

# **Activity level DNA evidence evaluation: on propositions addressing the actor or the activity**

Bas Kokshoorn<sup>a,#,\*</sup>, Bart J. Blankers<sup>a,#</sup>, Jacob C. De Zoete<sup>a,b</sup>, Charles. E.H. Berger<sup>a,c</sup>

<sup>a</sup> Netherlands Forensic Institute, P.O. Box 24044, NL-2490AA The Hague, The Netherlands.

<sup>b</sup> University of Amsterdam, Korteweg de Vries Instituut voor Wiskunde, P.O.Box 94248, NL-1090GE Amsterdam, The Netherlands.

<sup>c</sup> Leiden University, Institute for Criminal Law and Criminology, P.O. Box 9520, NL-2300RA Leiden, The Netherlands

\* corresponding author: [b.kokshoorn@nfi.minvenj.nl](mailto:b.kokshoorn@nfi.minvenj.nl)

# these authors contributed equally.

## **Highlights**

- Various proposition sets for evaluating DNA evidence at activity level are studied
- We demonstrate the relevance of distinguishing between a questioned actor or activity
- The sensitivity of the evidential value to hypotheses and model parameters is shown

## **Abstract**

More often than not, the source of DNA traces found at a crime scene is not disputed, but the activity or timing of events that resulted in their transfer is. As a consequence, practitioners are increasingly asked to assign a value to DNA evidence given propositions about activities provided by prosecution and defense counsel. Given that the dispute concerns the nature of the activity that took place or the identity of the actor that carried out the activity, several factors will determine how to formulate the propositions. Determining factors are (1) whether defense claims the crime never took place, (2) whether defense claims someone other than the accused (either an unknown individual or a known person) performed the criminal activity, and (3) whether it is claimed and disputed that the suspect performed an alternative, legitimate activity or has a relation to the victim, the object, or the scene of crime that implies a legitimate interaction. Addressing such propositions using Bayesian networks, we demonstrate the effects of the various proposition sets on the evaluation of the evidence.

**Keywords:** Hypotheses; activity level; weight of evidence; criminalistics; DNA; Bayesian network; evidence evaluation.

## 1 Introduction

In forensic science there are different levels at which questions can be asked and evidence can be interpreted. This hierarchy of propositions distinguishes between the sub-source level, the source level, the activity level, and the offense level [1,2]. At the sub-source level questions concern the donor of DNA. At source level inferences are made about the cellular origin of the DNA and the association with the DNA donor [3,4]. At the activity level issues relate to how or when a trace was deposited (e.g. through a criminal or legitimate activity, through direct or indirect transfer, etc.). At the offense level questions concern whether a crime has been committed and who is responsible. While the offense level questions are most often within the domain of the judiciary, forensic scientists can assist the courts in answering source level and activity level questions.

The levels in the hierarchy of propositions are strictly defined, but the issues that are addressed can be close together. For instance addressing a stabbing event at activity level may be very close to the offense level if there are no mitigating circumstances (e.g. self-defense). Similarly, associating semen found in intimate swabs of an under-aged girl to DNA of a male individual may be close to the offense level issues if there is no reasonable explanation for these results (e.g. contamination of the sample).

In the past three decades since the introduction of DNA profiling, some consensus has developed concerning the value of DNA evidence for questions at the source level. Probabilistic models have been developed that can assist in assigning a value to observed DNA profiles ([5-7], and others), even with complex mixtures of degraded DNA. These tools are now widely available to the forensic genetics community. But increasingly, questions about the donor(s) of the DNA in a trace are no longer relevant since the source is not disputed. The question then shifts to the activity by which the trace was transferred. As DNA profiling has become ever more sensitive and ever smaller traces can be successfully profiled, more modes of transfer have to be considered.

We cannot generally expect to be able to directly infer the mode of transfer from minute traces [8]. But if the prosecution and defense provide different propositions about the activities that took place, the probabilities of the results can be evaluated given those competing propositions. These results are not limited to the DNA profile, but can also include the cell type, amount, location, and distribution of the DNA. As long as the results are not equally likely under both propositions they provide evidence for one of the propositions.

During a trial, inferences about the evidence have to be made at the activity and offense level, since merely being the source of a DNA trace is not punishable by law. Legislation only penalizes criminal activities, and it is up to the trier of fact to finally make a decision at the offense level. In many cases the step from source level to activity level interpretation can be (or seem) trivial, when there is only one reasonable way the DNA could have been transferred (e.g. DNA on a cigarette butt, or blood on the edge of a glass shard from a window broken during a burglary). In such cases it could be safe to leave the activity level interpretation to the trier of fact.

In another large group of cases the step from the source level to the activity level is not straightforward and reasonable alternative activities or modes of transfer have been suggested by the defense. In such cases it is often the DNA expert that is in the best position to move from source to activity level. This requires the expert to consider probabilities of transfer, persistence, and recovery of DNA, and the circumstances of the case. The model for case assessment and interpretation [1] embedded within a Bayesian logical framework [9] provides the basis for these activity level interpretations.

This paper shares some insights about the formulation of propositions for activity level interpretation. These insights follow from experience with the application of the theoretical framework in casework, and discussions and sharing of ideas within both the forensic and legal community.

## **2 Formulating propositions**

For every evaluation of evidence the crucial first step is to identify the issue at stake. Based on this issue a minimum of two mutually exclusive propositions need to be formulated. These propositions need to address the issue while adhering to the underlying logic for the evaluation of evidence. The formulation of propositions at activity level has received some attention [10,11,2,12], and methods and guidelines for evaluating forensic evidence at activity level have been described [13-15]. Most recently, the European Network of Forensic Science Institutes (ENFSI) published a guideline for evaluative reporting in forensic science [16].

### **2.1 Guidelines for formulating propositions**

Ideally, the propositions in a case are entirely based on the positions taken by prosecution and defense. While the final propositions might look rather straightforward, the process of obtaining those propositions can take quite some effort. Both parties are usually able to

define the key issues to be considered and will formulate propositions that address these issues. But fulfilling the logical requirements for the propositions requires some knowledge about probabilistic reasoning in a Bayesian logical framework that is not always sufficiently available within the legal community. In these instances prosecution and defense can benefit from the assistance of the forensic expert when formulating the propositions.

The ENFSI guideline for evaluative reporting in forensic science states:

“The key issue(s) in the case will be established by:

- Considering all available, relevant information and, where necessary, requesting additional information
- Agreeing by discussing - when possible or necessary - with the relevant mandating authority or party (e.g., magistrate, prosecution or defence team)

On the basis of the case circumstances and the agreed key issue(s), competing propositions at a given level in the hierarchy are set.”

This echoes what the Association of Forensic Science Providers (AFSP) “Standards for the formulation of evaluative forensic science expert opinion” said:

“The key issue(s) in the case will be established and agreed by:

- Considering all available, relevant information and, where necessary, requesting additional information.
- Discussing with the relevant client e.g. police, CPS/PPS/COPFS and defence team.

On the basis of the case circumstances and the agreed key issue(s) identify:

- The relevant prosecution proposition.
- The relevant defence proposition.”

At the department of human biological traces of the Netherlands Forensic Institute (NFI), formulating the propositions is preferably done in close cooperation with prosecution and defense. For many cases standard sets of propositions can be used. In a small proportion of cases the propositions are discussed in a meeting with all parties involved (i.e. the examining judge, the prosecution, and the defense). The issue of what to do when the defense (or prosecution) chooses not to come forward with a proposition was discussed elsewhere [17]. Often an appropriate alternative proposition will follow from the positions of both parties.

Useful and relevant sets of propositions address one or more of the issues that are disputed in the case. The propositions must also be mutually exclusive. This means that if one proposition is true the other one must be false. It would not make sense to try to use evidence to distinguish between propositions that overlap. Propositions do not need to be exhaustive: they do not need to cover all possibilities [16], but if possible this is preferred. The search for the truth in a legal setting is not absolute but conditional on the positions taken by the parties and the information available. An exclusive alternative proposition could quickly be obtained by using the negation of the prosecution proposition (i.e. “The suspect strangled the victim” vs. “The suspect did *not* strangle the victim”), but this would violate another criterion for useful propositions. Propositions should be specific enough to allow assigning a probability for obtaining the evidence when either proposition is true.

Sometimes forensic scientists focus on what their analytical method can do rather than the relevant issues that are disputed in the case. This can result in activity level propositions that include words related to the analytical method or type of trace (i.e. “DNA of the accused was transferred through [...]”, which can distract from the relevant issues at activity level that exist irrespective of the traces chosen. A closely related problem arises when results of the analysis are featured in the propositions. In that case the proposition can become an explanation: a proposition which makes the probability of the resulting evidence one. The results of the examination have no place in the propositions [10]. Preferably the propositions are formulated before full examinations are carried out and their results are known, to prevent proposition formulation to be influenced by the results [1]. In practice however, (many of) the findings are usually known before activity level propositions are phrased, and often even before there is an accused and a defense.

We stress that this differs from the approach at (sub-)source level. At (sub-)source level, results of the biological traces and DNA analysis generally need to be known (number of contributors, individuals included in the mixture, cell type, etc.) to be able to formulate relevant propositions.

In the course of our casework we found that there are two identifiable and distinct situations in activity level interpretation. Either the actor is disputed while the activity is not, or the activity is disputed while the actor is not.

## 2.2 The disputed activity versus the disputed actor

The prosecution proposition is often well-defined: the accused has carried out a criminal activity. The prosecution - for example - states that (a) “the accused has stabbed the victim with a knife”, or that (b) “the accused has strangled the victim”. If the defense accepts that the criminal activity<sup>1</sup> took place, the criminal activity becomes a given for the interpretation of the evidence: “the victim has been stabbed”, or “the victim has been strangled”. In such cases the criminal activity is not disputed, but the identity of the actor (perpetrator) is.

Other examples of prosecution propositions are: (c) “the accused sexually assaulted the victim”, or (d) “the accused smothered the victim with a pillow”. If the defense disputes that the activity ever took place while proposing an alternative activity by the accused, then the activity rather than the actor is disputed.

In either situation, an alternative (non-criminal) activity or interaction between the suspect and the victim may be proposed. This aspect may also be disputed by the prosecution. This leads to three elements that are recurring themes in any set of propositions at activity level. These elements can be used in different combinations in either the prosecution or defense proposition, or be left out. They are (1) whether the defense claims a criminal activity did not take place, or (2) whether the defense claims someone else (either an unknown individual or a known person) performed the criminal activity, and (3) whether it is claimed and disputed that the accused performed an alternative, legitimate activity or has a relation to the victim, item of evidence, or the scene of crime that implies a legitimate interaction. Table 1 lists the combinations of propositions with these three key elements that are most commonly encountered in casework.

---

<sup>1</sup> We define ‘criminal activity’ as the activity that is alleged by the prosecution. The term ‘criminal activity’ is not intended to mean ‘offense’, and as such is not referring to whether an offense took place.

**Table 1** List of 6 sets of propositions based on the three key elements

<b>Dispute focused on activity (criminal / alternative activity)</b>		
1	H1	Accused did criminal activity, accused did not do alternative activity
	H2	Accused did alternative activity, criminal activity never happened
2	H3	Accused did criminal activity, accused did alternative activity
	H2	Accused did alternative activity, criminal activity never happened
<b>Dispute focused on actor (accused / unknown person)</b>		
3	H3	Accused did criminal activity, accused did alternative activity
	H4	Someone else did criminal activity, accused did alternative activity
4	H1	Accused did criminal activity, accused did not do alternative activity
	H4	Someone else did criminal activity, accused did alternative activity
5	H1	Accused did criminal activity, accused did not do alternative activity
	H5	Someone else did criminal activity, accused did not do alternative activity
6	H3	Accused did criminal activity, accused did alternative activity
	H5	Someone else did criminal activity, accused did not do alternative activity

Proposition sets 1 and 2 focus on a disputed activity. This means that in the context of the case it is unclear whether an alleged activity took place. In proposition sets 3 to 6 it is not disputed that the activity took place. These proposition sets instead focus on the person who performed the alleged activity.

Of these six sets of propositions, sets 5 and 6 are unlikely to be proposed in actual casework. These two sets do not provide a legitimate means by which DNA of the suspect may have been transferred to the location of interest. Effectively, only weight of evidence at source level and laboratory error and/or contamination will be relevant factors to consider under these two sets of propositions. We will therefore not address sets 5 and 6 any further.

Yet another proposition, H6, might be phrased as “Accused did not do alternative activity, criminal activity never happened”, leading to more proposition sets that could become relevant when no DNA traces are found. In casework practice, activity level questions are usually asked after the DNA results have become known. When the results are an absence of DNA this does usually not lead to a request to assess the results at activity



level. Though we note that ideally the activity level assessment should be done before the results have become known, we have chosen not to consider H6 for the purposes of this paper.

### 2.3 Proposition sets and Bayesian networks

In order to demonstrate the effect of the format of the propositions on the interpretation and weight of evidence we present four Bayesian networks that model the proposition sets 1 to 4 given in Table 1. All four sets of propositions can be derived from the following generic case description.

*The prosecution states that the accused performed the criminal activity. This is disputed by defense counsel who state that the accused did not perform the criminal activity. Defense also state that the suspect did in fact handle the item on which his DNA was found prior to the incident explaining the presence of his DNA on the item.*

Since the defense states nothing on whether criminal activity ever happened, and the prosecution states nothing on whether accused did the alternative activity, proposition sets 1 to 4 from Table 1 might apply. In Figures 1 to 4 four Bayesian networks are shown that might be used to assess this case depending on the applicable proposition set. The figures include the conditional probability tables for the network nodes. The third Bayesian network has been published previously (Evetts et al. 2002).

All four networks use the following nodes to evaluate the findings.

- **Hypotheses:** States: two competing propositions, defined by the specific positions taken in the case. The Bayesian network needs prior probabilities to operate. Since the prior probabilities of the propositions being true are in the domain of the judiciary, we refrain from addressing these. The value of 0.5 is only used to obtain the likelihood-ratio, which is in this instance the same as the posterior odds.
- Transfer from **Crime:** DNA was transferred by the criminal activity, persisted, and was recovered. States: yes / no. The probability of transfer, persistence and recovery of DNA through the criminal activity is designated  $P_{\text{crim}}$ .
- Transfer from **Alternative** activity<sup>2</sup>: DNA of the accused was transferred by the alternative activity, persisted, and was recovered. States: yes / no. The probability of

---

<sup>2</sup> Theoretically this alternative activity may be conceded, and the actor may be questioned. This is an exceptional situation, since the alternative –legitimate– activity is generally brought forward by

transfer, persistence and recovery of DNA through the alternative activity is designated  $P_{alt}$ .

- **Background:** This node provides the possibility of encountering background DNA. States: yes / no. Background DNA is defined as the presence of DNA of one or more unknown individuals within the context of the case that is unrelated to the criminal or alternative activity. It could have been transferred before, during (e.g. through the hands of the perpetrator), or after the incident. DNA is present and recovered from at least one unknown person. States: yes / no. The probability of recovery of background DNA is designated  $P_{bg}$ .
- **DNA result:** States:
  - Single source DNA profile with the accused not excluded;
  - Mixed DNA profile with the accused not excluded;
  - Other DNA profile (any profile, single source or mixed, which excludes the accused);
  - No DNA profile.

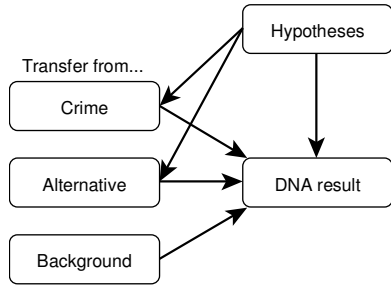
For all nodes the model only accounts for the presence or absence of DNA. We refrained, for this example, from including information on the quantity of DNA, or the quality or mixture ratio of the obtained DNA profiles.

### 2.3.1 Proposition set 1

The network for proposition set 1 in Figure 1 explicitly considers the alleged criminal activity versus the alternative activity. It is disputed by the defense that the criminal activity took place, while the prosecution disputes the legitimate activity proposed by defense. The propositions are H1: ‘Accused did criminal activity, accused did not do alternative activity’, vs. H2: ‘Accused did alternative activity, criminal activity never happened’. The probability tables for the nodes are also shown in Figure 1.

---

defense counsel. This activity may be disputed by prosecution, but another actor for this legitimate activity is generally not introduced.



Hypotheses	
H1: Accused did criminal activity, accused did not do alternative activity	0.5
H2: Accused did alternative activity, criminal activity never happened	0.5

Crime	H1	H2
Yes	$P_{\text{crim}}$	0
No	$1 - P_{\text{crim}}$	1

Alternative	H1	H2
Yes	0	$P_{\text{alt}}$
No	1	$1 - P_{\text{alt}}$

Background	
Yes	$P_{\text{bg}}$
No	$1 - P_{\text{bg}}$

Hypotheses		H1								H2							
Transfer	Crime	Y				N				Y				N			
	Alternative	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
	Background	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
DNA result	Single source	*	*	0	1	*	*	0	0	*	*	*	*	0	1	0	0
	Mixture	*	*	1	0	*	*	0	0	*	*	*	*	1	0	0	0
	Other	*	*	0	0	*	*	1	0	*	*	*	*	0	0	1	0
	No Profile	*	*	0	0	*	*	0	1	*	*	*	*	0	0	0	1

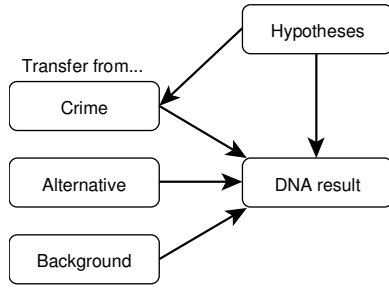
**Figure 1.** Bayesian Network and associated conditional probability tables (CPTs) for proposition set 1. In this example the alternative activity does not take place under H1, hence the probability for state ‘Yes’ for node ‘Alternative’ is zero under H1. Similarly, since the defense disputes that the criminal activity ever took place, the probability for transfer, persistence, and recovery of DNA via the criminal activity (state ‘Yes’ in node ‘Crime’) is zero under H2. (\*) Some entries in the CPTs correspond with ‘impossible’ situations. For example, the top left entry corresponds to the situation where *the accused did criminal activity, accused did not do alternative activity, there was transfer and persistence during the criminal activity and there was transfer and persistence due to an alternative explanation*. Due to the hypotheses, this cannot occur. Within a Bayesian network software package, the numbers in these columns are irrelevant.

This proposition set would be appropriate, for example, in a case with a disputed cause of death. The accused is suspected of smothering a victim with a pillow from the victim's bed. The accused claims the person died of natural causes instead, and states that he changed the sheets of the bed a few days before. This alternative activity is disputed by the prosecution.

The propositions for this particular case example would be  $H_p$ : 'The accused smothered the victim with the pillow and did not change the sheets a few days before', vs.  $H_d$ : 'The victim died of natural causes, the accused changed the sheets a few days before the incident'.

### **2.3.2 Proposition set 2**

The network for proposition set 2 (Figure 2) describes the situation where the occurrence of the criminal activity is disputed, but where the relation or alternative activity for the accused is undisputed. The propositions for this network are  $H_3$ : 'Accused did criminal activity, accused did alternative activity' vs.  $H_2$ : 'Accused did alternative activity, criminal activity never happened'. The probability tables for the nodes are also shown in Figure 2.



Hypotheses		
H3: Accused did criminal activity, accused did alternative activity		0.5
H2: Accused did alternative activity, criminal activity never happened		0.5

Crime	H3	H2
Yes	$P_{\text{crim}}$	0
No	$1 - P_{\text{crim}}$	1

Alternative	
Yes	$P_{\text{alt}}$
No	$1 - P_{\text{alt}}$

Background	
Yes	$P_{\text{bg}}$
No	$1 - P_{\text{bg}}$

Hypotheses		H3								H2							
Transfer	Crime	Y				N				Y				N			
	Alternative	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
	Background	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
DNA result	Single source	0	1	0	1	0	1	0	0	*	*	*	*	0	1	0	0
	Mixture	1	0	1	0	1	0	0	0	*	*	*	*	1	0	0	0
	Other	0	0	0	0	0	0	1	0	*	*	*	*	0	0	1	0
	No Profile	0	0	0	0	0	0	0	1	*	*	*	*	0	0	0	1

**Figure 2.** Bayesian network and node tables for proposition set 2. (\*) As in the CPT in Figure 1, some entries correspond with impossible situations and probabilities for those cells are irrelevant.

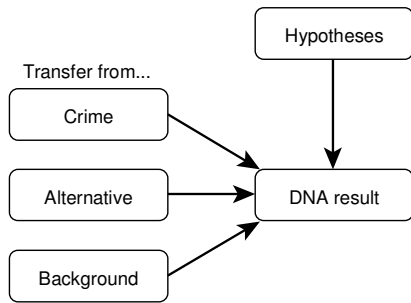
An example case for these propositions could be an alleged sexual assault after the complainant and the accused had dinner together. The legitimate activity (dinner) is not disputed, but the sexual intercourse as the criminal activity is. The propositions for this case example would be Hp: ‘The accused had sexual intercourse with the complainant, the accused had dinner with the complainant’, vs. Hd: ‘The accused did not have sexual intercourse with the complainant, the accused had dinner with the complainant’.

Alternatively, the undisputed legitimate activity (dinner) may be seen as part of the case circumstances, resulting in more concise propositions, Hp: ‘The accused had sexual

intercourse with the complainant’, vs. Hd: ‘The accused did not have sexual intercourse with the complainant’.

### **2.3.3 Proposition set 3**

The network for the third proposition set (Figure 3) uses propositions H3: ‘Accused did criminal activity, accused did alternative activity’, vs. H4: ‘Someone else did criminal activity, accused did alternative activity’, and considers the alternative activity (or relation between accused and item) under both propositions, while there is no dispute that the criminal activity occurred. Unless both persons performed the criminal activity together, these propositions are mutually exclusive. The probability tables for the nodes are shown in Figure 3.



Hypotheses	
H3: Accused did criminal activity, accused did alternative activity	0.5
H4: Someone else did criminal activity, accused did alternative activity	0.5

Crime	
Yes	$P_{\text{crim}}$
No	$1 - P_{\text{crim}}$

Alternative	
Yes	$P_{\text{alt}}$
No	$1 - P_{\text{alt}}$

Background	
Yes	$P_{\text{bg}}$
No	$1 - P_{\text{bg}}$

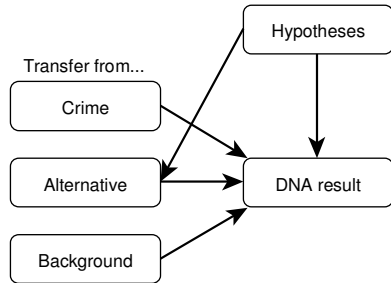
Hypotheses		H3								H4							
Transfer	Crime	Y				N				Y				N			
	Alternative	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
	Background	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
DNA result	Single source	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	0
	Mixture	1	0	1	0	1	0	0	0	1	1	0	0	1	0	0	0
	Other	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0
	No Profile	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1

Figure 3. Bayesian network and node tables for proposition set 3.

An example could be a strangulation case (the victim is deceased, and strangulation as the cause of death is not disputed) where the accused was living with the victim in the same house (a legitimate relation between accused and victim that is not disputed by either prosecution or defense). The associated propositions would be formulated as  $H_p$ : The accused strangled the victim, the accused and the victim were housemates, vs.  $H_d$ : someone else strangled the victim, the accused and the victim were housemates. Again, the undisputed part of the propositions might be moved to the conditioning case circumstances.

### 2.3.4 Proposition set 4

The network for proposition set 4 (Figure 4) has propositions H1: ‘Accused did criminal activity, accused did not do alternative activity’ vs. H4: ‘Someone else did criminal activity, accused did alternative activity’. The legitimate activity (or relation) of the accused is disputed by the prosecution, and so is the perpetrator of the crime, but the occurrence of the crime is undisputed. The probability tables for the nodes are shown in Figure 4.



Hypotheses	
H1: Accused did criminal activity, accused did not do alternative activity	0.5
H4: Someone else did criminal activity, accused did alternative activity	0.5

Crime	
Yes	$P_{crim}$
No	$1 - P_{crim}$

Alternative	H1	H4
Yes	0	$P_{alt}$
No	1	$1 - P_{alt}$

Background	
Yes	$P_{bg}$
No	$1 - P_{bg}$

Hypotheses		H1								H4								
Transfer	Crime	Y				N				Y				N				
	Alternative	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N			
	Background	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	
DNA result	Single source	*	*	0	1	*	*	0	0	0	0	0	0	0	0	1	0	0
	Mixture	*	*	1	0	*	*	0	0	1	1	0	0	1	0	0	0	0
	Other	*	*	0	0	*	*	1	0	0	0	1	1	0	0	1	0	0
	No Profile	*	*	0	0	*	*	0	1	0	0	0	0	0	0	0	0	1

**Figure 4.** Bayesian network and node tables for proposition set 4. The tables for the nodes ‘Alternative’ and ‘Background’ are as for proposition set 1. In this example the accused did not do the alternative activity under H1, hence the probability for state ‘Yes’ for node ‘Alternative’ is zero under H1 (as in proposition set 1).



An example case could be a stabbing incident where the accused states that someone else stabbed the victim, but that he and the victim had some drinks together that afternoon. The prosecution disputes that victim and accused did the alternative activity (having drinks). This results in the propositions Hp: ‘The accused stabbed the victim, the suspect did not have drinks with the victim’, vs. Hd: ‘An unknown individual stabbed the victim, the accused had drinks with the victim’.

## 2.4 Value of the evidence and sensitivity analyses based on a case example

The effect of the assumptions and the disputed elements for the four possible sets of propositions on the value of the evidence will be demonstrated using a single case example. Consider the following fictitious case.

*The prosecution states that a crime was committed by a perpetrator wearing a baseball cap. The cap was left at the scene. The accused states that he is in fact the owner of the cap, but it was stolen from him 2 weeks prior to the incident. The cap looks like it was thoroughly washed recently. The inside surface of the cap was sampled in its entirety by a single tape lift.*

For this case, values for the different probabilities were assigned for illustration purposes<sup>3</sup>:

- $P_{\text{crim}} = 0.9$
- $P_{\text{alt}} = 0.1$
- $P_{\text{bg}} = 0.5$

The likelihood ratio for this case example will, of course, depend on the DNA results. Four different outcomes are considered again: (1) a single source DNA profile that matches the profile of the accused, (2) a mixture profile that does not exclude a contribution by the accused, (3) other, which covers all situations in which a (single source or mixture) profile was observed but where the result excludes the accused as a contributor to the stain, and (4) no profile. Each network (corresponding to another set of propositions), each of which could be relevant to the generalized case example at hand, results in a different value for the likelihood ratio (LR). An overview is presented in Table 2.

---

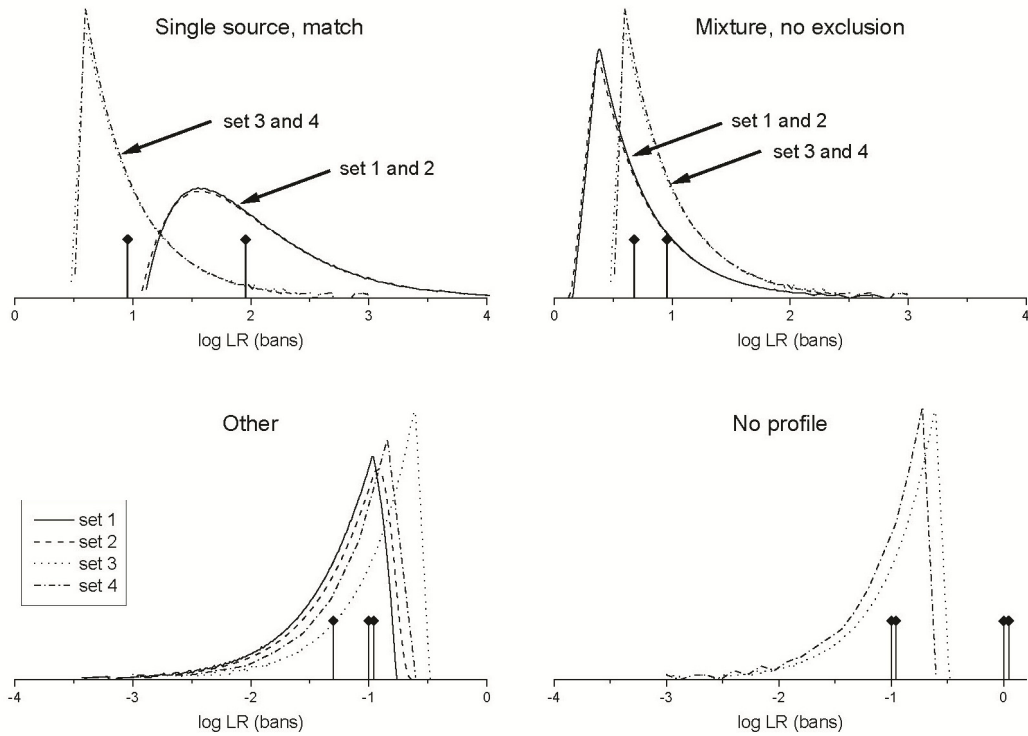
<sup>3</sup> These values are not based on experimental data. The values were chosen for illustration purposes. Data acquisition and its use in casework is outside the scope of the paper.

**Table 2.** Likelihood ratio values for the four situations and four profiling results, given the values  $P_{\text{crim}} = 0.9$ ,  $P_{\text{alt}} = 0.1$ , and  $P_{\text{bg}} = 0.5$ .

	Single source, match	Mixture, no exclusion	Other	No profile
Proposition set 1	9	9	0.11	0.11
Proposition set 2	9.1	9.1	0.1	0.1
Proposition set 3	91	4.79	0.05	1
Proposition set 4	90	4.74	0.06	1.11

The value of the likelihood ratio further depends on values assigned to the probabilities  $P_{\text{crim}} = 0.9$ ,  $P_{\text{alt}} = 0.1$ , and  $P_{\text{bg}} = 0.5$ . When such values are assigned based on very limited data we might wonder whether to carry out (more) experiments. That choice will also depend on the sensitivity of the resulting LR values to the assigned values for the probabilities. By performing a sensitivity analysis, the influence of varying the probabilities on the likelihood ratio can be examined. When, instead of a single value, the probabilities are assigned a distribution, a distribution of LRs will be obtained. So, instead of  $P_{\text{crim}} = 0.9$  we use a normal distribution  $P_{\text{crim}} \sim N(\mu_{\text{crim}}, \sigma_{\text{crim}}^2)$ , where  $\mu_{\text{crim}} = 0.9$  and  $\sigma_{\text{crim}} = 0.05$ . Similarly,  $P_{\text{alt}} \sim N(0.1, 0.05^2)$  and  $P_{\text{bg}} \sim N(0.5, 0.05^2)$ . Here,  $\sigma_{\text{crim}}$  represents the variation around the value 0.90. Since we know how to express the value of the likelihood ratio in terms of  $P_{\text{crim}}$ ,  $P_{\text{alt}}$ , and  $P_{\text{bg}}$  (see the Appendix) the distribution of LRs resulting from simultaneous variation of those transfer probabilities can be obtained<sup>4</sup>. The likelihood ratio distributions for the different outcomes of the DNA profile and the different networks can be found in Figure 5. These figures can aid the forensic practitioner in identifying the order of magnitude of the LR that characterizes the evidence. Especially in situations where the LR is expressed verbally, the sensitivity analysis can assist in assigning an evidential value on a verbal scale. Furthermore, if the assigned values for the probabilities  $P_{\text{crim}}$ ,  $P_{\text{alt}}$ , and  $P_{\text{bg}}$  are based on experiments, this analysis can suggest which type of additional experiments are most informative for the assignment of the likelihood ratio.

<sup>4</sup> In theory, we can express the value of the likelihood ratio in terms of the parameters of the normal distributions for the different events. However, this often (like in this scenario) results in an overly complex expression.



**Figure 5.** The likelihood ratio sensitivity to the described variation in the transfer probabilities, for the different states of the ‘DNA results’ node and the various proposition sets. Vertical lines denote the values shown in Table 2.

In the networks for proposition sets 1, 2 and 4 one or both of the nodes ‘Crime’ and ‘Alternative’ are dependent on the propositions. This means that the probabilities of transfer, persistence and recovery may depend on the propositions in these situations. This might for instance become relevant if the known accused under  $H_p$  would differ from the general population in shedding propensity (e.g. a skin condition). Another reason could be that persistence of material through the relation is (strongly) influenced by the criminal activity or vice versa. In the example used here, we assume that the probabilities of persistence are the same for both propositions.

### 3 Discussion and conclusion

The results of the analyses using the case example allow for some general and case specific observations. There are two points that we wish to address specifically.

First, if the criminal activity is disputed (proposition sets 1 and 2) the presence or absence of DNA of unknown(s) is irrelevant to the weight of the evidence, since the dispute

focuses on the presence or absence of DNA of the accused. However, if the criminal activity is conceded, the dispute focuses on the actor of that activity (proposition sets 3 and 4). Under these conditions the presence or absence of DNA of unknown(s) does influence the likelihood ratio.

Second, the difference between proposition sets 1 and 2 is the alternative activity that is questioned by the prosecution under set 1 and conceded under set 2. The value for the likelihood ratio calculated for each state of the node 'DNA result' for those proposition sets is nevertheless nearly the same or identical to those calculated for proposition set 2. This is probably due to the relative values chosen for  $P_{crim}$  and  $P_{alt}$ . Given that the probability for finding DNA of the accused on the cap is low under the alternative activity (0.1) compared to the criminal activity (0.9), the fact that a match is found practically makes the presence of DNA through the alternative activity irrelevant.

For the transfer probabilities chosen for this particular case example, we have shown that the value of the likelihood ratio may vary up to an order of magnitude depending on the propositions that are selected to assess the case. From the differences in the likelihood ratios for proposition sets 1 and 2 versus 3 and 4 shown in Table 2 and the likelihood ratio distributions shown in Figure 5 - which are based on the same generalized case example - it is clear that whether the dispute in the case concerns the activity or the actor can make a crucial difference in the evaluation of the evidence.

It should also be noted that - unless the prosecution disputes the alternative activity ever took place - the proposed alternative activity could also have taken place under the prosecution proposition. Depending on the circumstances of the case, it might be undisputed that e.g. the accused and the victim did have a meeting earlier that day, regardless of whether the accused stabbed the victim or not. If the criminal activity is a given, and the prosecution does not dispute the alternative activity taking place or the relation existing, then this can also be part of the agreed case circumstances rather than the propositions.

We used simplified Bayesian networks, focusing on presence or absence of DNA, with the exclusion of other potentially relevant information like amount of DNA, cell type, distribution of traces, etc. Complex models that take one or more of these factors into account have been published and are available for use in casework (e.g. [18-20]). While these complex networks take more information into account, the general principle that we demonstrated holds true.

In practice we notice that sometimes parts of the legitimate activity are conceded, while other elements are disputed. An example may be a meeting between accused and

victim prior to the incident. The defense counsel may suggest that the suspect and victim shook hands and kissed when meeting, while prosecution agrees that hands were shaken, but disputes the kiss. Our study shows how important it is to get to the core of the issue and find out which elements are disputed and which are not.

## **Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## **References**

- [1] Cook, R., Evett, I.W., Jackson, G., Jones, P.J., Lambert, J.A. (1998) A model for case assessment and interpretation. *Science and Justice* 38: 151-156.
- [2] Evett, I.W., Gill, P.D., Jackson, G., Whitaker, J., Champod, C. (2002) Interpreting Small Quantities of DNA: the Hierarchy of Propositions and the Use of Bayesian Networks. *Journal of Forensic Sciences* 47: 520-530.
- [3] de Zoete, J., Oosterman, W., Kokshoorn, B., & Sjerps, M. (2016) Cell type determination and association with the DNA donor. *Forensic Science International: Genetics* 25: 97-111.
- [4] Taylor, D., Abarno, D., Hicks, T., & Champod, C. (2016) Evaluating forensic biology results given source level propositions. *Forensic Science International: Genetics* 21: 54-67.
- [5] Haned, H., Slooten, K., Gill, P. (2012) Exploratory data analysis for the interpretation of low template DNA mixtures. *Forensic Science International: Genetics* 6: 762-774.
- [6] Taylor, D., Bright, J.-A., Buckleton, J. S. (2013) The interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics*. 7: 516-528.
- [7] Perlin, M.W., Legler, M.M., Spencer, C.E., Smith, J.L., Allan, W.P., Belrose, J.L., Duceman, B.W. (2011) Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences* 56: 1430-1447.
- [8] Meakin G.E., Jamieson, A. (2013). DNA transfer: review and implications for casework. *Forensic Science International Genetics* 7: 434-443.
- [9] Jackson, G., Jones, S., Booth, G., Champod, C., Evett, I.W. (2006) The nature of forensic science opinion—a possible framework to guide thinking and practice in investigation and in court proceedings. *Science and Justice* 46: 33-44.

- [10] Hicks, T., Biedermann, A., Koeijer, J.A. de, Taroni, F., Champod, C., Evett, I.W. (2015) The importance of distinguishing information from evidence/observations when formulating propositions. *Science and Justice* 55: 520-525.
- [11] Gill, P. (2014) *Misleading DNA evidence. Reasons for miscarriages of justice.* pp 1-173. Academic Press, Elsevier. London, UK.
- [12] Evett, I.W., Gill, P.D., Jackson, G., Lambert, J.A., McCrossan, S. (2000a) The impact of the principles of evidence interpretation on the structure and content of statements. *Science and Justice* 40: 233-239.
- [13] Association of Forensic Science Providers (2009). Standards for the formulation of evaluative forensic science expert opinion. *Science & Justice* 49: 161–164.
- [14] Kokshoorn, B., Aarts, L.H.J., Blaeij, T. de, Maaskant-Van Wijk, P.A., Blankers, B.J. (2014) Bewijskracht van onderzoek naar biologische sporen en DNA. Deel 3: Activiteitniveau. *Expertise & Recht* 2014-6: 213-219.
- [15] Jackson, G., Aitken, C., Roberts, P. (2015) *Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice.* 4. Case Assessment and Interpretation of Expert Evidence. pp 3-133. Royal Statistical Society. UK.
- [16] Aitken, C.C.G., Barrett, A., Berger, C.E.H., Biedermann, A., Champod, C., Hicks, T.N., Lucena-Molina, J., Lunt, L., McDermott, S., McKenna, L., Nordgaard, A., O'Donnell, G., Rasmusson, B., Sjerps, M.J., Taroni, F., Willis, S.M., Zadora, G. (2015) ENFSI guideline for evaluative reporting in forensic science, European Network of Forensic Science Institutes (ENFSI).
- [17] Evett, I.W., Jackson, G., Lambert, J.A. (2000b) More on the hierarchy of propositions: exploring the distinction between explanations and propositions. *Science and Justice* 40: 3-10.
- [18] Wieten, R., de Zoete, J.C., Blankers, B.J. & Kokshoorn, B. (2015) The interpretation of traces found on adhesive tapes. *Law, Probability & Risk* 14: 305-322.
- [19] Taylor, D. (2016) The evaluation of exclusionary DNA results: a discussion of issues in *R v. Drummond*. *Law, Probability & Risk* 15: 175-197.
- [20] Taylor, D., Biedermann, A., Samie, L., Pun, K., Hicks, T. & Champod, C. (2017) Helping to distinguish primary from secondary transfer events for trace DNA. *Forensic Science International: Genetics*. 28: 155-177.

## Appendix

These are the equations for the likelihood ratio in terms of  $P_{crim}$ ,  $P_{alt}$ , and  $P_{bg}$  for each of the four states of the node ‘DNA result’ in the Bayesian networks for proposition sets 1 to 4 as described in Section 2.3.

### Network 1

With  $E$  = single source profile, accused not excluded:

$$LR = \frac{P_{crim} | Hp (1 - P_{alt} | Hp)(1 - P_{bg})}{(1 - P_{crim} | Hd)P_{alt} | Hd (1 - P_{bg})} = \frac{P_{crim}}{P_{alt}}. \quad (1)$$

With  $E$  = mixture profile, accused not excluded:

$$LR = \frac{P_{crim} | Hp P_{bg} (1 - P_{alt} | Hp)}{(1 - P_{crim} | Hd)P_{alt} | Hd P_{bg}} = \frac{P_{crim}}{P_{alt}}. \quad (2)$$

With  $E$  = other profile, accused excluded:

$$LR = \frac{(1 - P_{crim} | Hp)(1 - P_{alt} | Hp)P_{bg}}{(1 - P_{crim} | Hd)(1 - P_{alt} | Hd)P_{bg}} = \frac{(1 - P_{crim})}{(1 - P_{alt})}. \quad (3)$$

With  $E$  = no profile:

$$LR = \frac{(1 - P_{crim} | Hp)(1 - P_{alt} | Hp)(1 - P_{bg})}{(1 - P_{crim} | Hd)(1 - P_{alt} | Hd)(1 - P_{bg})} = \frac{(1 - P_{crim})}{(1 - P_{alt})}. \quad (4)$$

### Network 2

With  $E$  = single source profile, accused not excluded:

$$LR = \frac{P_{crim} | Hp(1 - P_{bg}) + (1 - P_{crim} | Hp)P_{alt}(1 - P_{bg})}{(1 - P_{crim} | Hd)P_{alt}(1 - P_{bg})} = \frac{P_{crim} + (1 - P_{crim})P_{alt}}{P_{alt}}. \quad (5)$$

With  $E$  = mixture profile, accused not excluded:

$$LR = \frac{P_{crim} | Hp P_{bg} + (1 - P_{crim} | Hp)P_{alt} P_{bg}}{(1 - P_{crim} | Hd)P_{alt} P_{bg}} = \frac{P_{crim} + (1 - P_{crim})P_{alt}}{P_{alt}}. \quad (6)$$

With  $E$  = other profile, accused excluded:

$$LR = \frac{(1 - P_{crim} | Hp)(1 - P_{alt})P_{bg}}{(1 - P_{crim} | Hd)(1 - P_{alt})P_{bg}} = 1 - P_{crim}. \quad (7)$$

With  $E$  = no profile:

$$LR = \frac{(1 - P_{crim} | Hp)(1 - P_{alt})(1 - P_{bg})}{(1 - P_{crim} | Hd)(1 - P_{alt})(1 - P_{bg})} = 1 - P_{crim}. \quad (8)$$

### Network 3

With  $E$  = single source profile, accused not excluded:

$$LR = \frac{P_{crim}(1-P_{bg}) + (1-P_{crim})P_{alt}(1-P_{bg})}{(1-P_{crim})P_{alt}(1-P_{bg})} = 1 + \frac{P_{crim}}{P_{alt} - P_{alt} \cdot P_{crim}}. \quad (9)$$

With  $E$  = mixture profile, accused not excluded:

$$LR = \frac{P_{crim}P_{bg} + (1-P_{crim})P_{alt}P_{bg}}{P_{crim}P_{alt} + (1-P_{crim})P_{alt}P_{bg}}. \quad (10)$$

With  $E$  = other profile, accused excluded:

$$LR = \frac{(1-P_{crim})(1-P_{alt})P_{bg}}{P_{crim}(1-P_{alt}) + (1-P_{crim})(1-P_{alt})P_{bg}} = \frac{(1-P_{crim})P_{bg}}{P_{crim} + (1-P_{crim})P_{bg}}. \quad (11)$$

With  $E$  = no profile:

$$LR = \frac{(1-P_{crim})(1-P_{alt})(1-P_{bg})}{(1-P_{crim})(1-P_{alt})(1-P_{bg})} = 1. \quad (12)$$

### Network 4

With  $E$  = single source profile, accused not excluded:

$$LR = \frac{P_{crim}(1-P_{alt} | Hp)(1-P_{bg})}{(1-P_{crim})P_{alt} | Hd (1-P_{bg})} = \frac{P_{crim}}{(1-P_{crim})P_{alt}}. \quad (13)$$

With  $E$  = mixture profile, accused not excluded:

$$LR = \frac{P_{crim}(1-P_{alt} | Hp)P_{bg}}{P_{crim}P_{alt} | Hd + (1-P_{crim})P_{alt} | Hd)P_{bg}} = \frac{P_{crim}P_{bg}}{P_{crim}P_{alt} + (1-P_{crim})P_{alt}P_{bg}}. \quad (14)$$

With  $E$  = other profile, accused excluded:

$$LR = \frac{(1-P_{crim})(1-P_{alt} | Hp)P_{bg}}{P_{crim}(1-P_{alt} | Hd) + (1-P_{crim})(1-P_{alt} | Hd)P_{bg}} = \frac{(1-P_{crim})P_{bg}}{P_{crim}(1-P_{alt}) + (1-P_{crim})(1-P_{alt})P_{bg}}. \quad (15)$$

With  $E$  = no profile:

$$LR = \frac{(1-P_{crim})(1-P_{alt} | Hp)(1-P_{bg})}{(1-P_{crim})(1-P_{alt} | Hd)(1-P_{bg})} = \frac{1}{1-P_{alt}}. \quad (16)$$