the means of the variables are given per participant.

A sensitivity power analysis was conducted using G*Power 3.1 (Faul et al., 2007) to estimate the sample size required to achieve adequate statistical power for the present study. Sensitivity analysis is appropriate in cases where no a priori power analysis was performed, as it allows for an evaluation of the sample size needed to detect effects of a given magnitude under conventional statistical criteria. Assuming a mixed-design repeated-measures ANOVA with one within-subject factor (time: T1 vs. T2) and one between-subject factor (expertise), an alpha level of .05, and desired power of .80, approximately 30–50 participants would be required to detect medium-sized effects ($f = 0.25$). Even for large effects ($f = 0.40$), a minimum of approximately 16–24 participants would be necessary. Given that the present study included only four participants, it was substantially underpowered, particularly for detecting between-subject and interaction effects, and the results should therefore be interpreted with severe caution.

Table 1

Analysed data per participant per fragment

	T1		T2	
Participant	Duration	Fixations / Saccades	Duration	Fixations / Saccades
Participant 1	28m 29s	4474	11m 55s	1911
Participant 1 Alternative	28m 29s	4474	9m 20s	1522
Participant 2	27m 32s	3282	19m 29s	3282
Participant 3	27m 20s	1766	10m 19s	1766
Participant 4	27m 23s	1529	8m 53s	1529

Difference in gaze behaviours

In order to test the first hypothesis on difference in gaze behaviour, paired samples t-tests were performed. The assumption of normality was checked through the Shapiro-Wilk test and Q-Q plots, which did not indicate substantial deviations from normality. The Shapiro-Wilk test showed for fixation duration p = .990, for saccade duration p = .112 and for saccade amplitude p = .497. There were no outliers detected in the data, as boxplots of the data showed no extreme data points. The descriptive statistics of the fixation duration, saccade duration and saccade amplitude are shown in Table 2.

No statistically significant differences were observed between T1 and T2 for any of the gaze measures. For mean fixation duration, the effect did not reach significance, $t(3) = 2.50$, $p = .088$, although the effect size was large (Cohen's $d = 1.25$), with fixation duration decreasing over time. Similarly, saccade duration did not show a significant difference, $t(3) = -1.99$, $p = .140$, but yielded a large effect size (Cohen's $d = -1.00$), indicating an increase over time. For saccade amplitude, no significant effect was found, $t(3) = 1.37$, $p = .263$, with a moderate-to-large effect size (Cohen's $d = 0.69$), suggesting a decrease over time.

Given the very small sample size, the study was substantially underpowered to detect statistically significant effects, and these results should therefore be interpreted with caution. The pattern seen within the effect sizes is illustrated in Figure 1. While the observed effect sizes suggest potentially meaningful changes in gaze behaviour over time, these patterns should be considered explorative and require replication in a larger sample to determine their reliability.

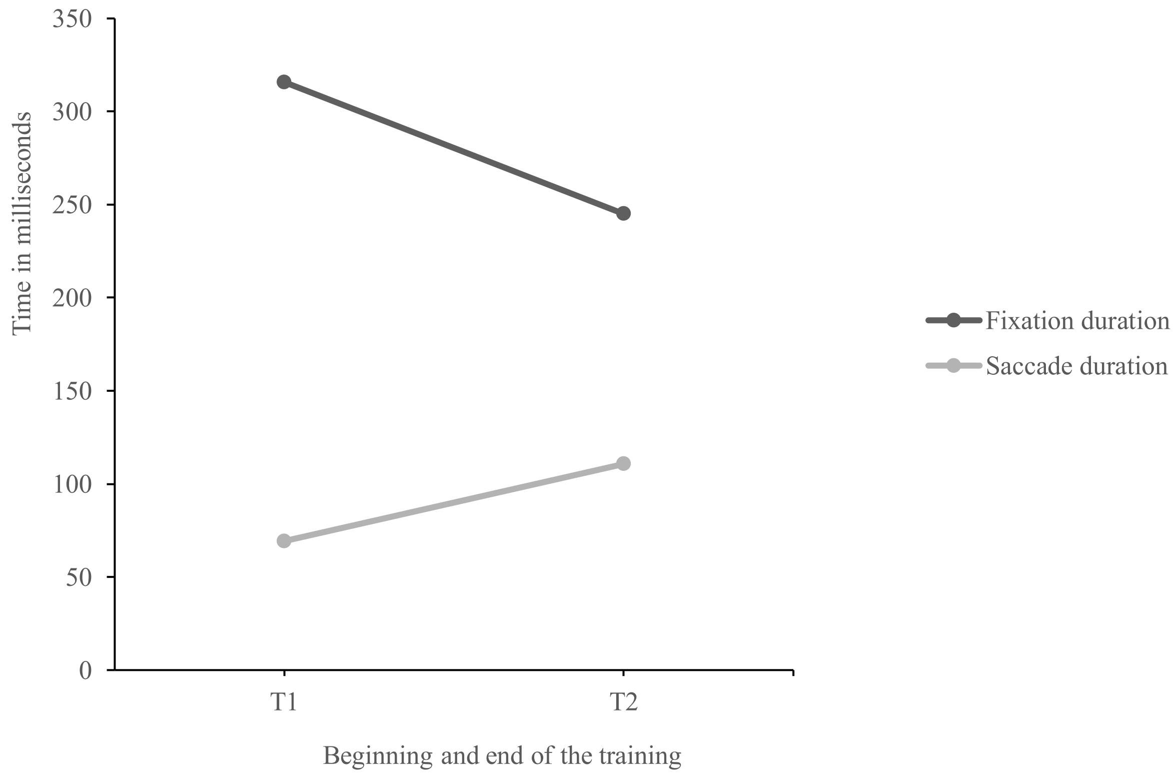
Table 2

Descriptive statistics on the first and last day of the training

Gaze behaviours	T1 (mean (SD))	T2 (mean (SD))
Fixation duration (ms)	315.77 (37.46)	245.02 (47.92)
Saccade duration (ms)	69.30 (6.62)	110.69 (43.50)
Saccade amplitude (px)	13173958592859950.00 (117316805518862.02)	12777180422717850.00 (483013308000824.30)

Figure 1

Descriptive statistics on the first and last day of the training



Alternative location

One participant reported doubts regarding the reliability of their data due to gaining familiarity over time with both the type of evidence as well as the locations of the mock crime scenes. However, they did report they were unfamiliar with the mock crime scene on the second to last day of the training where they believed to have behaved differently when investigating this scene. It was believed by the participant this could have been seen in the measured gaze behaviours.

To investigate these claims, an alternative T2 fragment was chosen from the mock crime scene at a different location on the second to last day of the training for this specific participant. When looking at the difference between the alternative T2 measurements, it is seen that there is a small difference between the two. Where the participants fixation duration decreased with 0.59% between T1 and the original T2, it decreased 0.68% between T1 and the alternative T2. This is in line with the claims of the participant, however the difference is too small to make any reliable claims regarding the origin of this difference. Also, the small amount of data does not allow for statistical testing of the results.

Furthermore, it should be mentioned that this was the participant with the most experience and the least difference in gaze behaviours over time compared to the other participants in this study. The data of the other participants showed larger trends, with the percentual change over time being 17.43%, 30.58% and 38.75%. Therefore, these results should be interpreted with caution as these results could vary greatly when considering a participant with less experience or different gaze behaviours on crime scenes, regardless of familiarity.

The effect of expertise on gaze behaviours

Participants showed a difference in their experience, with two participants reporting they only saw one real crime scene prior to the training, with the other two reporting to have seen multiple crime scenes with experienced colleagues as well as having done investigations independently. Participant data was thus separated into two groups, these being "experienced" and "novice". Participants who had experience with three or more real crime scenes were categorised as experienced. This resulted in two participants being categorised as "experienced" and two participants as "novice", of which one participant had only done internships and the other had experienced one real crime scene. To illustrate the differences in performance, participants' means on the variables and their expertise categorisation were given in Table 3.

Table 3

Means (SD) and expertise categorisation per participant

	Fixation duration (ms)		*Saccade duration (ms)*		*Expertise*
Participant	*T1*	*T2*	*T1*	*T2*	
Participant 1	293.76 (231.20)	292.03 (262.14)	74.80 (69.13)	82.20 (80.56)	Experienced
Participant 2	356.67 (537.17)	218.47 (203.96)	65.13 (60.03)	137.40 (178.37)	Novice
Participant 3	336.92 (400.99)	278.19 (193.52)	62.21 (51.47)	66.04 (50.91)	Experienced
Participant 4	275.71 (263.88)	191.41 (154.53)	75.08 (82.33)	157.13 (193.59)	Novice

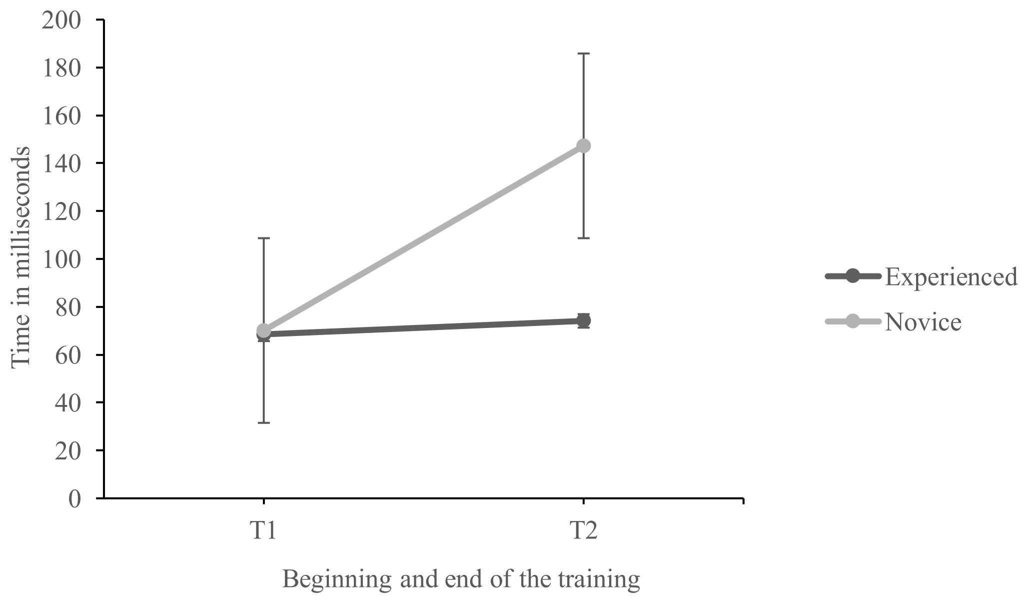
To examine differences in gaze behaviour as a function of time and expertise, a mixed-design ANOVA was conducted with time (T1, T2) as a within-subject factor and experience (novice, experienced) as a between-subject factor. Given the extremely small sample size, this analysis should be interpreted as exploratory, particularly for between-subject and interaction effects.

For mean fixation duration, there was no significant main effect of experience, $F(1, 2) = 2.01$, $p = .292$, $\eta^2 = .50$, and the interaction between time and experience was also not statistically significant, $F(1, 2) = 4.27$, $p = .175$, $\eta^2 = .68$. Similarly, for mean saccade duration, no statistically significant main effect of experience was observed, $F(1, 2) = 13.08$, $p = .069$, $\eta^2 = .87$. The interaction between time and experience reached statistical significance, $F(1, 2) = 189.15$, $p = .005$, $\eta^2 = .99$. This effect is illustrated in Figure 2. For mean saccade amplitude, there was no significant main effect of experience, $F(1, 2) = 3.71$, $p = .194$, $\eta^2 = .65$, and no significant interaction between time and experience, $F(1, 2) = 1.14$, $p = .398$, $\eta^2 = .36$.

Overall, although only the interaction effect for saccade duration reached statistical significance, all observed patterns were broadly consistent in suggesting that experienced participants showed smaller changes in gaze behaviour over time compared to novices. However, given the limited sample size, these findings should be interpreted with caution, and the observed patterns should be considered explorative until replicated in a larger sample.

Figure 2

Mean saccade time differences in milliseconds



Note. Experienced = 2, novice = 2, total $N = 4$

Feedback on the use of eye-tracking during training

To examine whether eye tracking technology enhanced participants' learning experience, qualitative data from the semi-structured interviews were analysed using thematic analysis. Several themes were identified regarding how participants experienced the use of eye-tracking feedback during the training.

A main theme in the data was that participants found the eye-tracking data interesting, but difficult to use on their own to improve their learning. Participants mentioned that although the data showed their gaze behaviour, they did not know how to interpret it or what to change:

"I looked at the data and found it really interesting to see the many movements my eyes made during investigating. However, whether that behaviour was right or not, I have no idea. It was cool to see but I had no indication on if I should change anything and how I would do that."

Participants also reported that the data would be more useful if it was combined with feedback from an experienced teacher or colleague. All participants indicated that guidance would help them understand how to improve their performance:

"If my teacher would look at the recording and suggest a different strategy, that would help for sure! Now, it's more guesswork for them on how we performed our investigation, so I think it would make it easier for them too."

In addition, participants mentioned that comparing their gaze behaviour to that of experienced investigators could help their learning. This suggests that eye tracking may be more useful when it is used together with feedback, rather than on its own.

Some participants also described using the eye-tracking recordings to reflect on their performance, especially after receiving feedback from trainers. In these cases, participants went back to the recordings to understand how they missed certain pieces of evidence:

"I was told the phone was under the bed, but I was sure I looked there. Later I viewed the recordings to see if I had seen the phone but simply not registered it or if I had truly completely missed that area."

These findings suggest that eye tracking can support reflection, particularly when combined with feedback.

Overall, the results show that eye tracking technology may enhance participants' learning experience, but mainly when guidance is provided. When participants had to interpret the data on their own, they found it difficult to use it to improve.

Contextual factors influencing the learning experience

In addition to the main themes, participants mentioned several factors that may have influenced their learning experience. These findings are important to consider when interpreting the results.

Firstly, participants reported that they approached the mock crime scenes differently compared to real-life situations. Because they had seen the training environments multiple times and knew what kind of evidence was expected, they developed expectations about where to look:

"Yes, I have seen those houses so often that week, you simply know what they want to know: the way in and out of the perpetrator. They also want to test shoeprints and blood samples. Thus, I figured if the door was not reported broken then the window would most likely have either blood or fingerprints on it, as that is an easy spot to plant such things and is usually what they choose to do."

This may have influenced their gaze behaviour and how realistic the eye-tracking data is.

Participants also reported that gaze behaviour would likely differ depending on the type of crime scene. For example, burglary scenes require a broader search, while other types of crimes may involve focusing on specific areas. This suggests that the usefulness of eye tracking may depend on the context.

Lastly, participants mentioned that they were working towards a final exam, which influenced their behaviour. They reported focusing on showing the expected steps rather than naturally exploring the scene, which was described as "performing the trick they want to see." In addition, once they felt confident in certain skills, they shifted their focus to other aspects of the task.

Together, these factors may have influenced both gaze behaviour and how useful participants found the eye-tracking feedback. These points should therefore be considered when interpreting the results of this study.

Discussion

In this study, I investigated whether participants demonstrated significant differences in gaze behaviours between the beginning (T1) and the end (T2) of their training in investigating burglary crime scenes. I also examined whether these differences were moderated by participants' prior knowledge and experience in the field. Furthermore, I explored whether eye-tracking feedback was perceived as useful for supporting learning. Overall, no significant differences were observed between T1 and T2 in terms of fixation duration, saccade duration, or saccade amplitude. However, a significant interaction was found between prior experience and changes in saccade duration, suggesting that experience may influence how gaze behaviour changes over the course of training. Participants also reported finding the eye-tracking feedback interesting, but many reported that the data alone was insufficient for learning without additional guidance from instructors or observation of expert behaviour. These findings should be interpreted with caution due to the limited sample size. This study was substantially underpowered, meaning that subtle but potentially meaningful changes in gaze behaviour may have gone undetected.

Regarding the first hypothesis, which predicted differences in gaze behaviours over time, the analyses showed no statistically significant changes in fixation or saccade durations. Importantly, a non-significant result does not imply that learning- or experience-related changes did not occur. The lack of significance may reflect insufficient statistical power, masking subtle trends that could be meaningful in a larger sample. Moreover, individual differences in investigative strategies, task approaches, and attention allocation may have introduced variability that further obscured patterns. Therefore, caution is required when interpreting these findings as indicative of no change.

Although the changes over time were not statistically significant, effect sizes for fixation and saccade metrics ranged from moderate to large. This suggests that the training may have had some influence on participants' gaze behaviours, even if the current study lacked the sample size to detect it definitively. Interestingly, a general decrease in fixation duration was observed across participants over time, which contrasts with prior research showing that experts typically demonstrate longer fixations than novices, alongside shorter saccades (Durugbo, 2021; Li et al., 2023; Prytz et al., 2018). Such patterns have been observed across multiple domains, including surgical training and emergency response, as well as in forensic contexts, where increased fixation duration is associated with expertise (Durugbo, 2021). This discrepancy highlights the need to interpret trends cautiously and consider context-specific factors that may influence gaze behaviours.

Within the forensic domain, it is particularly important to consider task-specific and situational factors. Previous research on evidence recognition using eye-tracking has indicated that while experts may show more similarity in search sequences, they exhibit greater variation in search duration, reflecting flexibility in adapting to different scenes (Watalingam et al., 2017). Experts also tend to demonstrate more complex scan paths compared to novices (Durugbo, 2021). In the current study, participants investigated the same locations for mock crime scenes at T1 and T2, and fragments analysed at T2 were shorter than at T1, suggesting some change in search behaviour. However, the study did not examine search sequences or scan paths, which could further explain observed differences. Expectations of what evidence might be present and prior information also influence search strategies, indirectly shaping gaze behaviours (Lidén & Almazrouei, 2023; Van Den Eeden et al., 2016). Participants were trained specifically on burglary investigations, which likely shaped their expectations and gaze patterns. Additionally, the mock crime scenes became familiar over time, potentially leading participants to scan for anomalies rather than investigating thoroughly. This combination of prior expectations and scene familiarity may explain why trends observed here diverge from previous findings.

When considering the interaction between experience and changes in gaze behaviours, a pattern showed. More experienced participants demonstrated relatively stable gaze behaviours over time, whereas novices exhibited increased saccade durations. This partially contradicts the typical expert-novice patterns observed in other fields, where experts generally show longer fixations and shorter saccades (Durugbo, 2021; Li et al., 2023; Prytz et al., 2018). One explanation may relate to confidence in investigative skills. Experienced participants reported feeling familiar with the evidence and confident in their ability to identify it, leading them to search less thoroughly over time. In contrast, novices often conducted a final walkthrough of the scene to ensure they had not missed evidence. These walkthroughs may have increased saccade durations while decreasing fixation times, as novices scanned larger areas rather than concentrating on individual pieces of evidence. This behavioural pattern aligns with qualitative reports indicating that novices spent more time practicing, whereas experienced participants approached the tasks with confidence that additional practice was unlikely to change outcomes. These qualitative insights highlight how procedural behaviours and confidence levels interact with gaze metrics, illustrating the importance of integrating qualitative data when interpreting eye-tracking results.

The perceived usefulness of eye-tracking feedback was analysed through qualitative data. Thematic analysis revealed that participants found the gaze data interesting but insufficient alone for improving their skills. Participants suggested that pairing eye-tracking data with instructor feedback or expert examples would enhance learning. This aligns with previous research demonstrating that feedback can improve learning outcomes in eye-tracking tasks (Carroll et al., 2013). Some reported using the data to identify evidence they had overlooked, while others reflected on differences between seeing and processing visual information. Participants' suggestions point to practical implications for training design: eye-tracking feedback alone may be insufficient to support learning, however, it can be a valuable tool when integrated with structured feedback, modeling, and reflective exercises.

Participants also noted several contextual factors that likely influenced their gaze behaviours, including the type of crime scene, preparation for an exam versus a real investigation, knowledge of the crime type, and familiarity with the mock scenes. Such factors suggest that gaze behaviours in controlled training environments may differ substantially from those in real investigations. While trends in the current study did not fully align with previous research, the qualitative feedback indicates that context, task goals, and participant expectations are critical when interpreting gaze data. This has implications for future research: future research should consider the influence of context, task demands, and participant goals when assessing eye-tracking data in investigative training.

This study also has practical and social implications. Participants' behaviours during training differed from those expected in real-life investigations. For example, participants often focused on passing the exam rather than approaching the scene as a real investigation. While exam-oriented training can teach important procedural skills, it may not fully prepare participants for the variability and complexity of real crime scenes. While the training imparted valuable skills relevant to the exam and certain investigative tasks, it raises questions about whether broader investigative abilities are being enhanced. Eye-tracking data could potentially be leveraged to provide more individualized feedback and to measure performance relative to real-life scenarios. Furthermore, insights about novice behaviours could inform training curricula to inhance targeted feedback to develop more efficient search strategies and confidence in their investigative decisions.

Several limitations must be considered. The small sample size limited statistical power and the ability to account for individual differences in strategy, attention, and experience. Areas of Interest (AOIs) could not be analyzed due to differences across mock crime scenes, restricting analysis to gaze durations rather than fixation content. Additionally, participants were trained on specific evidence types and locations typical of burglary scenes, which may not generalize to other crime scene types or real-life investigations. These constraints show the need for cautious interpretation of the results and the importance of complementary qualitative insights to provide context to the eye-tracking data.

In conclusion, while this study did not find statistically significant changes in gaze behaviours across the training, the observed effect sizes, interaction patterns, and participants' qualitative feedback suggest meaningful trends that warrant further investigation. The findings highlight that training context, prior experience, and feedback when interpreting eye-tracking data are critical factors influencing investigative behaviour and the learning experience. Future work should aim to combine quantitative and qualitative measures to better capture the complexity of learning and performance in real crime investigations, ensuring that eye-tracking feedback is used effectively to support investigative skill development.

References

Baber, C., & Butler, M. (2012). Expertise in Crime Scene Examination: Comparing Search Strategies of Expert and Novice Crime Scene Examiners in Simulated Crime Scenes. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(3), 413–424. https://doi.org/10.1177/0018720812440577

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*(2), 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Carroll, M., Kokini, C., & Moss, J. (2013). Training effectiveness of eye tracking-based feedback at improving visual search skills. *International Journal of Learning Technology*, *8*(2), 147. https://doi.org/10.1504/IJLT.2013.055671

Centraal Bureau voor de Statistiek. (2024, August 29). *125 mensen vermoord in 2023*. Centraal Bureau Voor de Statistiek. https://www.cbs.nl/nl-nl/nieuws/2024/35/125-mensen-vermoord-in-2023

Centraal Bureau voor de Statistiek. (2025). *Fewer crimes recorded in 2024*. https://www.cbs.nl/en-gb/news/2025/10/fewer-crimes-recorded-in-2024?utm_source=chatgpt.com

Chang, R., & Tsai, M. (2022). Visual behavior patterns of successful decision makers in crime scene photo investigation: An eye tracking analysis. *Journal of Forensic Sciences*, *67*(3), 1072–1083. https://doi.org/10.1111/1556-4029.14970

Chowdhury, M. (2021). A broken system? Examining the perilous state of quality assurance in crime scene practice. *Science & Justice*, *61*(5), 564–572. https://doi.org/10.1016/j.scijus.2021.07.001

Cooper, G. S., & Meterko, V. (2019). Cognitive bias research in forensic science: A systematic review. *Forensic Science International*, *297*, 35–46. https://doi.org/10.1016/j.forsciint.2019.01.016

De Gruijter, M., De Poot, C. J., & Elffers, H. (2016). The Influence of New Technologies on the Visual Attention of CSIs Performing a Crime Scene Investigation. *Journal of Forensic Sciences*, *61*(1), 43–51. https://doi.org/10.1111/1556-4029.12904

De Rechtspraak. (2023, October 11). *Aantal strafzaken bleef in 2022 op hetzelfde niveau*. Rechtspraak.Nl. https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Nieuws/Paginas/Aantal-strafzaken-bleef-in-2022-op-hetzelfde-niveau.aspx#:~:text=In%2090%20procent%20van%20de,en%20geldboete%20(14%20duizend).

Dror, I., Melinek, J., Arden, J. L., Kukucka, J., Hawkins, S., Carter, J., & Atherton, D. S. (2021). Cognitive bias in forensic pathology decisions. *Journal of Forensic Sciences*, *66*(5), 1751–1757. https://doi.org/10.1111/1556-4029.14697

Durugbo, C. M. (2021). Eye tracking for work-related visual search: A cognitive task analysis. *Ergonomics*, *64*(2), 225–240. https://doi.org/10.1080/00140139.2020.1822547

Edmond, G. (2002). Constructing Miscarriages of Justice: Misunderstanding Scientific Evidence in High Profile Criminal Appeals. *Oxford Journal of Legal Studies*, *22*(1), 53–89. https://doi.org/10.1093/ojls/22.1.53

Emhardt, S. N., Kok, E., Van Gog, T., Brandt-Gruwel, S., Van Marlen, T., & Jarodzka, H. (2023). Visualizing a Task Performer's Gaze to Foster Observers' Performance and Learning—A Systematic Literature Review on Eye Movement Modeling Examples. *Educational Psychology Review*, *35*(1), 23. https://doi.org/10.1007/s10648-023-09731-7

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Goedhart, Y. R., Draxel, K., Katsikis, A. S. I., Jullens, I., De Ronde, A., Kokshoorn, B., & De Poot, C. J. (2025). Determination of relevant sampling locations for burglary investigations. *Forensic Science International*, *377*, 112615. https://doi.org/10.1016/j.forsciint.2025.112615

Jacob, R. J. K., & Karn, K. S. (2003). Eye Tracking in Human-Computer Interaction and Usability Research. In *The Mind's Eye* (pp. 573–605). Elsevier. https://doi.org/10.1016/B978-044451020-4/50031-1

Kok, E. M., Aizenman, A. M., Võ, M. L.-H., & Wolfe, J. M. (2017). Even if I showed you where you looked, remembering where you just looked is hard. *Journal of Vision*, *17*(12), 2. https://doi.org/10.1167/17.12.2

Lee, J. Y., Donkers, J., Jarodzka, H., & Van Merriënboer, J. J. G. (2019). How prior knowledge affects problem-solving performance in a medical simulation game: Using game-logs and eye-tracking. *Computers in Human Behavior*, *99*, 268–277. https://doi.org/10.1016/j.chb.2019.05.035

Li, S., Duffy, M. C., Lajoie, S. P., Zheng, J., & Lachapelle, K. (2023). Using eye tracking to examine expert-novice differences during simulated surgical training: A case study. *Computers in Human Behavior*, *144*, 107720. https://doi.org/10.1016/j.chb.2023.107720

Lidén, M., & Almazrouei, M. A. (2023). "Blood, Bucks and Bias": Reliability and biasability of crime scene investigators' selection and prioritization of blood traces. *Science & Justice*, *63*(2), 276–293. https://doi.org/10.1016/j.scijus.2023.01.005

Mele, M. L., & Federici, S. (2012). A psychotechnological review on eye-tracking systems: Towards user experience. *Disability and Rehabilitation: Assistive Technology*, *7*(4), 261–281. https://doi.org/10.3109/17483107.2011.635326

Meterko, V., & Cooper, G. (2022). Cognitive Biases in Criminal Case Evaluation: A Review of the Research. *Journal of Police and Criminal Psychology*, *37*(1), 101–122. https://doi.org/10.1007/s11896-020-09425-8

NU.nl. (2024, January 30). *14 Nederlandse zaken toegevoegd aan EU-database met rechterlijke dwalingen*. NU.Nl. https://www.nu.nl/buitenland/6299680/14-nederlandse-zaken-toegevoegd-aan-eu-database-met-rechterlijke-dwalingen.html?referrer=https%3A%2F%2Fwww.google.com%2F

Prytz, E. G., Norén, C., & Jonson, C.-O. (2018). Fixation Differences in Visual Search of Accident Scenes by Novices and Expert Emergency Responders. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *60*(8), 1219–1227. https://doi.org/10.1177/0018720818788142

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Shirkey, H. B. (2010). *Last attorney to the jury box is a rotten egg: Overcoming psychological hurdles in the order of presentation at trial*. *8*, 581–601.

Van Den Eeden, C. A. J., De Poot, C. J., & Van Koppen, P. J. (2016). Forensic expectations: Investigating a crime scene with prior information. *Science & Justice*, *56*(6), 475–481. https://doi.org/10.1016/j.scijus.2016.08.003

Van Putten, L. (2024, January). *Learning from miscarriages of justice with the new European Registry of Exonerations*. https://www.universiteitleiden.nl/en/news/2024/01/learning-from-miscarriages-of-justice-with-the-new-european-registry-of-exonerations

Watalingam, R. D., Richetelli, N., Pelz, J. B., & Speir, J. A. (2017). Eye tracking to evaluate evidence recognition in crime scene investigations. *Forensic Science International*, *280*, 64–80. https://doi.org/10.1016/j.forsciint.2017.08.012