Accepted Manuscript

Title: Forensic intelligence framework. Part II: Study of the main generic building blocks and challenges through the examples of illicit drugs and false identity documents monitoring

Author: Simon Baechler Marie Morelato Olivier Ribaux Alison Beavis Mark Tahtouh Paul Kirkbride Pierre Esseiva Pierre Margot Claude Roux

PII:	S0379-0738(15)00087-0
DOI:	http://dx.doi.org/doi:10.1016/j.forsciint.2015.02.021
Reference:	FSI 7924

To appear in: *FSI*

Received date:	1-10-2014
Revised date:	30-1-2015
Accepted date:	22-2-2015

Please cite this article as: S. Baechler, M. Morelato, O. Ribaux, A. Beavis, M. Tahtouh, P. Kirkbride, P. Esseiva, P. Margot, C. Roux, Forensic intelligence framework. Part II: Study of the main generic building blocks and challenges through the examples of illicit drugs and false identity documents monitoring, *Forensic Science International* (2015), http://dx.doi.org/10.1016/j.forsciint.2015.02.021

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Forensic intelligence framework. Part II: Study of the main generic building blocks and challenges through the examples of illicit drugs and false identity documents monitoring

Simon Baechler[a,b], Marie Morelato[c,*], Olivier Ribaux[a], Alison Beavis[c], Mark Tahtouh[d], Paul Kirkbride[e], Pierre Esseiva[a], Pierre Margot[a], Claude Roux[c]

[a] Ecole des Sciences Criminelles, Université de Lausanne, 1015 Lausanne, Switzerland
[b] Service forensique, Police neuchâteloise, Rue des poudrières 14, 2006 Neuchâtel, Switzerland
[c] Centre for Forensic Science, University of Technology, Sydney, Broadway, NSW, Australia
[d] Forensics, Australian Federal Police, Sydney, NSW, Australia
[e] School of Chemistry and Physics, Flinders University, Bedford Park, South Australia, Australia
[*] Corresponding author. Tel.: +61 40 535 55 49.
E-mail address: Marie.Morelato@uts.edu.au (M. Morelato)

Forensic intelligence framework. Part II: Study of the main generic building blocks and challenges through the examples of illicit drugs and false identity documents monitoring

Abstract

The development of forensic intelligence relies on the expression of suitable models that better represent the contribution of forensic intelligence in relation to the criminal justice system, policing and security. Such models assist in comparing and evaluating methods and new technologies, provide transparency and foster the development of new applications. Interestingly, strong similarities between two separate projects focusing on specific forensic science areas were recently observed. These observations have led to the induction of a general model (Part I) that could guide the use of any forensic science case data in an intelligence perspective. The present article builds upon this general approach by focusing on decisional and organisational issues. The article investigates the comparison process and evaluation system that lay at the heart of the forensic intelligence framework, advocating scientific decision criteria and a structured but flexible and dynamic architecture. These building blocks are crucial and clearly lay within the expertise of forensic scientists. However, it is only part of the problem. Forensic intelligence includes other blocks with their respective interactions, decision points and tensions (e.g. regarding how to guide detection and how to integrate forensic information with other information). Formalising these blocks identifies many questions and potential answers. Addressing these questions is essential for the progress of the discipline. Such a process requires clarifying the role and place of the forensic scientist within the whole process and their relationship to other stakeholders.

Keywords: forensic science; comparison process; organisation and decisions; model; integration

1. Introduction

The fundamental principle of forensic intelligence is that, instead of treating each case individually with the aim of assisting the court (i.e. evidential focus), a multi-case focus and more holistic approach based on the study of crime phenomena should be followed [1, 2]. The structured and systematic exploitation of crime traces is essential to produce knowledge that will guide strategic, operational and tactical decisions, in particular in models such as intelligence-led policing [3]. The main objective of such models is to monitor repetitive crimes that are evolving and complex due to their underlying organised nature. However, such clues do not represent the whole crime picture and a collaborative approach is required to provide actionable intelligence to decision makers.

In a previous paper [4], we described the induction process that led to the proposition of a forensic intelligence framework. Not only will the implementation of a general model break barriers between specific fields of study in forensic science and intelligence, but it will also help solve issues that are common across crime and trace types (hence the name 'transversal'). Indeed, a transversal model has the potential to offer a common vocabulary and an integrated framework, and will also assist in defining cross-discipline difficulties. It was observed that fundamental issues were treated in a similar way in two apparently different areas (i.e. illicit drugs and false identity documents). The general framework proposed in Part I [4] is a first significant step towards defining forensic intelligence, situating its role in policing, and exposing the potential opportunities and limitations.

The framework serves as a support for the development, implementation and evaluation of specific intelligence processes. It helps in the making of good and objective decisions about the way to elaborate and implement its particular components (or building blocks) and defining its relations with other information processes in order to maximise its overall efficiency. Part II further develops the general framework by exploring the main generic building blocks presented in Part I. The objective here is to further develop the modelling and generalisation efforts initiated in Part I and to provide illustrations of the potential use and limitations of the transversal model through two independent fields of application, i.e. illicit drugs and false identity documents. This contribution also aims to highlight the outcome of these formalisation efforts, which is to bring together in a common framework a qualitative approach (used to build the framework), quantitative approaches (i.e. metrics, scores, threshold values, error rates) and a Bayesian approach.

Part II proposes scientific and rational criteria useful to properly conceive and operate a forensic intelligence system, and to compare and assess alternatives. Based on these criteria, the paper then explores the decision points that are crucial in defining the process architecture as well as in assisting in making objective decisions in real caseworks. Building blocks related to the comparison process and the evaluation systems are first presented. The development of these particular blocks is mainly driven by forensic science and those blocks work relatively independently from general intelligence and investigative data. This separation is however neither logical nor practical when other building blocks are considered as they concern the many people and organisations that are involved in the whole forensic intelligence process. The reflections regarding these other components or building blocks must thus be seen as shared by all participants collaborating in the overall process.

2. Relevant criteria in conceiving and operating a forensic intelligence system

Forensic intelligence ultimately serves different objectives in a wide variety of operating contexts where decisions are often of a different nature than evidence-based court decisions [5]. Systems implementing the forensic intelligence process must be pragmatic enough to sustain uncertain reasoning while remaining scientifically rigorous and controllable. To cope with these constraints and manage risks of reasoning and acting under potentially false hypotheses, it is argued that a balance must be struck between four general parameters: *credibility*, *integrity*, *timeliness and flexibility* [6-9]. The performance of any forensic intelligence system as well as its building blocks can be assessed using these four parameters regardless of the nature of the trace considered. The notions are defined hereafter:

- *Credibility* depends on the system ability to limit the erroneous positive information it provides. In other words, credibility is related to the reliability of the positive results provided by the system. Credibility is measured through the rate of type I errors (i.e. to consider true a hypothesis that is actually false).
- *Integrity* depends on the system ability to limit the erroneous negative information it provides. In other words, integrity is related to the completeness of the positive results provided by the system (or wholeness, entireness, referring to the Latin origin of the word *integrity*). Integrity is measured through the rate of type II errors (i.e. to consider false a hypothesis that is actually true).
- *Timeliness* is associated with the system ability to provide information that can be used by decision-makers in a timely fashion. Time is critical when analysing criminal activity. Ideally, in order to be useful, the analysis response should be compatible with the rapid evolution of the phenomenon of interest [10]. Indeed, relevant information obtained at the wrong time would not only be useless but might be detrimental to the efficiency of further actions [11].
- *Flexibility* is the ability of the system to adapt to and account for the different contexts in which forensic intelligence may be applied [12]. The crime environment is dynamic and evolves rapidly. As a consequence, there is no universal system configuration that is adequate in every situation and flexibility is a key parameter of any evaluation system.

Flexibility and timeliness should both be maximised to provide actionable intelligence. In contrast, credibility and integrity cannot be maximised simultaneously since they evolve in opposite directions. For instance, a system that achieves high credibility but low integrity provides truthful but incomplete results, while a system that achieves low credibility and high integrity provides comprehensive but unreliable results. When considering the ability of the system to detect links among forensic case data, integrity is connected to the well-known risk of *linkage blindness* [13] while credibility is connected to the risk of detecting links that are actually absent. It is hypothesised that the credibility and integrity of the system are the driving factors for decision-making in performing any forensic intelligence task. Any selection of a metric or of an evaluation system, any queries in a database or any risk assessment are based on these criteria to balance the decision in order to fit the results to the expectations of operators. Finding the optimal trade-off between these criteria and the operational needs is a constant challenge for forensic scientist and intelligence analysts.

The following sections present the integration and role of the above criteria in regards to the different building blocks of the forensic intelligence process.

3. Comparison process: iterative selection of the 'best' metric

Once the features to be profiled are selected and extracted from specimens collected (see section 5.2), a process of comparing profiles and measuring their similarity must be selected [14]. Contrary to a common misconception, this choice, or decision point, is not only important when conceiving the system, but also arises each time the system is used. In fact, the choice of the best comparison process depends on the operator objectives and the kind of problem at stake, which may vary according to the context within which the forensic intelligence process is operated (see section 6).

Thus, the system must enable a flexible and dynamic selection of solutions to compare profiles and measure their similarity.

Metrics are the generic solution for the comparison process irrespective of the nature of the trace (i.e. visual, physical, chemical or even digital). A metric is defined as "*a transformation that adds a new layer of information since it starts with entities (i.e. profiles) and concludes with a measurement describing the degree of relationship between entities (i.e. scores)*" [4]. They are used to compare and measure the (dis)similarity between specimens. They have the critical advantage of relying on explicit, transparent and verifiable rules. They can be used with both quantitative and qualitative data, as demonstrated in previous work [15, 16]. Furthermore, metrics can be seamlessly integrated with additional statistical methods commonly used to manage and process big datasets typically encountered in forensic intelligence (e.g. principal component analysis).

Metrics, such as the Pearson correlation, the cosine function or the Manhattan and Euclidean distances have proven their relevance across different fields of study, such as illicit drugs, counterfeit medicines or false identity documents [14-19]. Figure 1 represents the distribution of score frequencies (i.e. intra- and inter-variability distributions) using organic impurities found in MDMA specimens (i.e. quantitative and continuous data) on the left and the visual characteristics of false identity documents (i.e. qualitative and discrete data) on the right. The individual results were discussed more thoroughly in previous research [15, 16]. As observed in Figure 1, once the scores are computed using metrics either as a degree of proximity or a distance, they can be treated independently of the nature of the specific trace and the type of data. Indeed, similar results are obtained, confirming the transversal nature of the comparison process.

Insert Fig. 1

The resulting information in terms of intra- and inter-variability can be managed and assessed through a common process and the most fit-for-purpose metric can be rationally chosen. This was conducted in previous articles and can be found in [15, 16]. Detection error trade-off (DET) and receiver operating characteristics (ROC) curves as well as Tippett plots are recommended tools to assess and compare scientifically and numerically the performance of different metrics [20-24]. Tippett plot are used for empirical performance assessment, in particular to assess the likelihood ratio (LR) accuracy. Figure 2 represents an example of a Tippett plot that compares the performance of the squared cosine correlation metric for illicit drugs and false identity documents datasets.

Insert Fig. 2

These tools assist in identifying the most suitable metrics since they allow the measurement of type I and type II error rates [25] and both are indicative of the credibility and integrity of the process. At this point, it is important to mention that error rates associated with metrics have to be re-evaluated on a regular basis to ensure their adequacy to the evolution of crime patterns and crime markets. Timeliness is not really an issue regarding metrics since rapid computation facilitates the generation of almost immediate results, independent of the metric selected.

The iterative selection of the 'best' metric is a pivotal decision point within the forensic intelligence process since it affects the scores that will go through the evaluation system, which critically influences the ultimate suitability of intelligence products.

4. Evaluation system: proposal of an integrated framework

Scores resulting from the comparison process must be interpreted as a *link value* through a formalised evaluation system. Approaches based on a deterministic or a Bayesian framework are common solutions across forensic literature and practice [25]. They are often regarded as antagonistic and mutually exclusive. In opposition to that view, an evaluation framework combining both approaches is advocated here to meet the flexibility criteria mentioned in section 2. The commonality between these evaluation approaches is their ability to assess credibility and integrity criteria through the expression of type I, and type II error rates. Both approaches have been presented and practically applied in [15]. Therefore, they will only be briefly described in the present article.

Let's consider two profiles, *A* and *B*, and the two hypotheses '*the profiles A and B are linked'*; '*the profiles A and B are not linked'* which represent *H1* and *H2*, respectively. The notion of link may accept different meanings depending on the type of trace, its origins, and the level of generality and inference considered.

In the deterministic approach, a binary classification is postulated and a link is either present or absent. The score is compared against a defined threshold value that is chosen in accordance with what is considered acceptable by the operator and the organisation in a given operating context. In that approach, either *H1* or *H2* is true. False positive (type I error – reporting that a link is present when in fact it is absent) and false negative (type II error – reporting that a link is absent when in fact it is present) rates depend on the selected metric and threshold.

$$Score \geq Threshold \Rightarrow Link$$
$$Score < Threshold \Rightarrow No\ link$$

In the Bayesian evaluation framework, a likelihood ratio (LR) is calculated by dividing the probability of obtaining a particular score given the presence of a link by the probability of obtaining this particular score given the absence of a link.

$$LR = \frac{Pr(Score|Link)}{Pr(Score|No\ link)}$$

The resulting LR is associated with a *rate of misleading evaluation given the presence of a link* (type I error) and a *rate of misleading evaluation given the absence of a link* (type II error). More elaborated LR formulas have been proposed but they are more appropriate to a criminal trial context where error rates have to be strictly measured and minimised, e.g. [26-28]. In contrast, in an intelligence perspective, pragmatism is necessary and there is a higher tolerance of risk [29].

Neither of these approaches can be considered to be systematically better. Ideally, both approaches need datasets of known sources to be properly calibrated. However, most of the time such datasets are unavailable and the common or different origin of specimens has to be hypothesised. The authors believe that the probabilistic approach has a pivotal advantage due to the fact that LRs are context independent, which is not the case of a deterministic classification. Being context independent is a major advantage in regards to the integration of *link values* in a formal memory. Indeed, *link values* expressed as LRs can be stored, searched and interpreted again in any new context. Furthermore, they can be quantitatively compared across different traces and combined with any other sort of forensic or alternative data. Figure 3 illustrates the different steps between the acquisition of profiles and decision-making according to the deterministic and Bayesian evaluation framework.

Insert Fig. 3

Expressing *link values* as a LR clarifies the limit between forensic science results and contextual information (prior probabilities and alternative intelligence) which are in principle within the competence of different professionals. Comparatively, the deterministic approach requires a mixing of both forensic and contextual information in order to define the decision threshold [25] but is more straightforward and pragmatic in the sense that interpretation delivers a '*black or white*' result associated with false positive and false negative rates. Such a result can be easily handled and communicated, and is therefore appreciated when subsequent timely decisions have to be made. The disadvantage of the Bayesian approach is the fact that no decision is made, which prevents reasoning a step further in the process (e.g. to form groups and classes among linked profiles that may facilitate the organisation and exploration of information as well as the elicitation of working hypotheses). Although assessing uncertainty may be important from a purely scientific point of view, it raises many difficulties in terms of information processing and requires sophisticated information management techniques such as fuzzy sets for instance [30]. This can be a drawback when facing large and and/or complex datasets where analysis and subsequent communication necessitate simplifications.

Given that both approaches have advantages and disadvantages and that forensic intelligence may serve very different and evolving security tasks [12], it is advocated to implement them both within a dynamic evaluation system. Their combination enables reasoning towards the analysis step by considering in parallel several working hypotheses. The elicitation and management of several hypotheses is a powerful and flexible form of reasoning practiced pragmatically and intuitively in every day intelligence and investigative routine activity. However, it is hardly ever formalised, which hampers its implementation in computerised systems. In that regard, the expression of a framework that accounts for that form of reasoning is pivotal to fill this gap and bring together deterministic and Bayesian sub-models. The proposed framework is presented in Figure 4.

Insert Fig. 4

In the deep level, the relationships between entities (i.e. scores) are represented by link values, that is the presence/absence of a link (deterministic sub-model) and the likelihood of a link (Bayesian sub-model). On that basis, analysts may elicit in parallel different working hypotheses. In the working level, these link values are materialised or not, depending on each relevant hypothesis (*WHa*, *WHb*, *WHc*, *WHn*) formulated by operators in the forensic intelligence process (e.g. forensic scientists, intelligence analysts or investigators). In Figure 4, *WHa* considers that the link detected between grey and white entities according to the deterministic threshold is not significant. *WHb* considers the links detected by the deterministic approach but disregards the potential link between grey and black entities supported by a limited or moderate likelihood ratio. *WHc* considers a link between grey and black entities that was not detected according to the deterministic threshold that is a link based on a moderate or limited likelihood ratio or based on the indirect link between grey and black entities. Analysis can be initiated based on these working hypotheses. At this stage, other forensic science information or alternative information may be leveraged to assess the relevance of each hypothesis. Thus, the boundary between the evaluation and analysis steps is not that clear within the forensic intelligence process [4].

A simple and routine example can be used to illustrate this framework. A burglary was committed in a factory and a DNA specimen was collected at the scene. When checked against the national database, a match was obtained with a profile related to one case included in a series of house burglaries committed two years earlier. In the deep level, this match can be assessed as the presence of a link (deterministic sub-model) and as a very high likelihood ratio in favour of a link (Bayesian sub-model). In the working level, two hypotheses were drawn from the analysis: '*WHa : there is a link between both traces since they have a common source' ; 'WHb : there is no link between both traces since the result is due to a random match'*. Two scenarios may explain WHa : *'a serial burglar deposited both traces'* or *'a person who has legitimate access to both locations (house and factory) deposited both traces'.* Each working hypothesis guided different decisions and operations. Investigations following *WHa* showed that the owner of the house was also an employee of the factory.

Through feedback and analysis of results and effects associated with the parallel hypotheses, an iterative process of adjustment and refining of the metric and evaluation system parameters is conducted. This requires the ability of the system to access the underlined information. Consequently, besides integrating deep and working model-related information, i.e. *link values* from both interpretation approaches in the memory and working hypotheses thereon, we support integrating the underlying *profiles* and *similarity scores* as well. This basic information is helpful to perform the continuous and dynamic reassessment required by the perpetual evolution of crime problems, and parallel hypotheses thereon [31].

To support decision-making, such an integrated framework enables the rapid optimisation of error rates according to accepted risks of reasoning falsely under each of the hypotheses considered. These risks are defined in relation to specific contextual intelligence challenges, objectives and policy decisions. Operators and their organisations may manage and endorse uncertainty and risks differently [6]. In that regard, the role of the forensic scientist within their organisation and within the forensic intelligence framework may guide the way both evaluation approaches are combined [32].

5. Discussion of the other building blocks

The comparison process and evaluation system are only a small part of the whole forensic intelligence process. They raise many questions downstream and upstream. It is essential to place them in context and discuss each step or building block. Indeed, decision points and organisational issues related to the forensic intelligence process must also be part of the modelling effort. It is always difficult to address these issues as they are more complex and are organisational dependent. Furthermore, they depend upon the organisational risk appetite relating to a particular offense category (e.g. the risk tolerance would probably be different in a counter-terrorism context in comparison to the monitoring of volume crime). However, the formalisation and modelling efforts initiate the discussion and offer possible solutions.

5.1. Detection and deciding which traces to profile

Available data that are used in the comparison process rely on the traces detected and collected [4]. At this early stage, data may already be biased in some ways depending on what is favoured consciously or not by detection stakeholders (police, border guards, community or any organisation facing the crime problem at stake). Their efforts may be oriented towards a particular phenomenon, specific traces or type of cases. For instance, police may concentrate their resources during a certain period of time on the methylamphetamine (MA) market to disrupt it and, as a result, MA seizures would increase. Meanwhile, MDMA, cocaine, and heroin seizures could apparently decrease. This increase and respective decrease do not reflect the actual criminal activity but rather the activity of the police [33]. In regards to false identity documents, alerts and instructions diffused to agents in the field will influence their detections. The choice of specific instrumentation will also make them able to detect specific features at the expense of others (e.g. ultraviolet light sources would assist them in detecting low quality counterfeits but could distract them from detecting higher quality forgeries). While targeted and intelligence-led operations are undisputedly vital to the success of policing [3, 34], the risk of over-focusing operations on known patterns exists. Given this risk, the ability of the forensic intelligence process to catch the unexpected must be developed. Rigid rules must be avoided and a more randomised or intelligent '*out of the box*' approach must be adopted on a regular basis.

Once crime traces are detected, a choice has to be made to decide what data are going to be profiled or not. This decision has a dramatic influence on the perception of crime problems that will result from forensic intelligence. The seriousness of the case is frequently viewed as the major criterion to guide this decision, mainly due to political or legal reasons. However, it might be more than often a misconception. Indeed, once properly aggregated, less serious and less spectacular cases, considered as fragmented and weak signals, may provide a considerable basis for the understanding of repetitive, prolific and organised crimes. For instance, the decision not to profile small illicit drug seizures (e.g. less than 10 grams) could critically neglect potential information on the retail levels of drug rings, or hide criminal organisations that intentionally work with small batches to reduce the judicial risk. Similarly, deciding not to profile false travel documents used by illegal immigrants may neglect human smuggling organisations from crime analysis. Furthermore, this also raises the question of representativeness. It is important to ensure that an adequate number of specimens are analysed to ensure they are representative of the crime market (or at least what is known to be the crime market) [10].

In selecting the traces to profile, the authors argue that the (apparent) seriousness of the case is not the driving factor and that more scientific and rational reasons prevail. Besides the trace relevance [32] and its quality [4], the main guiding criterion for forensic intelligence is to find a balance between a systematic approach and the available resources. Indeed, the more comprehensive the approach, the more significant the intelligence products will be. However, at the same time, more resources will be required and the management of information will become more complex. Knowing these limiting factors, a balanced approach that processes from the general to particular is advocated, as introduced in [31]. At a first level (*surface level profiling*), all traces should ideally be systematically profiled using surface features (effortless to extract and manage). This helps to point out and select sub-problems which deserve greater attention through a tighter and more resource-intensive follow-up, based on more detailed features (*modus operandi level profiling*) or on a restricted number of cases (*series level profiling*). For instance, the use of physical profiling or one chemical profiling technique to compare illicit drug seizures could be used first (surface level profiling) to obtain quick links between cases as demonstrated in [16, 35, 36]. The use of the whole set of chemical profiling techniques could then be applied when sub-problems are identified to obtain more detailed features or when a link is identified using circumstantial information and a confirmation of this link is required using chemical profiling (i.e. in this case, the entire set of chemical profiling techniques would be applied to a restricted number of seizures).

5.2. Deciding how to profile selected traces

From conceptual and practical viewpoints, it is not possible to extract all features of a trace. In fact, modelling traces is inherent to the recognition and extraction of their features. A model being inevitably a simplification of real objects, a conscious selection of features has to be made [37]. The criteria presented in Table 1 can be used as a guide to select relevant features. Some criteria are feature-intrinsic, i.e. they depend on the feature itself and its origin, while others are feature-extrinsic, i.e. they depend on the observation/measurement methods. These decision criteria do not depend on the level of detail at which the forensic intelligence process is engaged (general or particular). However, it is important to mention that it is difficult to fulfil all these criteria in practice. Furthermore, these criteria are most often based on hypotheses (e.g. investigation of the intra- and inter-variability of the different features is only possible if the hypothesis of a common/different source is inferred and assumed). As a consequence, a balance has to be struck between intrinsic and extrinsic criteria.

Insert Table 1

The example of the printing technique of personal data in a counterfeit identity document can be provided to illustrate these criteria. The printing technique has a priori a rather low intra-variability since there is no reason for the forger to regularly change his printer (additional costs). Although not necessarily high, inter-variability is judged higher than intra-variability since different forgers may have different printers (inkjet, toner, thermo transfer, thermo sublimation, etc.). Complementarity is more difficult to evaluate. For instance, the choice of a particular technique to print personal data has no relation (i.e. weak or no dependence) with the way the forger will cut the document's edges or how he will imitate the watermark. However, this choice will influence (i.e. stronger dependence) the technique chosen to print the document background, as it is easier to use only one printer rather than two. The printing technique is interesting as it enables the determination of type of equipment that was used (printer type, possibly its make and model) which is directly related to the representativeness. The forger has to have access to such a device, which describes the *modus operandi*. Comparability is not a problem when comparing the printing technique only. However, it may be an issue when comparing more specific features (such as the printing mode, colours or printing defects) since inkjet and toner printers might not be comparable for instance. In this case, availability and completeness are not an issue, as forgers will have no choice but to print personal data on every counterfeit document. These criteria could become an issue when considering forged documents where only the photograph was modified. Concerning the extrinsic criteria, observing and determining the printing technique of personal data can be done with a stereomicroscope or even a simple magnifying glass. Such an examination is not destructive, quick and does not require expensive equipment or extensive training. Its resource cost is low and its accessibility very high since every laboratory and almost every border control point possesses the required equipment and knowledge to perform this examination. However, it is difficult to rate sensitivity, specificity and reliability since such an examination is highly dependent on the operator competence. For a qualified examiner, recognising printing techniques is not a tough challenge but it may be less obvious to a less qualified operator. The last criterion, namely adaptability, is, a priori, not an issue since the development of new printing techniques is relatively slow and can be easily managed through a technological follow-up. The arrival of a new printing technique on the market would simply necessitate adding it to the previous list.

At this stage, it is important to develop the representativeness criterion in more details as the choice of particular features has a direct impact on the links that will ultimately be detected and inferred [15, 31]. For instance, analysing the organic impurities of illicit drugs present in the specimen or the physical characteristics of the specimen will result in different types of links. Indeed, a link obtained using the organic impurities will indicate a common origin at the early stage of production whereas a link obtained using the physical characteristics might indicate a common tableting process, which could take place in another location and laboratory. The combination of several features within the profile is essential to reach the optimal trade-off between the above-listed criteria. Some of these are antagonistic, such as representativeness and resource requirement. Indeed, it is always tempting to multiply the features to profile (visual, physical and chemical) in order to increase representativeness, but this has in turn a direct effect on the amount of resources needed to perform the profiling task. As demonstrated in [16], in an intelligence perspective, there is no need to apply multiple analytical methods for the chemical profiling of MDMA specimens. It was shown that the information gap resulting from using only one technique was negligible while the time needed to obtain results was greatly reduced. Furthermore, when combining features, their complementarity or correlation is an issue that must be investigated. Finally, finding the optimal trade-off between the intrinsic and extrinsic criteria is not a straightforward task and requires continuous research to remain adequate in regards to the evolution of crime patterns and markets.

5.3. Decisions in regards to the integration of information: organisation of the memory and fusion of alternative information

The integration of the different levels of forensic information presented in section 5.1 is difficult as is the management of parallel hypotheses discussed in section 4. In order to overcome these issues, the architecture of the memory must be flexible and dynamic and allow the operators to continuously '*zoom in - zoom out*' across information levels and navigate among hypotheses. If the memory is not flexible, it would run the risk of being obsolete, unable to support timely decisions and unable to adapt to the evolution of crime problems [4]. Besides its dynamic nature, the memory should always record the basic information that underlies *links* and *link values* (e.g. the profiles and the scores) in order to enable a backtracking capacity, as argued in section 4. This enables to re-build previous state of knowledge in light of new information at any time.

Recognising that forensic information is insufficient to understand criminal activity and that the aim of the memory is to gather information from different sources to obtain a holistic and shared view, should alternative information be integrated in the memory? If yes, how and to what extent should they be integrated?

These questions are difficult to answer as they require approaches to manage data from different sources, formats and values. While we believe that forensic science results are best used in combination with alternative information, we do not recommend building a unique database that centralises and captures all kind of data. Systematisation and computerisation of information play a critical role in the organisation and implementation of the memory. Indeed, it would be inconceivable to store and analyse numerous, diverse and complex information without the help of computers. However, it has to be complemented by the human capability of drawing inferences and thinking critically. The memory should be shared among the different units of the organisation [38], which fosters a collaborative approach to address crime problems. The memory and its architecture have to be thoroughly considered in regards to the organisation objectives and role. However, administrative issues should not be the main focus. Indeed, the memory should mainly be conceived and implemented in order to be able to answer the questions raised in regards to the crime problem at stake. Apart from the forensic scientist, different people, such as crime/intelligence analysts and decision makers will be involved in the decision concerning the architecture of the memory and the way information should be fused. Proper integration is a real challenge as people running the system may be disconnected from the collection, investigation and intelligence functions due to organisational, operational or political barriers.

6. Discussion: challenges and risks associated with forensic intelligence

Forensic intelligence suffers from limitations that can be concretely measured against two related risks that need to be mitigated and balanced: type I error and type II errors. To handle these limitations, the people in charge of generating forensic intelligence and those utilising it have to draw the right balance between two competing factors introduced above: credibility – limiting erroneous positive information – and integrity – limiting erroneous negative information. The notion of *expected utility* known to decision theory [39] could be a useful tool to approach the management of these risks in a transparent and objective manner. However, a completely formal approach should be avoided since intelligence tasks require pragmatism and flexibility, as presented in section 2. Besides technical limitations due to features measurement and the comparison metric used, finding the right compromise depends on the understanding of the environment, the context of the decision to be made, potential pitfalls of an erroneous decision, as well as on priorities and resources of the decision-maker and his organisation. These parameters affect the preference (or limited rejection) of type I versus type II errors and should guide the mindset of the decision-makers in regards to the kind of problem they are facing.

For instance, when the *problem to prove* is at stake (typical of an evaluative mindset [40]), the risk of asserting something that is absent (type I error) should be minimised since a higher expected utility is generally placed on certainty (i.e. credibility) than on integrity (i.e. completeness) of information. Accepting a false assertion is viewed in such contexts as the main issue. Indeed, in court, a type I error could lead to a false conviction (miscarriage of justice), an outcome that has to be avoided at all costs.

Conversely, when the *problem to find* is at stake (typical of an investigative [40] and intelligence mindset [10]), the risk of failing to assert something that is present or true (type II error) should be minimised since a higher expected utility is generally placed on having something to start to work with (a hypothesis to check, a lead to follow, a list of candidates resulting from a database search to process, a suspect to identify, arrest or interview) than on the reliability of information. Integrity is viewed as more important than credibility [41]. Indeed, the risk of missing what has to be found (i.e. suffering from an incomplete perception or linkage blindness) is perceived as the main issue in such contexts. For instance, when dismantling a forgery factory, investigators desire to check the profile of false documents found at the factory against the database in order to uncover previous cases that may be linked to the production of the factory, thus assisting in establishing the dimension of the production and the seriousness of the forger's offence (i.e. providing false documents to terrorist organisations or to underage alcohol drinkers). In this situation, type II errors would desirably be avoided (to avoid linkage blindness) and type I errors be accepted since the database search would only point to cases that should be examined more deeply. Indeed, the list of documents potentially connected that was highlighted through the database search may contain false positives. However, these false positives will be evaluated further and refined in the light of alternative information. A similar procedure is applied in DNA or fingerprint database searches where one expects the system to return a "hit" result if the donor is indeed recorded in the database (the opposite would comprise integrity). It is expected and accepted that the system returns a list of candidates which most probably contains some false positives. In such cases, the focus is generally placed on integrity rather that credibility, even if an adequate balance must be struck depending on the context of any database search.

The risks pertaining to different investigative and policing decisions may vary greatly and therefore requires a flexible and dynamic system. For instance, the decision to execute a search warrant (an invasive, overt and irrevocable measure) is associated with different risks than the decision of including a new case in a series at a preliminary stage of a criminal investigation. Both decisions might be made based on the analysis of the same set of data, but cannot afford the same error rates due to very different consequences. In tactical contexts, if the aim is to utilise resources efficiently and focus efforts by limiting results to the most concrete hypotheses/leads or, conversely, if the aim is to open and follow every possible lead, then the first option would need to minimise type I error (i.e. maximise credibility) whereas type II errors should be minimised in the second option (i.e. maximise integrity). In regards to strategic intelligence, a fine trade-off between error rates is not as important and could rather be solved by choosing a consensus known as the equivalent error rates (EER). There is indeed no need to finely balance error rates when considering a problem at a general level or over the long term, except when considering extreme situations in order to assess 'worst case' and 'best case' scenarios. However, credibility and integrity of the system should be explicitly stated since forensic intelligence can support critical strategic decisions (e.g. deciding if a particular crime problem is considered as a priority threat). For instance, if the profiling of illicit drugs or false identity documents reveals a high percentage of linked seizures, this might indicate that the crime market is highly structured and facilitated by organised crime. As a consequence, more resources may be allocated to address these crime problems.

It is important to underline at this point that decisions regarding the *problem to find* should not systematically be associated with a need for high integrity. Similarly, decisions related to the *problem to prove* do not constantly require maximising credibility. Such a dichotomous view would be simple but is inappropriate. Indeed, selecting the right trade-off relies not exclusively on the problem at stake and how it is perceived, but also on available resource, on the organisation strategy, on what solution is permitted or not according to the legal framework, and on the context in which the decision has to be reached. For instance, the problem *to prove* may be understood very differently when addressed during the first steps of an investigation (prove someone's likelihood to be a member of an organised crime group in order to wiretap them) or in court (prove their guilt). Therefore, decision issues in investigative and policing contexts must be flexible enough and take into consideration exceptions and special cases. Forensic scientists cannot be simply withdrawn in the laboratory as it is currently often the case, but have to collaborate with other stakeholders. The objectives of the forensic intelligence process and the context in which it will be used should be clear to the forensic scientists in order to tailor and fine-tune their contribution.

If a forensic intelligence unit exists in an organisation, its place and role should be well defined [6]. Its role should not be limited to the rather technical aspects of the forensic intelligence process. It must integrate the implementation and connection of all the building blocks described here and in [4]. It is, therefore, important to identify the different interactions with the other stakeholders in the organisation. The authors believe that collaboration is one of the most important elements of the process. As mentioned by Rossy and Ribaux [42] and Gray [43], "*collaboration is a process through which parties who see different aspects of a problem can constructively explore their differences and search for solutions that go beyond their own limited vision of what is possible*". The implementation of a forensic intelligence process is only possible if the different stakeholders have direct access to the different functional areas of the organisation.

7. Conclusions

Currently, different pieces of information belong to separate forensic science disciplines with little consideration given to their relation with the broader information management of the police and security [10]. Indeed, the current situation tends to restrain forensic scientists within their specialisation and reinforces the concept of centralised laboratories distant from and with no direct connections to police organisations [42]. The over-specialisation and the gap between forensic science and policing can be considered as obstacles in using the full potential of information conveyed by forensic case data through the development of a forensic intelligence framework. However, as demonstrated in [2] and here, forensic scientists should actively participate in the development of forensic intelligence models and also be responsible, in collaboration with other stakeholders, to define decision points that are crucial for a successful implementation of these models.

The modelling and generalisation efforts conducted in [4] and continued here offer a baseline to address the many issues, challenges and decision points raised by the conception and operation of an efficient forensic intelligence function. These formalisation efforts are pivotal since they contribute to:

- Structure and assist further developments by focusing attention on specific building blocks as well as on their common functioning. Formalisation provides modularity and reusability of the components and solutions, thus minimising the need to reinvent the wheel;
- Facilitate the implementation of forensic intelligence processes in education and training programs, in practice and in computerised systems;
- Monitor how information is processed, provide transparency and prevent what can be called risky *black box effects*;
- Offer tools to test, compare and evaluate scientifically the operation of any forensic intelligence process using explicit critical success factors;
- Bring together qualitative and quantitative approaches as well as Bayesian and non-Bayesian evaluation systems, showing that there is no need (and no use) to "choose a side". This development is necessary to widen and further define the role and contribution of forensic science in policing and security while maintaining its coherence as a discipline.

Forensic intelligence takes advantage of the wealth of information that result from the trace. Considered as the remnant of a punctual source, activity or event, traces are one of the most tangible and exploitable effects/results of crime phenomena [29, 44]. The scientific processing of traces as well as their proper fusion with alternative information in a collaborative approach offers a unique opportunity to understand criminal activities, in particular when they are repetitive, prolific and organised. Illicit drugs and false identity documents trafficking are only illustrations among others. A general forensic intelligence framework was presented. Whatever the type of trace considered, the successful implementation, integration and development of such a framework is complicated since many people (with often different objectives and interests) are involved at different levels of the process. We argue that forensic scientists have an important role to play in the implementation of such a framework. Furthermore, they should realise their key role in that endeavour and engage themselves in the forensic intelligence debate. Although more developments are required, the formalisation efforts presented in [4] and here initiate the discussion and offer possible solutions for the implementation of forensic intelligence on a routine basis. These developments are essential as they will ensure the progress of forensic science as a whole.

References

[1] O. Ribaux, A. Baylon, E. Lock, O. Delémont, C. Roux, C. Zingg, P. Margot, Intelligence-led crime scene processing. Part II: Intelligence and crime scene examination, Forensic Science International 199 (1-3) (2010) 63-71.

[2] O. Ribaux, P. Margot, La trace comme vecteur d'information au service du renseignement, in: M. Cusson, B. Dupont, and F. Lemieux (Eds.), Traité de sécurité intérieure, Presses polytechniques et universitaires romandes, Lausanne, 2008, pp. 300-321.

[3] J.H. Ratcliffe (Ed.), Intelligence-led policing, Willan Publishing, Cullompton, 2008.

[4] M. Morelato, S. Baechler, O. Ribaux, A. Beavis, M. Tahtouh, P. Kirkbride, C. Roux, P. Margot, Forensic intelligence framework—Part I: Induction of a transversal model by comparing illicit drugs and false identity documents monitoring, Forensic Science International 236 (2014) 181-190.

[5] O. Ribaux, P. Margot, R. Julian, S.F. Kelty, Forensic Intelligence, in: J.E. Siegel and P.J. Saukko (Eds.), Encyclopedia of Forensic Sciences, Academic Press, Waltham, 2013, pp. 298-302.

[6] P. Aepli, O. Ribaux, E. Summerfield, Decision Making in Policing, EPFL Press, Lausanne, 2011.

[7] J. Ratcliffe, The structure of strategic thinking, in: J. Ratcliffe (Ed.), Strategic thinking in criminal intelligence, The Federation Press, 2009, pp. 1-12.

[8] J. Grieve, Developments in UK criminal intelligence in: J. Ratcliffe (Ed.), Strategic thinking in criminal intelligence The Federation Press, 2009, pp. 28-46.

[9] O. Higgins, The theory and practice of intelligence collection, in: J. Ratcliffe (Ed.), Strategic thinking in criminal intelligence The Federation Press, 2009, pp. 85-107.

[10] M. Morelato, A. Beavis, M. Tahtouh, O. Ribaux, P. Kirkbride, C. Roux, The use of forensic case data in intelligence-led policing: The example of drug profiling, Forensic Science International 226 (1-3) (2013) 1-9.

[11] C. Zingg, The analysis of ecstasy tablets in a forensic drug intelligence perpective, PhD, Université de Lausanne, Faculté de Droit Ecole des Sciences Criminelles, Lausanne 183, 2005.

[12] Z. Geradts, P. Sommer, O. Ribaux, G. Edelman, G. Jacobusse, T. Gloe, M. Kirchner, S. Ioset, E. de Vries, F. Coudert, Forensic profiling, D6.7c: forensic profiling, Z. Geradts and P. Sommer (Eds.), Future of Identity in the Information Society, 2008.

[13] S.A. Egger, Working definition of serial murder and the reduction of linkage blindness, Journal of Police Science and Administration 12 (3) (1984) 348-357.

[14] O. Guéniat, P. Esseiva, Le profilage de l'héroïne et de la cocaïne. Une méthodologie moderne de lutte contre le trafic illicite, ed. P. Margot, Presses polytechniques et universitaires romandes, Lausanne, 2005.

[15] S. Baechler, V. Terrasse, J.-P. Pujol, T. Fritz, O. Ribaux, P. Margot, The systematic profiling of false identity documents: method validation and performance evaluation using seizures known to originate from common and different source, Forensic Science International 232 (1-3) (2013) 180-190.

[16] M. Morelato, A. Beavis, M. Tahtouh, O. Ribaux, P. Kirkbride, C. Roux, The use of organic and inorganic impurities found in MDMA police seizures in a drug intelligence perspective, Science & Justice 54 (1) (2014) 32-41.

[17] F. Been, Y. Roggo, K. Degardin, P. Esseiva, P. Margot, Profiling of counterfeit medicines by vibrational spectroscopy, Forensic Science International 211 (1–3) (2011) 83-100.

[18] P. Esseiva, L. Dujourdy, F. Anglada, F. Taroni, P. Margot, A methodology for illicit heroin seizures comparison in a drug intelligence perspective using large databases, Forensic Science International 132 (2) (2003) 139-152.

[19] R. Marquis, C. Weyermann, C. Delaporte, P. Esseiva, L. Aalberg, F. Besacier, J.S. Bozenko Jr, R. Dahlenburg, C. Kopper, F. Zrcek, Drug intelligence based on MDMA tablets data: 2. Physical characteristics profiling, Forensic Science International 178 (1) (2008) 34-39.

[20] P. Gill, J. Curran, C. Neumann, A. Kirkham, T. Clayton, J. Whitaker, J. Lambert, Interpretation of complex DNA profiles using empirical models and a method to measure their robustness, Forensic Science International: Genetics 2 (2) (2008) 91-103.

[21] A.F. Martin, G.R. Doddington, T. Kamm, M. Ordowski, M.A. Przybocki, The DET curve in assessment of detection task performance, in *Proc. Eurospeech*, 1997.

[22] C. Neumann, C. Champod, R. Puch-Solis, N. Egli, A. Anthonioz, A. Bromage-Griffiths, Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiæ, Journal of Forensic Sciences 52 (1) (2007) 54-64.

[23] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters 27 (8) (2006) 861-874.

[24] J.R. Beck, E.K. Shultz, The use of operating characteristics (ROC) curves in test performance evaluation, Archives of pathology & laboratory medecine 110 (1) (1986) 13-20.

[25] C. Champod, D. Meuwly, The inference of identity in forensic speaker recognition, Speech Communication 31 (2–3) (2000) 193-203.

[26] G.S. Morrison, Measuring the validity and reliability of forensic likelihood-ratio systems, Science & Justice 51 (3) (2011) 91-98.

[27] G. Pierrini, S. Doyle, C. Champod, F. Taroni, D. Wakelin, C. Lock, Evaluation of preliminary isotopic analysis (13C and 15N) of explosives. A likelihood ratio approach to assess the links between semtex samples, Forensic Science International 167 (1) (2007) 43-48.

[28] G.S. Morrison, Distinguishing between forensic science and forensic pseudoscience: testing of validity and reliability, and approaches to forensic voice comparison, Science & Justice 54 (3) (2014) 245-256.

[29] O. Ribaux, Police scientifique, le renseignement par la trace, Presses polytechniques et universitaires romandes, Lausanne, 2014.

[30] K. Stoffel, D. Han, P. Cotofre, Fuzzy methods for forensic data analysis. in International Conference SoCPaR 2010, 2010.

[31] S. Baechler, O. Ribaux, P. Margot, 2012 student paper: toward a novel forensic intelligence model: systematic profiling of false identity documents, Forensic science policy & management 3 (2012) 70-84.

[32] D. Hazard, P. Margot, Forensic science culture, in: G. Bruinsma and D. Weisburd (Eds.), Encyclopedia of Criminology and Criminal Justice, Springer, New York, 2014, pp. 1782-1795.

[33] M. Ouellet, C. Morselli, Precursors and prices: structuring the Quebec synthetic drug market, Journal of Drug Issues 44 (1) (2014) 37-55.

[34] J. Ratcliffe (Ed.), Strategic thinking in criminal intelligence, 2nd ed, The Federation Press, 2009.

[35] M. Lopatka, M. Vallat, Surface granularity as a discriminating feature of illicit tablets, Forensic Science International 210 (1-3) (2011) 188-194.

[36] J. Camargo, P. Esseiva, F. González, J. Wist, L. Patiny, Monitoring of illicit pill distribution networks using an image collection exploration framework, Forensic Science International 223 (1–3) (2012) 298-305.

[37] J. Rumbaugh, M. Blaha, E. Premerlani, E. Frederick, W. Lorensen, Object-oriented modeling and design, Prentice-Hall International, Englewood Cliffs, NJ, 1991.

[38] M. Borry, Le neuromanagement des connaissances: les science cognitives appliquées au knowledge management L'Harmattan, Paris, 2014.

[39] D.W. North, A tutorial introduction to decision theory, IEEE transactions on systems science and cybernetics 4 (3) (1968) 200-210.

[40] S. Kind, Crime investigation and the criminal trial: a three chapter paradigm of evidence, Journal of the forensic science society 34 (1994) 155-164.

[41] P. Esseiva, L. Gaste, D. Alvarez, F. Anglada, Illicit drug profiling, reflection on statistical comparisons, Forensic Science International 207 (1-3) (2011) 27-34.

[42] Q. Rossy, O. Ribaux, A collaborative approach for incorporating forensic case data into crime investigation using criminal intelligence analysis and visualisation, Science & Justice 54 (2) (2014) 146-153.

[43] B. Gray, Collaborating: finding common ground for multiparty problems, Jossey-Bass Inc., San Francisco, 1989.

[44] P. Margot, Traçologie: la trace, vecteur fondamental de la police scientifique, Revue Internationale de Criminologie et de Police Technique et Scientifique (2014) 72-97.

Table 1
Decision criteria to select the features to profile

Intrinsic criteria *(relative to the features themselves and their origin)*	*Extrinsic criteria* *(relative to the observation and measurement methods)*
Low intra-variability: no or low variation among traces of a same origin	Non destructiveness: must not alter the trace integrity
	Sensitivity: ability to return a positive result when the feature is present
High inter-variability: significant variation among traces of different origins	Specificity: ability to return a negative result when the feature is absent
Complementarity: the features must be as independent as possible	Reproducibility and reliability: consistency of results when the extraction/acquisition operation is executed by different operators or equipment, at different times or locations
Representativeness: reflects the features and traits of the source(s)/activity(ies) at the origin of the trace (materials, equipment and method used) – as far as a person comes into question, privacy issues must be taken into account	Low resource requirement in terms of costs, time, equipment, knowledge and training since the extraction/acquisition operation will be repeated for each new trace
Comparability: features of a given trace must be comparable to the features of the others to evaluate their similarity/dissimilarity	Accessibility: equipment and knowledge available to stakeholders that are prone to operate the profiling task – reciprocally, traces accessible to stakeholders who possess the required equipment and knowledge to operate the profiling task
Availability and completeness: the feature is constantly and completely observable/measurable in the population of interest (all the traces considered)	Adaptability: ability to follow-up on the evolution and mutations of the observed/measured feature

Acknowledgments

The authors gratefully acknowledge the support of the Australian Federal Police and staff.

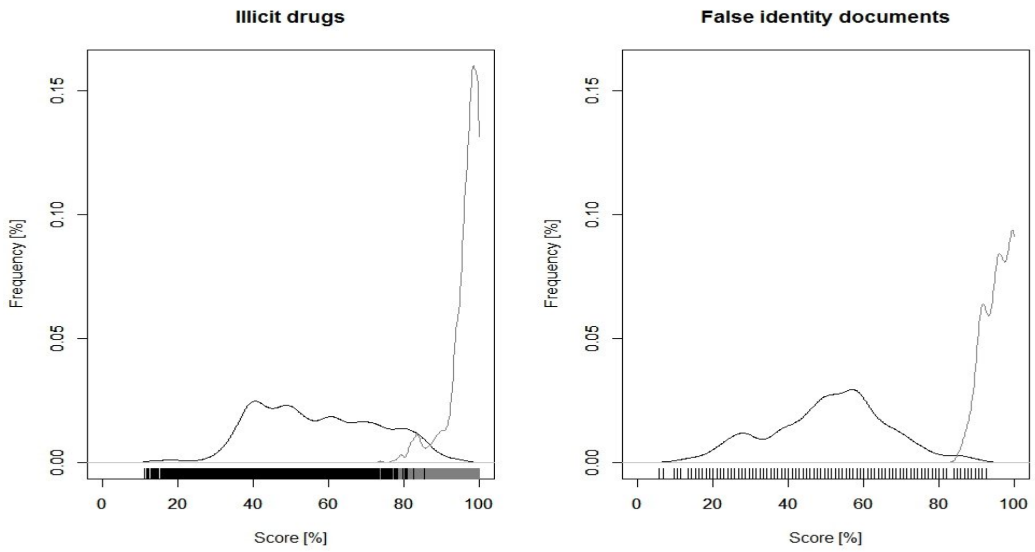

Fig. 1. Distributions of intra-variability (grey) and inter-variability (black) scores using the squared Cosine function and normalised on a 100% scale for MDMA seizures (left) and counterfeit identity cards (right).

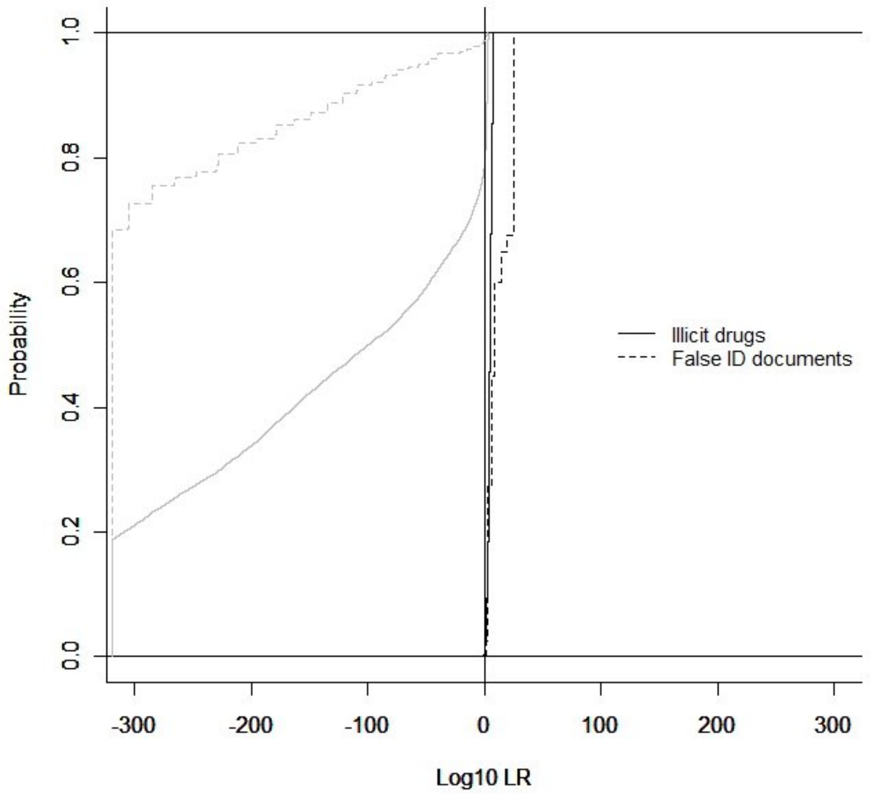

Fig. 2. Tippett plots presenting the Log_{10} likelihood ratios for MDMA seizures (solid line) and counterfeit identity cards (dashed line) according to the squared cosine correlation metric. Black curves (on the right) represent likelihood ratios when the hypothesis of a link is true, while grey curves (on the left) represent likelihood ratios when the hypothesis of the absence of a link is true

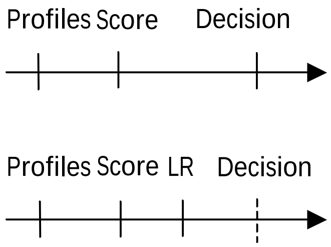

Fig. 3. Steps between the acquisition of profiles and decision-making according to the deterministic (top) and Bayesian (bottom) evaluation framework.

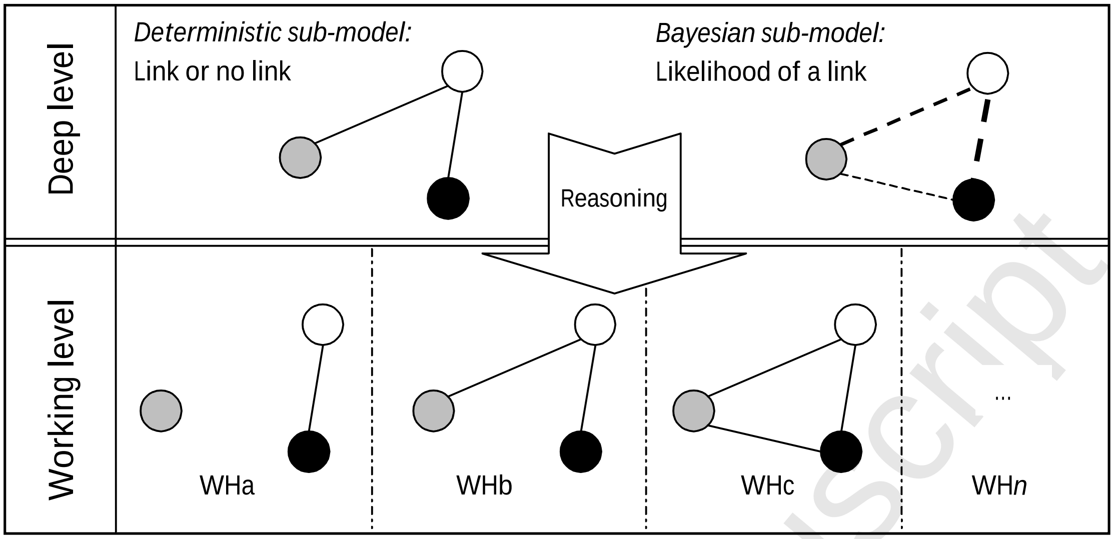

Fig. 4.Proposed framework regarding the elicitation and management of several parallel working hypotheses (WHa, WHb, WHc, WHn) based on link values evaluated through deterministic and Bayesian approaches. The grey, white and black dots represent profiles (i.e. forensic entities). Solid lines represent links considered as present, different levels of dashed lines represent different levels of likelihood of links.

Highlights

- Development of a transversal forensic intelligence process to guide the use of any forensic science case data in an intelligence perspective
- The comparison and evaluation processes that lay at the heart of the forensic intelligence framework are developed
- The issues and challenges raised by the conception and operation of an effective forensic intelligence function are addressed
- The article brings together qualitative and quantitative approaches as well as Bayesian and non-Bayesian evaluation systems
- These formalisation efforts are pivotal to ensure the progress of the discipline