

All Eyes on the Evidence: Classifying Expert and Novice Crime Scene Investigators on the basis of Eye Movement Patterns

Master Thesis - Data Science: Business and Governance  
The School of Humanities and digital Sciences

**Author:** Décio J. Da Silveira Quiosa  
**Studentnumber:** 2017433

**ANR:** u283203  
**Supervisor:** Frouke Hermens

**Second reader:**

**Theme:** Eye tracking in Crime scene investigation

December 2018

Tilburg University, Tilburg

# Preface

# Abstract

# 1 Introduction

## Context

Observation of the scenery around us does not happen in one glance, rather we observe our surroundings one small area at the time fixating our fovea onto different objects so we can allow our visual oculomotor system absorb the environment in a fine grained manner.

## Research questions

This research project aims to determine whether eye movement patterns can be used to distinguish between expert crime scene investigators and novices in a classification task. In addition, this project aims to examine the role of context (live crime scene investigation or based on images of the crime scene).

* ~~Which eye movement measurements contain the most information on expertise?~~
* ~~On the basis of which eye movements can we classify expertise?~~
* ~~Are these eye measurements consistent throughout different content and experimental setups?~~
* ~~Evaluate if simulation is a good proxy for live examination~~

## Findings

[**http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.67.6184&rep=rep1&type=pdf**](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.67.6184&rep=rep1&type=pdf)

# 2 Related work

Gaze behavior is an important component of attention and perception and has a substantial impact on information processing of an individual (Rizolatti, Riggio, Dascola & Umilta, 1987). The mechanisms that underlie our gaze behavior are generally classified into two categories. Whilst top-down mechanisms direct our visual attention to the task at hand, bottom up mechanisms ensure that our visual attention is directed to particular points of interest (e.g. noteworthy visual attributes of photographs) (Kollmorgen et al., 2010).The mechanism driven by stimuli is also known as visual saliency (Bruce & Tsotsos, 2009). Over the past decades, most research into gaze behavior seem to agree eye movements contain information in affluence about cognitive processes (Rayner, 1998). Hence, eye gaze behavior may be regarded as a valuable tool to understand expertise.

***Theories of visual behavior and eye tracking***

Eye-tracking allows for capturing where a person is looking at any given time. Various eye tracking studies have examined the differences in eye movement patterns between experts and novices in professional domains ranging from sports to medicine (Gegenfurtner, Lehtinen & Säljö, 2011). Examples of non-trivial findings have shown that elite golf players fixate their visual attention for a prolonged amount of time on the ball until they hit target, which is also known as the “Quiet Eye” concept (Vickers, 1992) another example in the surgical domain revealed that highly skilled surgeons as opposed to novices spend less time tracking the position of their tools during surgery (Vine et al., 2013).

Several theories that explain expertise through visual behavior have been outlined in a meta-analysis by Gegenfurtner et al., (2011): The information reduction theory (Haider & Frensch, 1999), the long-term working memory theory and the holistic model of image perception (Kundel et al. 2007). The holistic model of image perception proposes that experts are capable of widening their visual span by processing global information (Kundel et al. 2007). The long-term working memory suggests that experts, in contrast to, novices are capable of relying on their long term working memory as an extension of their working memory enabling faster processing of visual information due to retrieval structures (Ericsson & Kintsch, 1995). Finally, the most prevalent theory that offers an explanation for differences in gaze behavior between experts and novices is the information reduction theory (Haider & Frensch, 1999). This theory suggests that experts tend to reduce irrelevant information that needs to be processed by solely focusing on task-related visual stimuli and ignore unnecessary cues. Haider and Frensch (1999) showed in their work that expert air traffic controllers reduce the amount of visual information to be processed by solely focusing on task-related visual cues.

Often the aim of these studies is to understand the differences between experts and novices using their eye movement patterns. An illustration of why comparisons between experts and novices might be relevant can be found in laparoscopic surgery (surgery through small incisions). What makes laparoscopic surgery particularly interesting are the substantive results that lead to the contradictory situation of having favorable economic and societal outcomes related to fast recoveries for patients (due to small incisions), that comes at the cost of increased complexity of operations and higher skill demands for surgeons (Hermens, Flin and Ahmed, 2013; Reiley, Lin, Yuh and Hager, 2011).

These demands in combinations with technological improvements of eye movement trackers are a catalyst for increased interest in improvement of training for novice surgeons based on eye movements (Hermens et al., 2013). Substantial evidence of improved training as a result of mimicking visual behavior can be found in the work of Vine et al. (2013) who has shown that novices who received gaze training in order to adopt the same eye movements as experts obtained higher skill levels and moreover skills were retained for a longer period of time.

**Visual search within the CSI domain**

Reasons to believe that eye movement differences may extend to and have an important role within the CSI domain as a premise of the information reduction theory are the intertwined dynamics of selectively applying attention to objects, switching this attention between objects and fixating on certain areas of interest. These dynamics are inherent to examining a crime scene and are known as covert orientation (Hunt and Kingstone, 2003). The presence of these dynamics are corroborated by several studies. Baber and Butler (2012) showed that CSI experts tend to focus selectively on objects with “evidential value” and Watalingam, Richetelli, Pelz and Speir (2017) theorized that experts have a tendency towards hypothesizing in the process of interpreting evidence.

Given that CSI has a pivotal role to forensic analysis in the sense that it determines what elements of the crime scene is selected as evidence. Wrongful selection of evidence can lead to an increased workloads in laboratories, whereas a more effective CSI can speed up analysis (Watalingam et al., 2017).

Th

In addition, objective assessment of skill can lead to improved training as in Vine et al. (2013).

Whilst the aforementioned theories of visual behavior are corroborated in various domains such as chess (Sheridan & Reingold, 2014) and slalom skiing (Decroix et al., 2017), the applicability of the theories in each of these examples are dependent on the nature of the task in relation to domain specific characteristics (Prytz, Norén & Jonson, 2018).

Gegenfurtner et al. (2011) provides a taxonomy of task complexity drawn from the work of Wood (1986) and Campbell (1986) proposing that the complexity of a task can be ranked: only *viewing* is the simplest, followed by a moderately difficult *detection* task, a moderately difficult *decision* task and the hardest task a *problem solving* task. This suggests that the difference in expertise as measured by eye movements in the various domains may vary as a function of task complexity (e.g. eye-hand coordination in surgery, visual search).

An example can be found in the work of Prytz et al. (2018). They examined the differences between emergency professionals and novices in conducting a visual search task in order to identify an accident scene. They proposed that in the visual dynamics of selectively searching through task relevant areas are the most strongly linked to the information-reduction theory (Prytz et al., 2018).

However, investigation expertise within the CSI domain on the basis of eye tracking has been researched to a lesser extent in comparison to other domains. In Gegenfurtner’ et al. (2011) review of 71 studies, none was focused on the CSI domain. Previous research within the expert and novice paradigm with regards to visual behavior in a related field, forensic sciences, has studied expertise with regards to signature analysis (Dyer, Found & Rogers, 2006; Merlino, 2014) and fingerprint analysis (Busey et al., 2011).

More recently, Watalingam et al. (2017) examined whether eye movements could be used to distinguish expertise on the basis of sequences of eye fixation and found meaningful differences for sequences of eye fixations between expert and novice and the amount of time. In line with the information reduction hypothesis experts exhibit effective behavior in the sense that they spent a reported 88% of their time searching for evidence, whereas novices spent 86% of their time searching for evidence which is a small difference (Watalingam et al. 2017).

Around the same time, Ozger (2016), in the context of his PhD project, used eye tracking to understand expertise in the CSI field through eye movement patterns. In three experiments expert and novices were compared on search tasks differing in context (live examination versus static examination of photos). The focus of these experiments was to gain a better understanding of perceptual and cognitive expertise of crime scene investigators (See methods section for a more in depth explanation).

Additionally, a better understanding of expertise can lead to a more objective assessment of skill due to the ecological validity of eye measurements. Assessment based on results can have unfavorable and biased outcomes. It is reasonable that experienced CSI will take on the more challenging cases in the same way that more experienced surgeons are more likely the ones taking up more complex cases (Hermens, 2013).

**Gaze behavior and measures derived from Eye-tracking**

Gaze behavior is a complex source of information and analysis relies on a wide range of measures (Coutrot et al., 2018). The main measures found in literature belong to the oculomotor approach and include metrics such as fixations and saccades (Gegenfurtner et al., 2011; Ozger, 2016). An eye fixation is characterized by the maintenance of the visual focus on a single location and a saccade are small eye movements between fixations when a person relocates visual focus (Gegenfurtner et al., 2011). It is assumed that the lion’s share of visual information extraction takes place through covert attention during fixation (Ozger, 2016; Hunt & Kingstone, 2003).

One of the main favorable properties of oculomotor metrics according to Coutrot (2018) is the interpretability and the possibility to generalize these measures, whereas scan paths or graphs of eye movements are stimuli dependent. This allows for a data-driven approach, meaning that analysis can be done solely on eye metrics without considering the different points being looked at (Richstone et al 2010; Hermens, 2013). Table 1 shows an overview of the eye movement metrics related to the information reduction theory drawn from the overview found in the review of Gegenfurtner et al. (2011).

*Table 1*

*Operationalization of theories to eye movement measures*

*Adopted from Gegenfurtner et al. (2011)*

|  |  |  |  |
| --- | --- | --- | --- |
| **Explaining theory** | **Premise** | **Related eye movement metric** | **Definition** |
| Information reduction | Efficient allocation of attention by neglecting redundant visual information | Number of fixations  (relevant) | Number of fixations on areas that are relevant for the task. |
| Number of fixations  (irrelevant) | Number of fixations on area that are irrelevant for the task. |
| Fixation duration  (on relevant areas) | Time of one fixation on task relevant areas. |
| Fixation duration  (on irrelevant areas) | Time of one fixation on a task relevant area. |

**Current study**

Research on the subject has been mostly restricted to statistical comparisons of eye movement parameters, but ideally, one would like to determine how well skill can be measured using eye movement parameters (e.g., correct classification rates). The only work that uses classification methods in the context of eye movements and expertise appears to be that of Ahmidi et al. (2010; 2012) and Richstone et al. (2010), examining classification of expert and novice surgeons. While Watalingam et al. (2017) distinguished expertise using scan paths. No attempt has been undertaken to evaluate whether eye movement parameters can be used to classify.

Secondly, less attention went out to evaluate whether the assumptions found in the information reduction theory and corroborated in fields with similar dynamics (e.g. in Prytz et al. 2018) extend to the CSI domain.

The present research aims to examine whether eye movements between expert and novice CSIs differ, whether eye movements can be used to classify expert and novice CSIs, and whether such classification depends on the task (real world eye tracking, on-screen crime scene investigation, or a change detection task). Central to this thesis are the premises found in the Information-reduction theory. Based on the information reduction theory (reference) we expect that experts show shorter durations on task irrelevant areas and longer fixations on task relevant areas (evidence and exit & entry points). Another objective of this research is to evaluate whether eye movements of experts and novices contain enough information to distinguish between the two groups by applying machine learning techniques.

The overall goal is to contribute to our understanding of what constitutes an expert in the field of CSI in terms of eye movement parameters, possibly leading to the objective evaluation of skills and improved training of novices by the means of gaze behavior.

**Problem statement and research questions**

This thesis follows an exploratory approach of evaluating expertise effects in the context of CSI through eye movement metrics. The problem statement is formulated as followed:

*Can eye movement parameters be used to distinguish between experts and novices in the CSI domain?*

First of all, we need to evaluate whether the premises of the information reduction theory extends to the CSI domain. These premises withheld that experts have the tendency to focus their attention selectively to task relevant objects and ignore redundant objects. This resonates into visual behavior as a difference between numbers of fixations. In accordance with Gegenfurtner et al. (2011) we hypothesize that experts have more fixations on relevant objects and fewer fixation on redundant objects. This assumption is leads us to the following research question:

RQ 1) *Are there differences in the number of fixations on task relevant and task redundant areas between CSI experts and novices as a consequence of the information reduction hypothesis?*

* Data-Exploration
* Evaluate this question on Eyelink dataset.

The findings in the first research question guide our exploration with machine learning techniques in order to evaluate whether eye movement measures contain enough information to distinguish expertise. Hence, the second research question is as follows:

RQ2) *Do eye movement measures within the CSI domain contain enough information in order to predict Skill (Expert or novice) on the basis of eye movement parameters and across tasks?*

* *-> Random-Forrest for variable importance & feature selection.*
* *-> Classification with*
* *Classification for Change - blindness*

*&*

* *Visual search task*

The research follows an exploratory approach employing machine learning techniques. Therefore it may be relevant to state the last research question.

RQ3) *which patterns prevail in the eye moveme**nt data for experts and novices?*

# Methods

**Dataset**

Collection of data and Lincolnshire Experiments  
 This research utilizes data provided by Ozger (2016) from his PhD thesis project (Lincoln University, UK). The data consist of eye movement recordings from three experiments: (1) real world crime scene investigation (simulated crime scene), (2) on-screen crime scene investigation (photographs), (3) an on-screen change detection task (comparing crime scenes and other scenes). The first experiment measured eye movements of crime scene investigators (15 in total, 8 with expertise in violent crimes and 7 with mostly expertise in non-violent crimes, such as burglary) while exploring a staged crime scene wearing a mobile eye tracker. Participants were randomly assigned a story and then examined the same scene with no time limitation, after which they were asked to retrospectivelythinkaloud while watching their recorded eye movements**.** Eye movements for each video frame were manually assigned to one of severalAreas of interest (AOI), which included specific objects such as phone and shoes, which were sorted into broader categories, such as entry and exit points and evidence. These areas were marked as an AOI if a participant viewed it for longer than 50 ms.

The second experiment was conducted in a lab using photographs presented on a computer screen. This provided a more stable environment and ensured that participants viewed exactly the same stimuli. 36 participants took part across four groups: nine experts, nine intermediate investigators, nine first years’ forensic students (novices) and nine participants without experience. The participants viewed three different photos on a computer screen and their task was to point out evidence by clicking with a mouse on objects. Afterwards they had to formulate their thought process. Scenes were color coded for the various AOIs and eye movements assigned automatically. In subsequent analyses AOIs where grouped into evidence, entry and exit point of a scene.

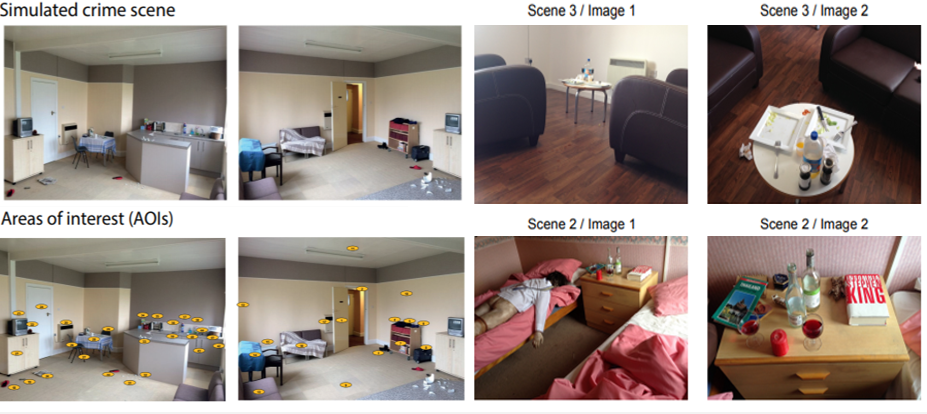
Lastly, the third experiment examined change blindness in two types of scenes: domain specific (14 images) and a domain unspecific (14 images) across nine experts and eleven novices. Change blindness can be described as a failure to notice an otherwise blatant change in a scene or environment (Mack and Rock, 1998). Within the change blindness experiment the flicker approach (display of short blank screen) was used to switch between the photos. In order to avoid motion cues, a blank screen was shown between the photos, each shown for 500 ms. Participants were instructed to click as fast as possible on the button as soon as they detected a change. Table 2 shows an overview of each experimental set-up.

*Table 2*

*Overview of experiments*

|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment** | **Participants** | **Stimuli** | **Type of data** |
| 1 | 8 experts and 7 novices | Real world scene with 2 scenarios | Frame by frame ROI of fixation |
| 2 | 36 across 4 levels | Photographs of crime scenes | Fixation by fixation ROI codes |
| 3 | 9 experts and 11 novices | Crime scene and 14 other change blindness images | Fixation on ROIs for crime scenes, accuracy, fixation duration |

*Stimuli*



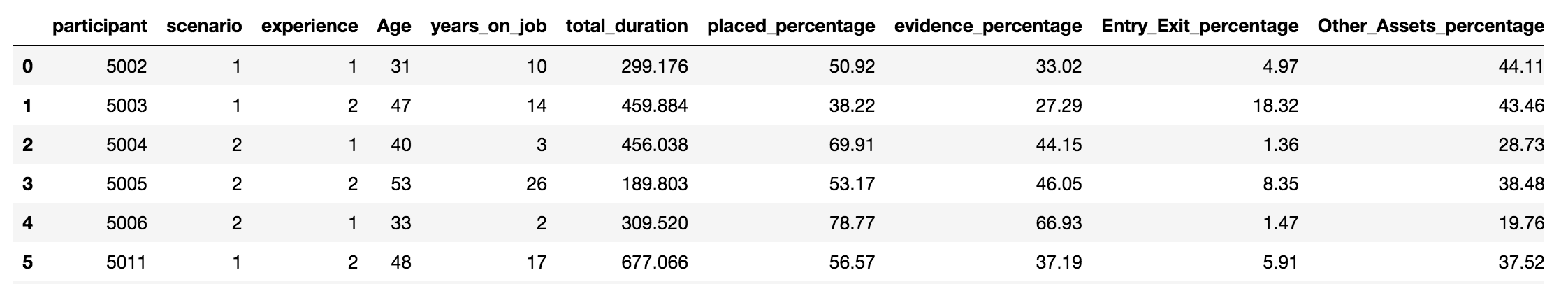
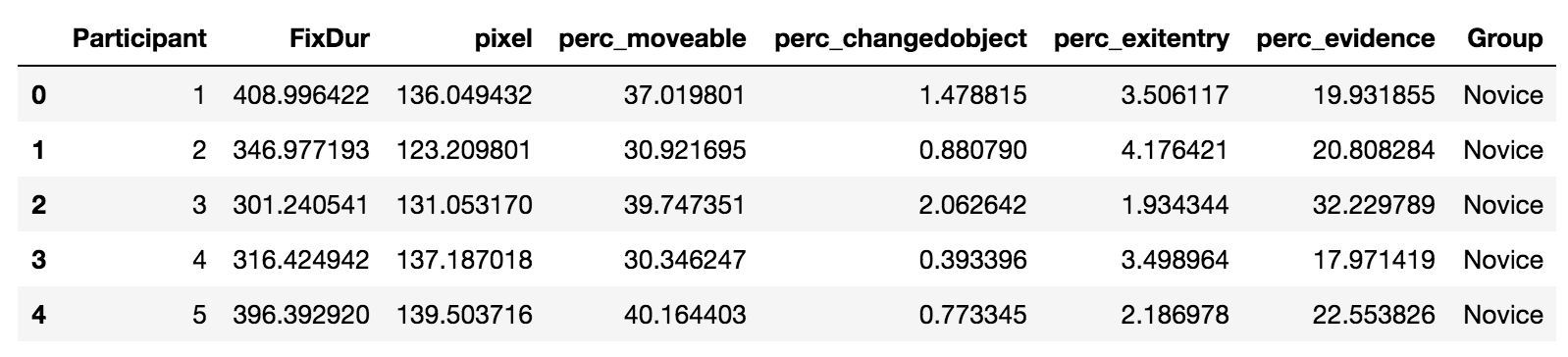
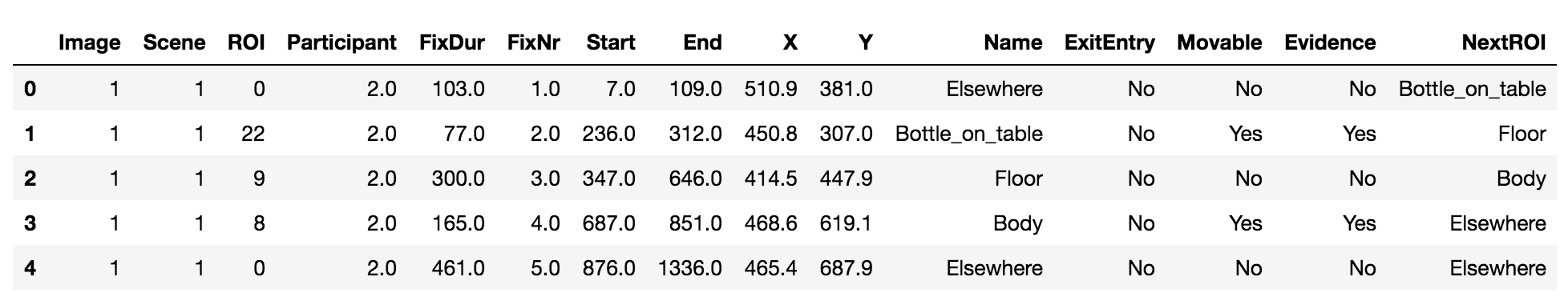
*Figure 1.* Images of visual stimuli in experiment 1 and 2. Figure 1a illustrates the live examination task. ‘live crime scene’ is one of the scenes which participants roamed and ‘AOIs’ shows the marked areas. Figure 1b are the visual stimuli of experiment 2 in which photographs where shown. From left to right the same photos are shown from different angles. Adopted from M. Ozger, (2016), *Paying attention to the evidence: a comparison of perception and decision making processes in novice and experienced scene of crime officers using eye tracking in simulated crime scene scenarios*, PhD thesis, University of Lincoln.



*Figure 2.* 3 Stimuli used in the change blindness experiment. Red circles indicate the objects or area that where modified during the change. Adopted from M. Ozger, (2016), *Paying attention to the evidence: a comparison of perception and decision making processes in novice and experienced scene of crime officers using eye tracking in simulated crime scene scenarios*, p.84. PhD thesis, University of Lincoln.

*Table 3*

*Overview of datasets*

****

Algorithms

**Classification and Regression Trees (CART)**

*Decision Trees*

The goal of classification with decision trees is to partition the data into disjoints sets (Rafalab, @@). Decision trees learn by inducing rules from observations in order to achieve this goal (Quinlan, 1985). In the classification setting, each observation has features that takes on a set of discrete or continues values (Quinlan, 1985). A decision tree splits based on question with regards to these variables.

Table 1.

Example training set

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Observation** | **features** | | | | **class** |
|  | Windy | Temparature | Humidity | Rain |  |
| 1 | Yes | Hot | High | Yes | Play outside |
| 2 | No | Cold | Normal | No | Don’t play |
| 3 | Yes | Mild | Low | yes | Play outside |

These features and their values propose an inference rule to classify the object. The weather example found in the work of Quinlan (1985) is adopted in table 1 with some modifications to illustrate this. For an example if it rains outside and the temperature is hot and windspeed is 11 and it rains outside. We can see that observation plays outside. Essentially these rules can be expressed as a decision tree. See figure 3.

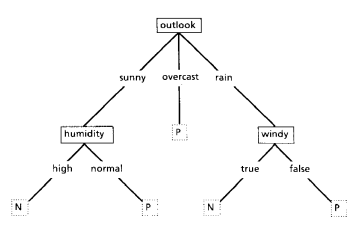


Figure 3

The N and P are called leafs and the nodes represent a feature and a “test” too which an answer decides which edge should be followed. Following from the top of the tree, the root, to a leaf. Each node represents a question through which an observation flows through downwards until it comes to a halt in an end leaf which represents the class of the observation. Adapted from *Induction of Decision Trees,* by R. Quinlan, 1986, Machine Learning. 1. P. 87.

In essence, decision trees capture relevant relations (rule) between the attributes and a class in order to generalize to unseen observations.

GINI - coefficient & TREE CONSTRUCTION

As mentioned before, a decision tree classifies by assigning a new observation to the most occurring class in a leaf (James, Witten, Hastie & Tibshirani, 2014). By maximizing the homogeneity at each node split (question) we can be more certain about the classification (James et al., 2014).

Homogeneity is also defined as pureness in terms of classes. In this regard the Gini coefficient is a metric to asses how purely a node splits the classes and resembles the quality of the split. The goal here is to obtain unmixed groups with respects to the classes. The Gini-coefficient is defined as followed:

C:\Users\u283203\Downloads\CodeCogsEqn.png

***Eq. 1***

*PmK* denotes the fraction of observations in the *M*th region from the *K*th class (James et al., 2014). The Gini-index is computed by multiplying the *PmK* by the inverse of the *Pmk* and sum it for all classes. A small value for the Gini index points at that observations in the nodes are predominantly homogeneous in terms of the class and a high index relates to mixed classes in the training data after a split (James et al., 2014).

The concept of Information gain is related to decrease in impurity in the sense that features that decrease impurity as measured by the Gini-index are informative (Breiman, 2001). The Decision Tree algorithm recursively tests all features at each node and prioritize on Informativeness.

The scikit-learn implementation computes feature importance as the normalized reduction in Impurity as measured by the Gini-coefficients (sklearn, decisiontreeclf documentation).

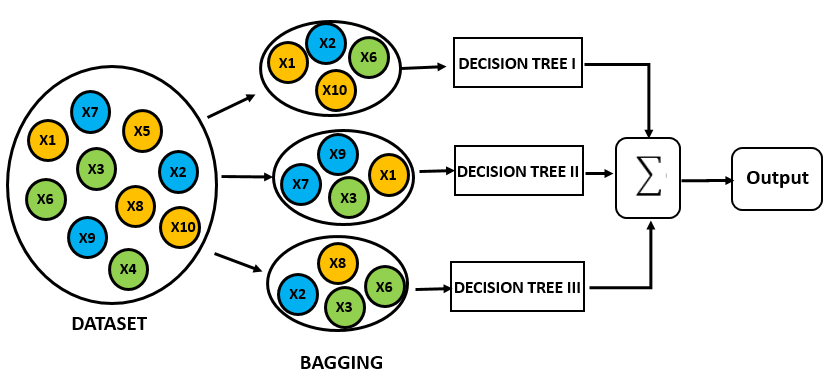
*Ensemble methods: Exta trees and Random forests*

The Random Forest is an ensemble method that combines multiple Decision Tree classifiers (or Regressors) trained on random samples from the training data and averages the output (Breiman, 2001; James et al., 2014). Figure In addition, the Random Forest decorrelates features by considering only a subset of the predictors in each tree that is part of the ensemble (Breiman, 2001). A common threshold is to use the square root of the total predictors in the predictor space (James et al., 2014)

**C:\Users\u283203\AppData\Local\Microsoft\Windows\INetCache\Content.Word\CodeCogsEqn.png**

**Eq.2**

By constantly considering a random sample of predictors and not the full available predictor space the random forest ensures that predictors with high information gain are not consistently used as the root for every tree (James et al., 2014). This would lead to highly correlated trees and a smaller reduction in variance (James et al., 2014)



*Figure 4*

*The Random Forest reduces variance, by bootstrapping subsets of the training data, building a decision tree and aggregating these results. This procedure is also known as bagging (James et al., 2014).*

In terms of variable importance, the Random Forest averages the amount that the Gini Index is decreased as a consequence of using that feature in a split averaged over all trees in the ensemble (James et al., 2014).

The ExtraTrees classifier is another ensemble method that is very similar to the Random Forest (Geurts, Ernst & Wehenkel, 2005). It differs in the way it considers its splits. Whereas the Random Forest splits @@

K Nearest Neighbors

KNN

* ISL
* Mastering ML
* Sklearn implementation

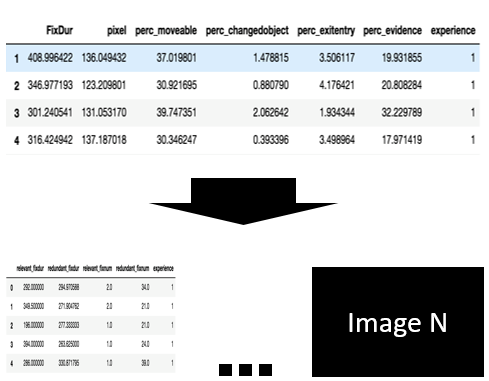
**Experimental Procedure**

As Coutrot (2017) mentioned, eye movements are dynamic and subject to “idiosyncratic” influences. In addition, we’ve established that saliency strengthens bottom-up effects (Gegenfurtner et al., 2011). Images with salient features are known to attract attention as opposed to images with no salient features. Within crime scene investigation the saliency of a scenery can be objects such as knives or for an example a puddle of blood. Eye coordinates and fixation durations are therefore image dependent.

Based on that knowledge, a twofold approach, globally and locally is justified. Figure 5 illustrates how the analysis was conducted. Models were fit both on aggregated data and the relevant and redundant fixation duration and number of fixations per images (locally).

Figure 5

Global versus Local approach



Preprocessing and Feature selection

The variables used in this study were selected in accordance with the premises of the information reduction theory and adopted from Gegenfurtner et al. (2011) with some augmentations. As mentioned before we hypothesize that experts have more fixations on relevant objects and fewer fixation on redundant objects. Relevant fixations (duration and number of fixations) are computed by amounting the numbers or the time spent looking at evidence. Redundant fixations (duration and number) is computed by amounting the number and time spent at not looking at evidence.

*Table 2*

Following the knowledge that saliency and bottom effects can differ per visual stimuli. Measures were computed per image. Figure 1 shows this schematically.

|  |  |  |
| --- | --- | --- |
| Measures | Global | * Fixation duration * Pixel coordinates * % Looked at Moveables * % Looked at Changed objects * % Looked at Evidence * % Looked at Exit and Entry points |
| **Local** | * Relevant fixation duration * Relevant number of fixations * Redundant fixation duration * Redundant number of fixations |

Principal component analysis

* *Compared minmax versus standard scaled. On both tasks. As can be seen in the picture 98% of the variance is contained in 4 PCA’s for both MM and STD scaled PCA’s. The same threshold is contained within 6 principal components.*
* *KNN and PCA are distance based methods and therefore perform better if everything is scaled. Two methods were performed. MM versus STD and while the trajectory looks different. They arrive at the same amount of PCAs.*

PCA’s

* Dimensionality reduction
* Maximizes variance [stackexchange]. With unscaled features the loadings of the principal components can

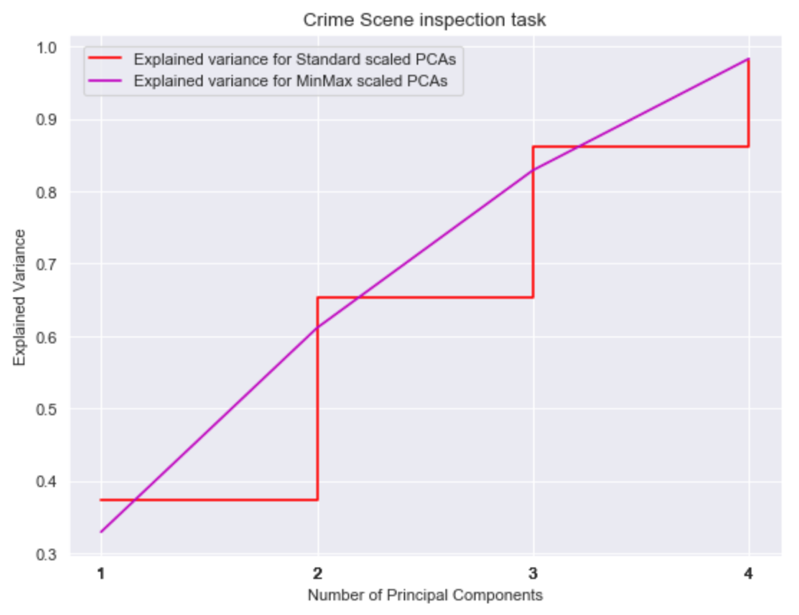
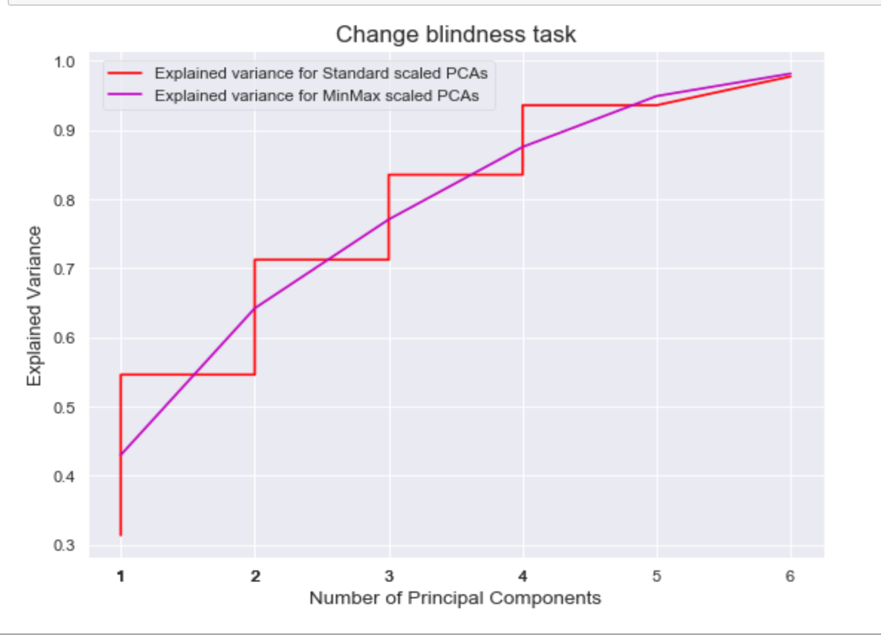
<https://stats.stackexchange.com/questions/69157/why-do-we-need-to-normalize-data-before-principal-component-analysis-pca>

* Plots revealed that data was non-linear
* So I relied on nonlinear methods.
* KNN.

*MinMax and StandardScaled.*

* + And it’s influence on PCA
  + ADD formulas for MinMax and Standard Scaled.

Screeplots were used for both datasets in order to determine the amount of PCA’s needed for prediction. Figure @@ shows these plots.

******>>> Tekst vergroten op het plaatje

***Hyperparameters and Algorithm training***

Following the aforementioned similarity between both classifiers and the possibility to run the treebased models in parallel. Crossvalidation was performed using a gridsearchCV functionality in sklearn. Table 3 shows the parameter tuning for both the Tree Classifiers and the KNN- score.

The Random Forest and Extra Trees classifier were trained using Grid search cross validation. The parameters that were observed can be found in table 3.

*Parameter tuning in tree-based methods*

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Parameters** | **Definition** | **Grids for both classifiers** |
| Max\_depth |  | [2,3,4,5] |
| Min\_samples\_split |  | [2,3,4,5] |
| Min\_samples\_leaf |  | [2,3,4,5] |
| Max\_features |  | [2,3,4,5] |
| N\_estimators |  |  |
|  |  |  |

* Run model again with a lot of trees. The bigger the better apparentely. Detect when it converges.

On random forests with small data.

<https://stats.stackexchange.com/questions/192310/is-random-forest-suitable-for-very-small-data-sets>

Why to not OVERTUNEEEE

<http://www.jmlr.org/papers/v3/reunanen03a.html>

*Change blindness and visual search parameter tuning for KNN*

The most important parameter in the K nearest neighbor’s algorithm is obviously the K. Parameter tuning for K showed that in the Change Blindness task the MinMax scaler seem to classify the objects with maximum accuracy.

Table 4:

Parameter tuning for KNN. For both datasets the data was standard scaled to mean and standard deviation and MinMax scaled between 0 and 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Value of K | Task: Change Blindness | | Task: Visual Search | |
|  | Standardized | MinMax | Standardized | MinMax |
| 3 | 0.6 | 1 | 0 | 0.0 |
| 5 | 0.2 | 1 | 0 | 0.0 |
| 7 | 0 | .6 | 0.1 | 0.1 |
| 9 | 0 | 0 | 0.1 | 0.1 |

In order to evaluate the best K. The data was divided in a train/validation set 67% (22 instances) and a testset 33% (11 instances) for the Visual search task. In order to evaluate the best K for the CB dataset the dataset was split into 67% train/validation set (13 instances) and a testset 33% (7 instances)

5-fold cross validation has shown that K=3 for the CB task and K= 7 for both standard and minmax scaled for the visual search task. The best K is used for predictions afterwards.

**Scoring and classification**

**Metrics**

**On F1 measures:**

[**https://stats.stackexchange.com/questions/49226/how-to-interpret-f-measure-values**](https://stats.stackexchange.com/questions/49226/how-to-interpret-f-measure-values)

**+**

**Data-skeptic podcast.**

**Measure of preciseness but also robustness\**

**Results**

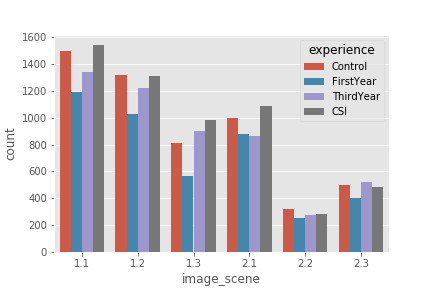
RQ 1) *Are there differences in the number of fixations on task relevant and task redundant areas between CSI experts and novices as a consequence of the information reduction hypothesis?*

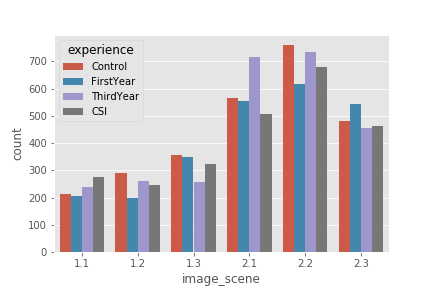
1. **RQ1**

*Figure @@*

*Relevant fixations (on Evidence) between Control, First year Third Year and CSI*

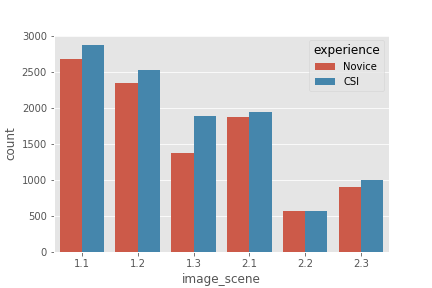
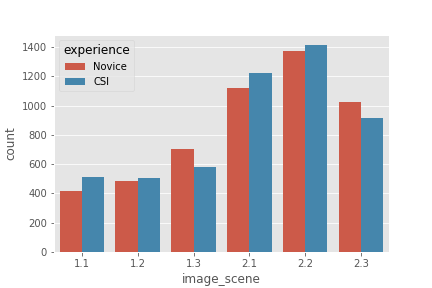
*Left is relevant and right is redundant.*



Figure @@

Comparison of number of fixations between both experts and novices.

Left is relevant, right is redundant



Here we saw that fixation duration and, percentage of time looked at exitentry points, percentage of time looked at moveables and pixel coordinates where the best predictors for a certain class.

**RQ2) *Do eye movement measures within the CSI domain contain enough information in order to predict Skill (Expert or novice) on the basis of eye movement parameters ?***

* In terms of Feature importance relevant fixation number doesn’t do very well globally. Both with 4 groups as with 2 groups. Fixation duration actually contains more information looking at Eyelink.

Classification algorithms

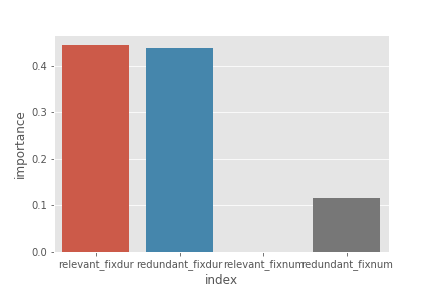
From the decision tree family the sci-kit implementation of the Random Forrest classifier.

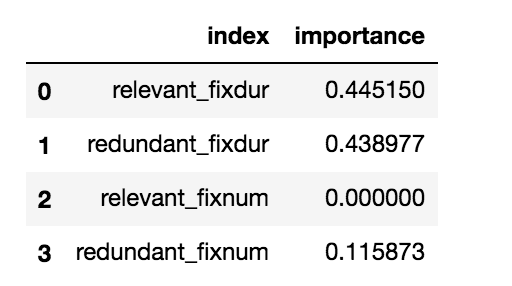
The methods and code can be found on Github @@Link.

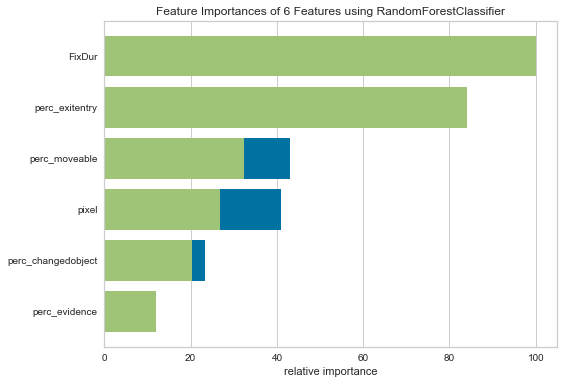
***Global feature importance -- DT***

Figure @@@

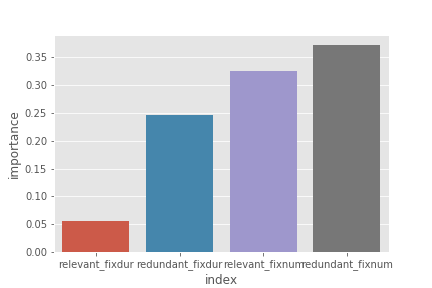
With gini



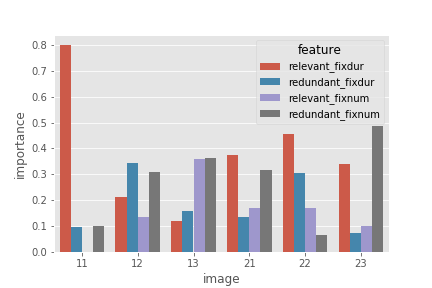
2



With entropy:



* ***Local feature importance***



The findings in the first research question guide our exploration with machine learning techniques in order to evaluate whether eye movement measures contain enough information to distinguish expertise. Hence, the second research question is as follows:

<http://gree2.github.io/python/2015/05/05/python-seaborn-tutorial-controlling-figure-aesthetics>

The research follows an exploratory approach employing machine learning techniques. Therefore it may be relevant to state the last research question.

RQ3) *which patterns prevail in the eye movement data for experts and novices?*

**Discussion**

* It is unclear whether the eye measurement differences are really symptoms of a better csi. While the methods used in this research point out which eye movement patterns are important and which eye movement pattern measures are consistently different between experts and novice csi. It remains unclear to which extent these measurements reflect upon certain acquired skill.
* The premises of these theories resonates by moderating the expertise effect. When visual stimuli contain an abundance of information. In contrast to novices, experts can rely on mental cues encapsulated in the long term working memory while processing information (Gegenfurtner, Lehtinen & Saljo, 2011). Hence redundant information and in visual stimuli is disadvantageous to novices.
* A second aspect that can modify the gap between experts and novice eye movement patterns is the nature of the task. In eyetracking research a variety of different tasks have been used without keeping account for the influence of task complexity on detected difference sizes ( Gegenfurther et al., 2011).
* Why is such an approach to skill assessment important? It is critical to assess surgeon skill for a number of reasons. Objective measurement of skill is necessary to monitor the progression throughout surgical training programs and a prerequisite for meaningful credentialing. Ensuring surgical skill and competency is critical to limit the incidence of iatrogenic injury and medical errors, which are alarmingly common.1 Unfortunately, however, the assessment of surgical skill remains a rudimentary science. Currently, the best available method appears to be direct observation of operative performance with a global rating scale.32 While this method has validity, direct observation is subject to biases, requires expert “judges” to be present, and is associated with significant time demand. Richstone et al (2010)

**Possible implications**

* : information reduction theory (might extend to csi since dynamics are similar). Lessons can be drawn from studies in which scanning behaviors can be used for novices to adopt. Mostly oculomotor eye movement parameters can be used in teaching.

**Further research**

* Evaluate the measures on a bigger dataset with more participants and see if they are also consistent there.
* include event-driven approach in your research.

**Limitations**

* First, **participants in this study were sampled from a fixed geographical** location, which confines the above conclusions to a single cohort, and therefore not necessarily representative of the entire population of crime scene examiners.
* One of the main problems with research in the expert novice paradigm is the scarcity of experts.

Mills et al., (good example paper to describe methods section)

<http://www.educationaldatamining.org/EDM2016/proceedings/paper_143.pdf>

* Feature selection was used on the training set of each crossvalidation fold (see below). Features were ranked using correlation-based feature selection (CFS) [15] from Weka and the top 30%, 50%, or 80% of features ranked were retained.

Other methods for feature\_selection could have been tried.

* **variability of real crime scenes** Third, a homicide/shooting reconstruction scene was evaluated, and as such, the results found here may not be transferable to all types of crimes (note that 22 of the 32 participants in this study (69%) reported experience with shooting scenes).
* Only focus on subset of eye gaze information the direct parameters. while there are others such as spatial distribution, string-based and geometric and probabilistic approaches.
* this was data driven maybe because (pro’s of event driven as opposed to data driven [find out what they are]) were prevalent in this study.
* [why the mobile eyetracking might not show good results.] such as the spatial bias to the center region [[7]](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000791#pcbi.1000791-Tatler1) and geometric properties of saccades [[8]](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000791#pcbi.1000791-Brockmann1)

# 

*Implications:*

* *Standardized surgical skills test in a trainer or simulator or alternatively submit eye metric performances for certification or review Richstone et al 201*

*[Extra]*

In accordance with the long-term working memory theory we hypothesize that experts in the CSI domain display overall shorter fixation duration **h1**.

In accordance with the holistic approach to perception we estimate that expert CSIs will have shorter saccade amplitudes and need shorter time to fixate their first.

# References

Ahmidi, N., Hager, G. D., Ishii, L., Fichtinger, G., Gallia, G. L., & Ishii, M. (2010). Surgical task and

skill classification from eye tracking and tool motion in minimally invasive surgery. *In Medical image computing and computer-assisted intervention*. Vol 13, 295–302 Retrieved from

Ahmidi N, Ishii M, Fichtinger G, Gallia GL, Hager D. (2012). An objective and automated

method for assessing surgical skill in endoscopic sinus surgery using eye‐tracking and tool‐motion data *Int Forum Allergy Rhinol*. 2012;2:507–515 Retrieved from <https://doi.org/10.1002/alr.21053>

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), 413–24. Retrieved from <http://journals.sagepub.com/doi/10.1177/0018720812440577>

[Breiman L](https://en.wikipedia.org/wiki/Leo_Breiman) (2001). "Random Forests". [*Machine Learning*](https://en.wikipedia.org/wiki/Machine_Learning_(journal)). **45** (1): 5–32. [doi](https://en.wikipedia.org/wiki/Digital_object_identifier):[10.1023/A:1010933404324](https://doi.org/10.1023%2FA%3A1010933404324).

Turvey, B.E. Freeman, J. (2017). Chapter 5 - Crime Scene Investigation and Analysis. *Forensic.*

*Investigations.* 91-123. Retrieved from <https://doi.org/10.1016/B978-0-12-800680-1.00005-5>.

Coutrot, A., Hsiao, J.H. & Chan, A.B. (2018). Scanpath modeling and classification with hidden

Markov models. *Behavior Research Methods* 50: 362. Retrieved from <https://doi.org/10.3758/s13428-017-0876-8>

Busey, T.A. & Parada, F.J. (2010). The nature of expertise in fingerprint examiners. *Psychonomic*

*Bulletin & Review* 17: 155. https://doi.org/10.3758/PBR.17.2.155

Decroix, M., Norjali Wazir, M.R.W., Zeuwts, L., Deconinck, F. F.J.A., Lenoir, M.,

Vansteenkiste, P. (2017). Expert – Non-expert differences in visual behaviour during alpine slalom skiing, *Human Movement Science*. 55, 229-239. Retrieved from <https://doi.org/10.1016/j.humov.2017.08.012>

Dyer, A. G., Found, B., & Rogers, D. (2006). Visual Attention and Expertise for Forensic

Signature Analysis. Journal of Forensic Sciences, 51(6), 1397-1404. Retrieved from doi:10.1111/j.1556-4029.2006.00269.x

Campbell, D. J. (1986). Task complexity: A review and analysis. Academy of Management Review, 13,

40– 52. Retrieved from <https://doi.org/10.5465/amr.1988.4306775>

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of

visualizations: A meta-analysis of eye-tracking research in professional domains. Educational Psychology Review, 23(4), 523-552.

<http://dx.doi.org/10.1007/s10648-011-9174-7>

Hermens, F., Flin, R., & Ahmed, I. (2013). Eye movements in surgery: A literature

review. Journal of eye movement research, 6(4), [4]. Retrieved from <http://dx.doi.org/10.16910/jemr.6.4.4>

Hunt, A. R.; Kingstone, A. (2003). "Covert and overt voluntary attention: Linked or

independent?". Cognitive Brain Research. **18** (1): 102–105.Retrieved from <https://doi.org/10.1016/j.cogbrainres.2003.08.006>

Jarodzka, H. Scheiter, K. Gerjets, P., Gog., T. (2010) In the eyes of the beholder: How experts and

novices interpret dynamic stimuli. *Learning and Instruction*. Volume 20, Issue 2, Pages 146-154.

Retrieved from https://doi.org/10.1016/j.learninstruc.2009.02.019

Rayner, K. 1998. Eye Movements in Reading and Information Processing: 20 Years of Research.

Psychological Bulletin. 124, 3 (Nov. 1998), 372. Retrieved from <http://dx.doi.org/10.1037/0033-2909.124.3.372>

Richstone L, Schwartz MJ, Seideman C et al: Eye metrics as an objective assessment of surgical

skill. Ann Surg, 2010; 252(1): 177–82. Retrieved from [http://dx.doi.org/**10.1097**/**SLA**.**0b013e3181e464fb**](http://dx.doi.org/10.1097/SLA.0b013e3181e464fb)

Rizzolatti G, Riggio L, Dascola I, Umilta C (1987) Reorienting attention across the horizontal and vertical

meridians: evidence in favor of a premotor theory of attention. Neuropsychologia 25: 31–40.

Ozger, M. (2016 ) Paying attention to the evidence: a comparison of perception and decision

making processes in novice and experienced scene of crime officers using eye tracking in simulated crime scene scenarios. PhD thesis, University of Lincoln. Retrieved from <http://eprints.lincoln.ac.uk/26649/>

Prytz., E.G., Norén, C., Johnson., C. (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* Retrieved from [https://doi.org/10.1177/0018720818788142](https://doi.org/10.1177%2F0018720818788142)

Vickers, J.N. (1992). Gaze control in putting. *Perception 21, 117-132.* Retrieved from

<https://doi.org/10.1068/p210117>

Vine, S. J., Chaytor, R. J., McGrath, J. S., Masters, R. S., and Wilson, M. R. (2013). Gaze

training improves the retention and transfer of laparoscopic technical skills in novices. *Surgical Endoscopy*. 27(9), 3205–3213. Retrieved from <https://doi.org/10.1007/s00464-013-2893-8>

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. Retrieved from <https://doi.org/10.1016/j.forsciint.2017.08.012>

Wood, R. E. (1986). Task complexity: Definition of the construct. Organizational Behavior and

Human Decision Processes, 37, 60–82. Retrieved from <https://doi.org/10.1016/0749-5978(86)90044-0>

Bruce and tsosos

chrome-extension://laookkfknpbbblfpciffpaejjkokdgca/dashboard.html

CLF and gini impurity

<https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier.feature_importances_>

Jocelyn D’souza picture of randomforrests

<https://medium.com/greyatom/a-trip-to-random-forest-5c30d8250d6a>

Quinlan, Ross. (1986). Induction of Decision Trees. Machine Learning. 1. 81-106. Retrieved from <https://doi.org/10.1023/A:1022643204877>

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2014. An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Incorporated.

P. Geurts, D. Ernst., and L. Wehenkel, “Extremely randomized trees”, Machine Learning, 63(1), 3-42, 2006.