

# Interpreting a problematic DNA profile

[gill@math.leidenuniv.nl](mailto:gill@math.leidenuniv.nl)

<http://www.math.leidenuniv.nl/~gill>

# Contents

- The crime, the evidence, and the trials
- Statistical paradigms
- Bayes' nets
- DNA profiles (genetics and biochemistry)
- Application to Tamara Wolvers case
- Conclusions

A classical painting of a woman, likely a personification of Justice or Truth, holding a scale of justice. She is depicted from the waist up, wearing a green and blue robe with a red sash. Her right arm is raised, holding the top of the scale, while her left hand rests on a dark, draped object. The background is dark and textured. The text "The crime, the evidence, and the trials" is overlaid in white, centered on the image.

The crime,  
the evidence,  
and the trials

# Tamara Wolvers murder case (Alphen on the Rijn)

- 2006: The murder
- 2008: Lower court: acquittal
- 2010: Appeal court: acquittal
- 2012: Supreme court: acquittal
- 2013: New law may allow reopening

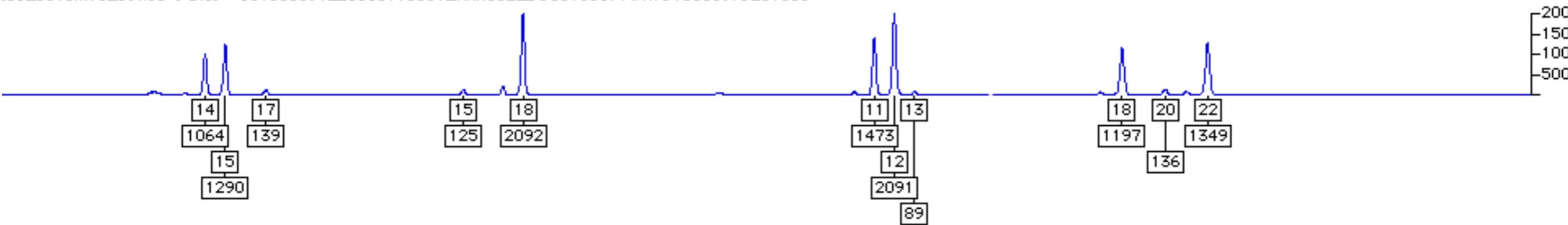
# Evidence

- Autosomal DNA traces
- Y-chromosome DNA
- Mitochondrial chromosome DNA
- The shoe-print
- The computer
- Opportunity

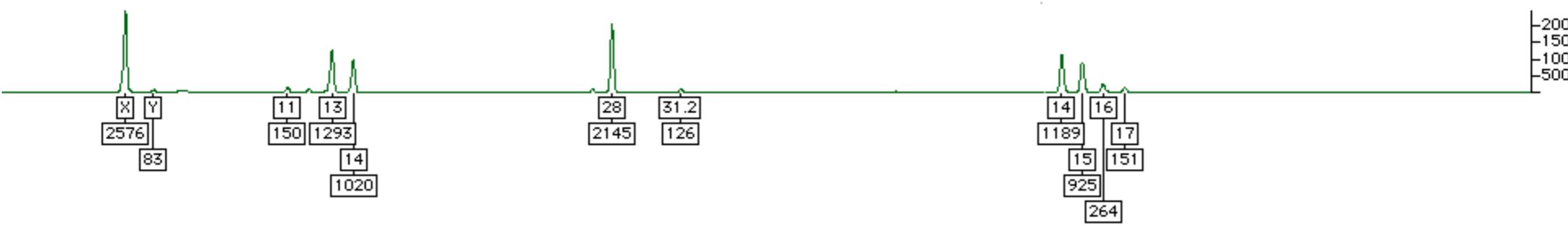
Peak: Scan 5520 Size 286.83 Height 68 Area 638 Category: D18S51:13

30 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380

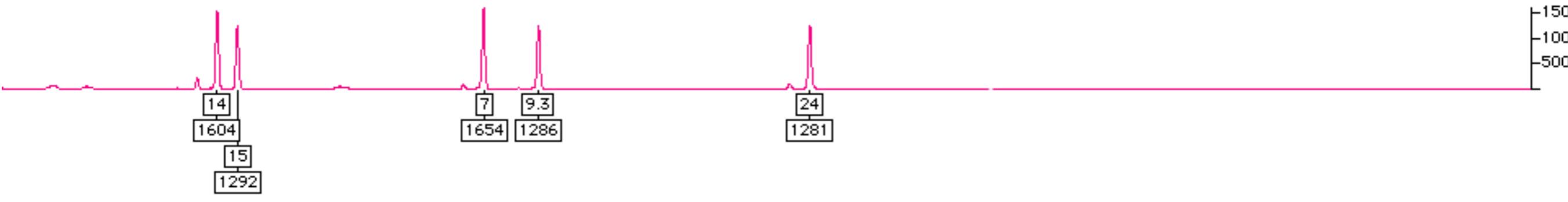
:03\_3015...15\_09.fsa 9 Blue 301583097\_20060713001\_AHH352\_2 D3S1358/VWA/D16S539/D2S1338



:03\_3015...15\_09.fsa 9 Green 301583097\_20060713001\_AHH352\_2 AMEL/D8S1179/D21S11/D18S51



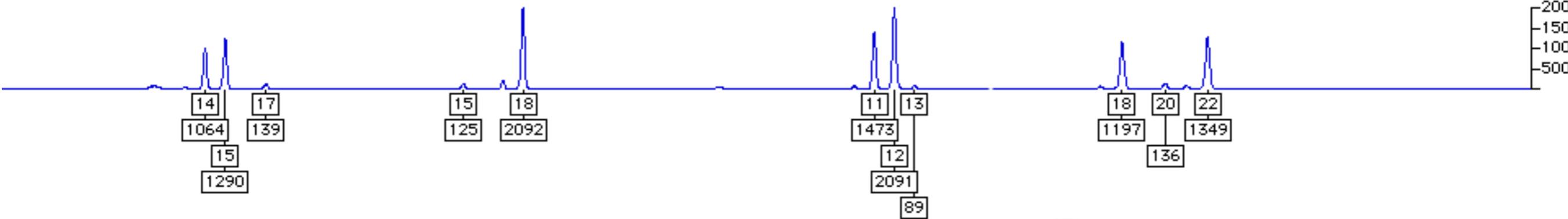
:03\_3015...15\_09.fsa 9 Yellow 301583097\_20060713001\_AHH352\_2 D19S433/TH01/FGA(-\*P06-971/019)



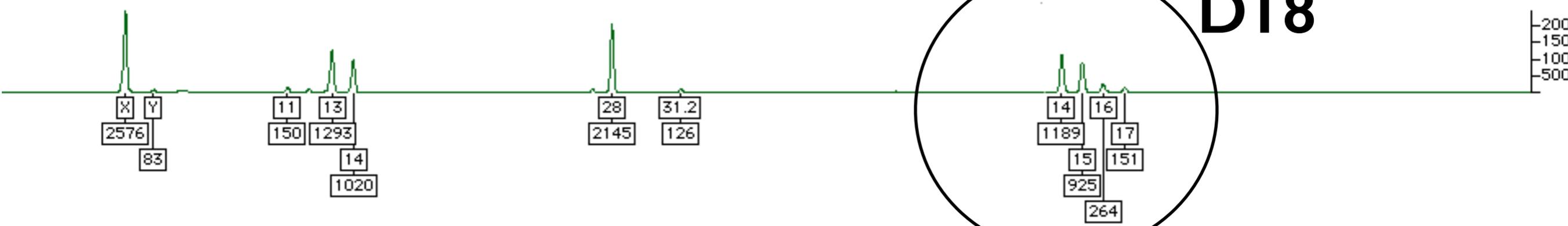
Peak: Scan 5520 Size 286.83 Height 68 Area 638 Category: D18S51:13

30 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380

:03\_3015...15\_09.fsa 9 Blue 301583097\_20060713001\_AHH352\_2 D3S1358/VWA/D16S539/D2S1338



:03\_3015...15\_09.fsa 9 Green 301583097\_20060713001\_AHH352\_2 AMEL/D8S1179/D21S11/D18S51

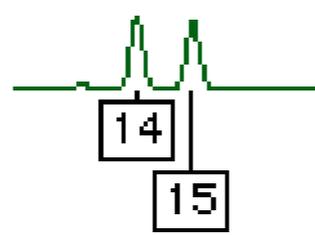


**D18**

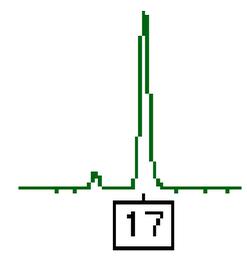
:03\_3015...15\_09.fsa 9 Yellow 301583097\_20060713001\_AHH352\_2 D19S433/TH01/FGA(-\*P06-971/019)



**D18: victim**



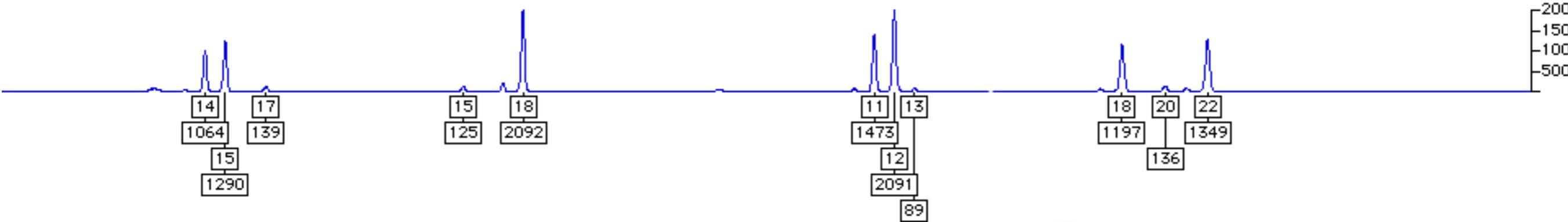
**, suspect**



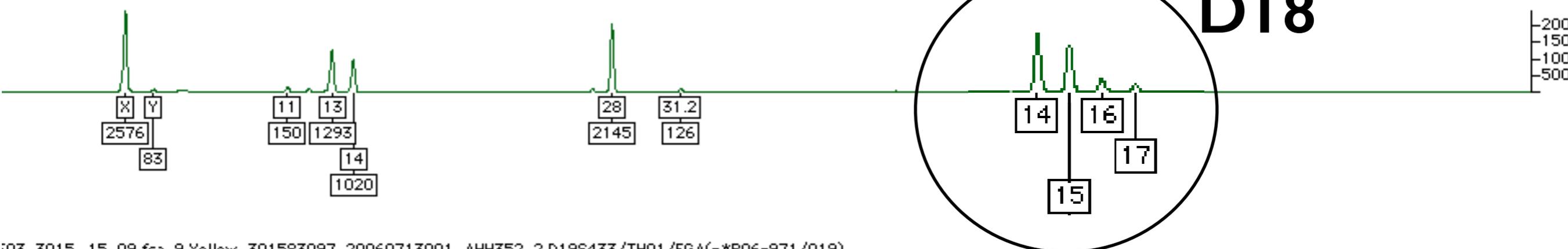
Peak: Scan 5520 Size 286.83 Height 68 Area 638 Category: D18S51:13

30 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380

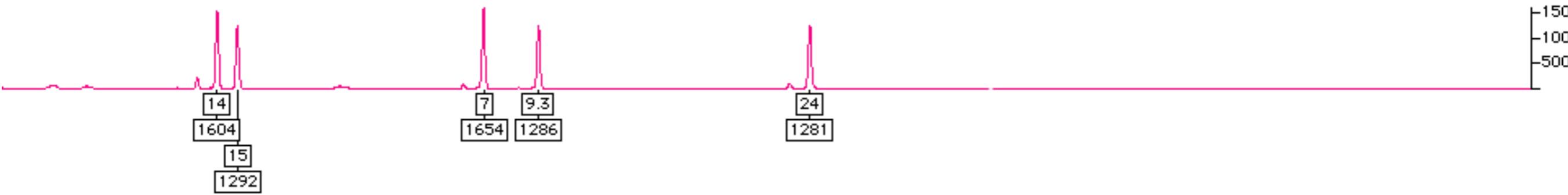
:03\_3015...15\_09.fsa 9 Blue 301583097\_20060713001\_AHH352\_2 D3S1358/VWA/D16S539/D2S1338



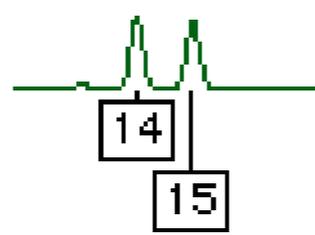
:03\_3015...15\_09.fsa 9 Green 301583097\_20060713001\_AHH352\_2 AMEL/D8S1179/D21S11/D18S51



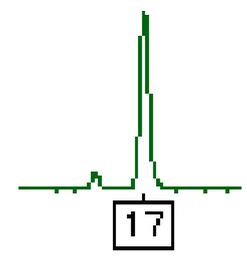
:03\_3015...15\_09.fsa 9 Yellow 301583097\_20060713001\_AHH352\_2 D19S433/TH01/FGA(-\*P06-971/019)



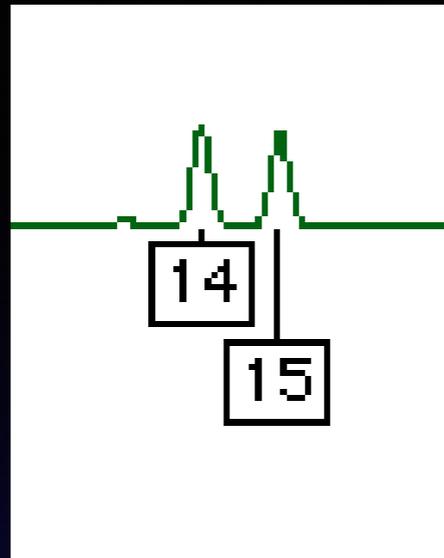
**D18: victim**



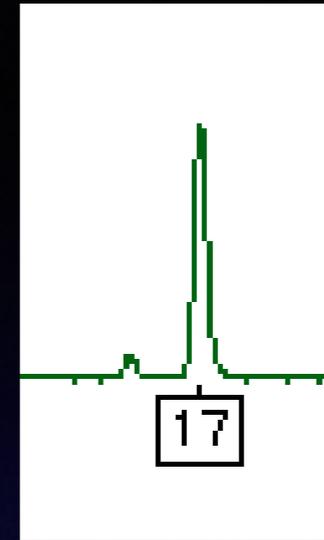
, **suspect**



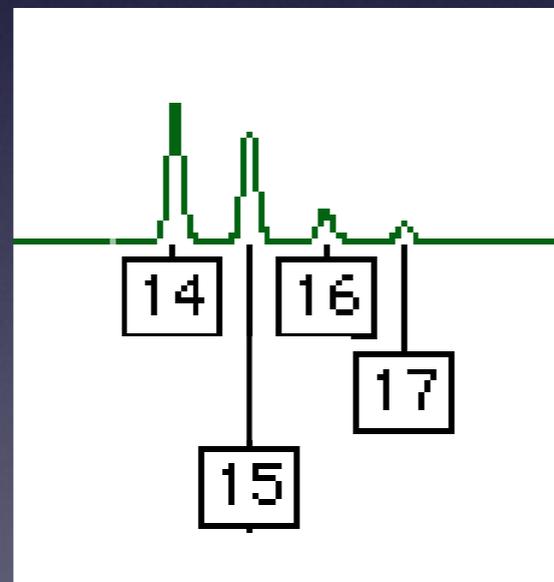
D18: victim



suspect



D18: “mixture profile” (fingernail scaping of victim)



Netherlands Forensic Institute: “If this trace is from two persons, we can exclude the suspect.

If it’s from three, we can’t”

(2008)

NFI expert , following official (deterministic)  
interpretation rules:

“If this trace is from two persons, we can exclude the  
suspect. If it’s from three, we can’t”

The rest of the profile, visually, matches a 90-10  
mixture victim-suspect rather well !

D18 is the most well-known locus for *mozaicism*  
which occurs there in *at least* 1 in 5000 persons !?

---

# Review

---

## Role of Short Tandem Repeat DNA in Forensic Casework in the UK—Past, Present, and Future Perspectives

*BioTechniques* 32:366-385 (February 2002)

**Peter Gill**

Forensic Science Service,  
Birmingham, UK

### ABSTRACT

*The analysis of short tandem repeat (STR) DNA sequences is of fundamental importance to forensic science because they have become the recognized standard in constructing national public databases.*

### DEVELOPMENT OF MULTIPLEXED SYSTEMS

Early multiplexes consisted of few loci that were based on simple short tandem repeats (STRs). The four-locus “quadruplex” was probably the first to be widely used (44); because it consisted of few STRs, the match probability was consequently high—1 in 10 000. In 1996, a six-locus STR system (57,58) combined with the amelogenin sex test (61) was introduced—known as the

ous years, all six loci of the older SGM system were retained in the new *AmpFl* STR SGM Plus system.

### Development and Harmonization of National DNA Databases

The harmonization of STR loci has been achieved by collaboration at the international level. Notably, the European DNA profiling group (EDNAP) carried out a series of successful studies to identify and recommend STR loci for the

## Somatic Mutation

If a somatic mutation occurs during embryological development, then two types of cells with different genotypes may coexist, and this leads to a three-banded profile (Figure 4). The peak areas will depend on the relative proportion of the mutant cell and will not be equivalent. This is arguably the most difficult condition to elucidate because it is possible that not all tissues will demonstrate somatic mutation. The incidence of somatic mutation is variable—out of 120 000 samples, not one has been observed at the HUMTH01 locus, whereas the incidence is approximately 1 in 5000 at the D18S51 and HUMFIBRA loci. It is possible that some somatic mutations will be indistinguishable from stutters; therefore, these figures are probably underestimates because they are only recorded if unambiguous.

The genetic phenomena described (trisomy, translocation, and primer

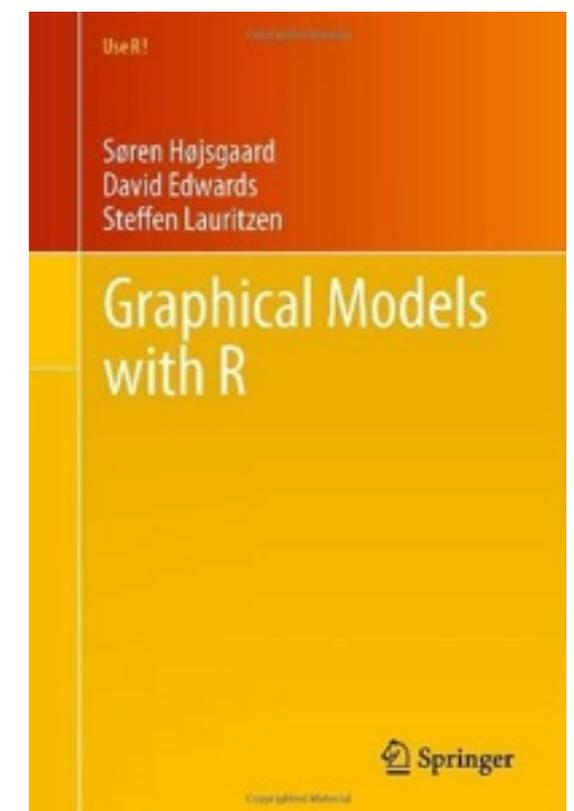
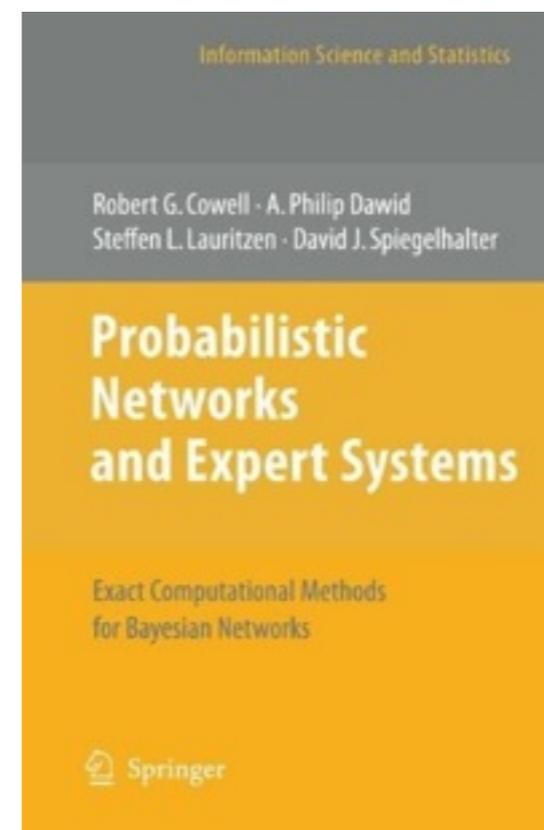
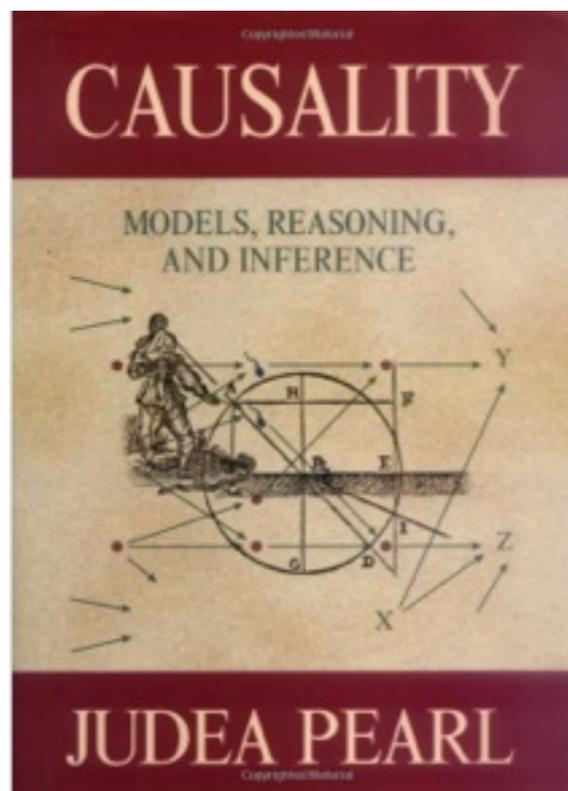
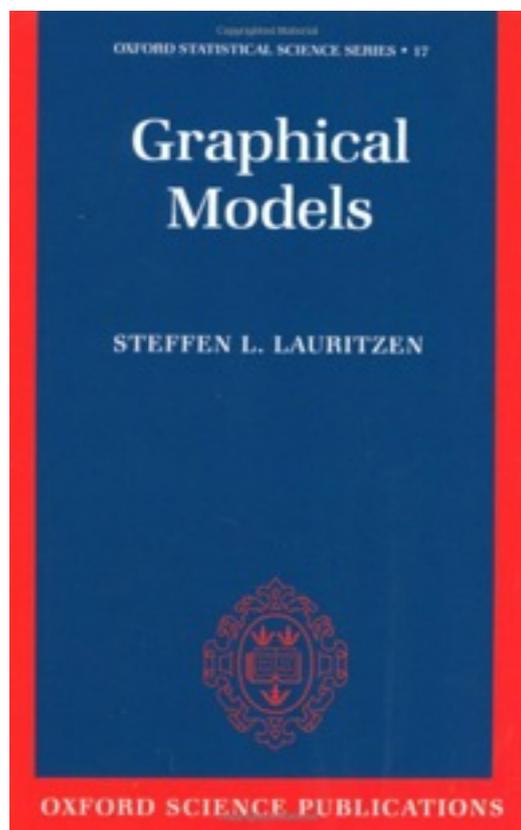
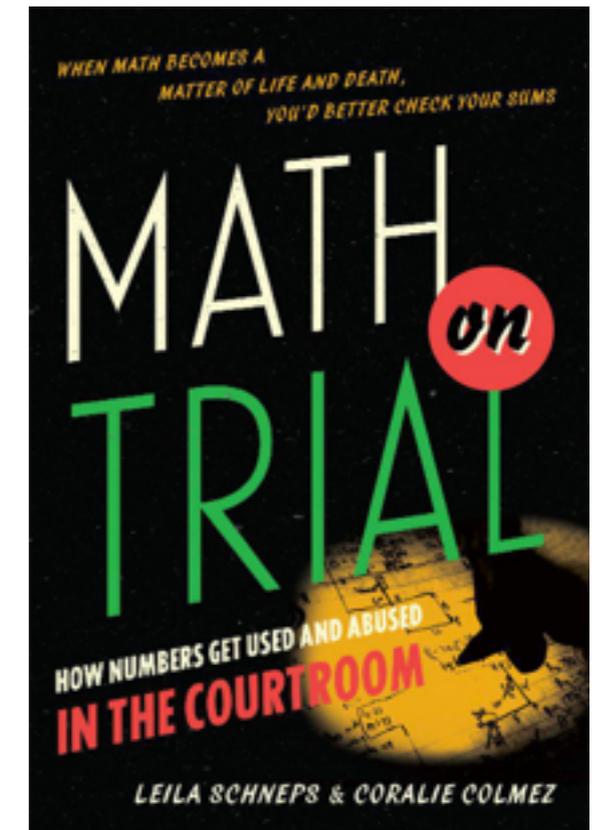
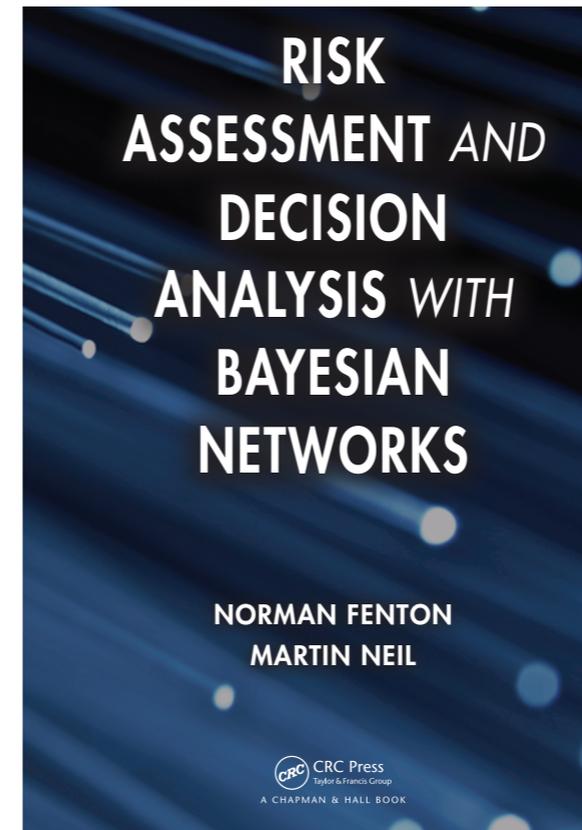
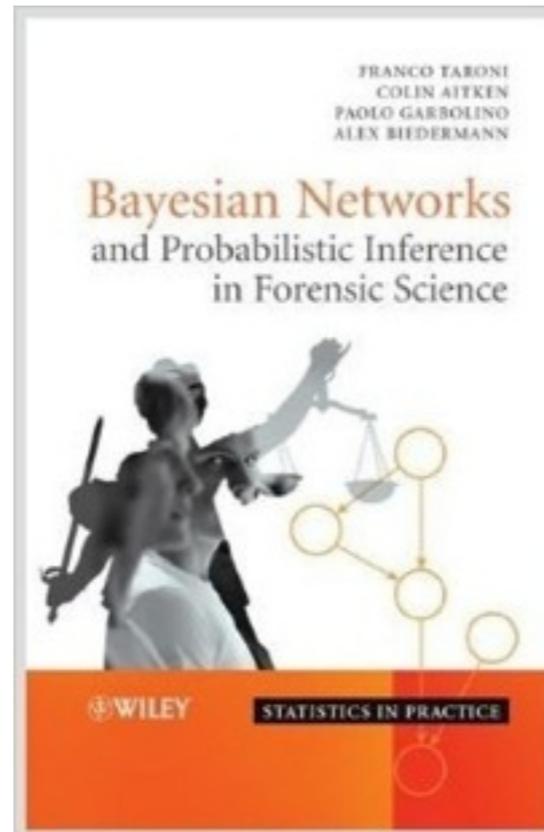
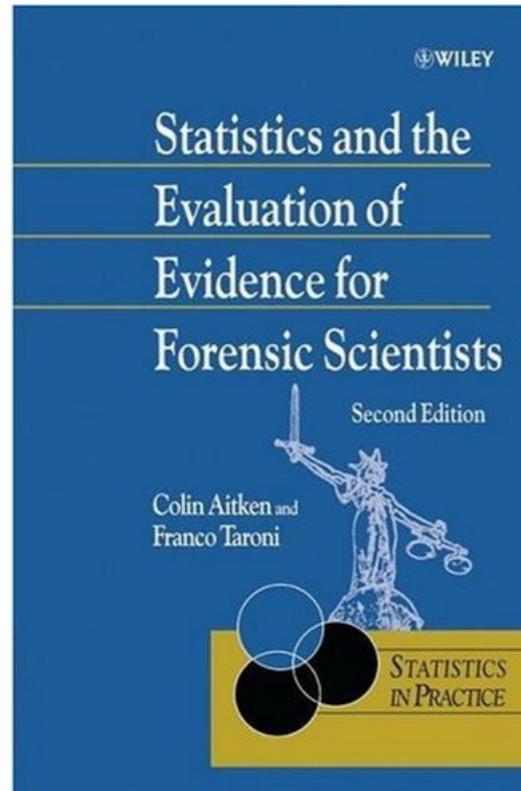
binding site mutations) can be verified by the analysis of the reference sample, which should also demonstrate the same anomaly unless a tissue-specific somatic mutation has occurred. In the latter case, confirmation may depend on a reference sample that has the same origin as the case stain, although we cannot completely rule out the possibility that the appearance of somatic mutations could vary over time within tissues such as the buccal lining, which consists of rapidly dividing cells.

To summarize, an understanding of the behavior of the DNA profiling system is important to assess potential mixtures. Loci will behave somewhat differently from each other, but it is possible to generalize. Here are some of the key features: (i) the smallest peak area of a heterozygote will usually be greater than 60% of the size of its partner (peak area or peak height); (ii) within the previous guideline, the high molecular weight peak is often smaller than the low mol-

A classical painting of a woman, likely a personification of Justice or Truth, holding a scale of justice. She is depicted from the waist up, wearing a green and blue robe. Her right arm is raised, holding the top of the scale, while her left hand holds the weighing pan. The background is dark and textured. The text "Statistical Paradigms" is overlaid in white, sans-serif font in the center of the image.

# Statistical Paradigms

# Forensic statistics & Graphical Models



# Statistical paradigms

- Bayes (one person statistics)
- Frequentist (two person, collaborative statistics)
- Likelihood (avoiding the issue)

# Bayes' rule

- Posterior odds  
= prior odds  $\times$  likelihood ratio
- Likelihood ratio  
=  $\text{Prob}( \textit{Evidence} \mid H_P ) : \text{Prob}( \textit{Evidence} \mid H_D )$
- The likelihood ratio is a measure of the **evidential value** of the “Evidence” in question

Note: The prior is up to the judges ... they must combine all of the available legal evidence, not just the particular piece we are discussing here.

The scientist who is called to help the court interpret “Evidence” should merely inform the court of its evidential value

# Example: Monty Hall

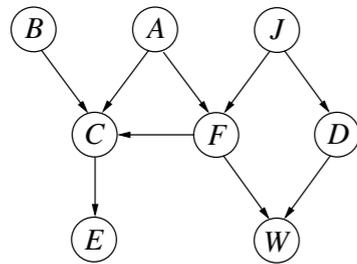
- Player chooses door 1
- At this point, **prior** to obtaining any evidence, **odds** on “car is behind door 1” : “car is behind door 2”  
= **1 : 1**
- Host opens door 3 and reveals goat ( = evidence )
- $\text{Prob}(\text{host opens 3} \mid \text{car behind door 1}) = 1/2$
- $\text{Prob}(\text{host opens 3} \mid \text{car behind door 2}) = 1$
- Therefore, **likelihood ratio** =  $1/2 : 1 = 1 : 2$
- Therefore (by Bayes), **posterior odds** = **1 : 2**

A classical painting of a woman, likely a personification of Justice or Truth, holding a pair of scales. She is depicted from the waist up, wearing a green and blue robe with a pink sash. Her right arm is raised, holding the top of the scales, while her left hand rests on a dark object, possibly a book or a tablet. The background is dark and textured. The text "Bayes' nets" is overlaid in white, centered over the woman's face and upper torso.

# Bayes' nets

Table 2.11 Definitions of the nodes used in the 'Lulu' Bayesian network in Figure 2.9.

- A*: Jack stabbed Lulu.  
*B*: Bloodstain at crime scene comes from offender.  
*C*: Bloodstain at crime scene comes from Jack.  
*E*: Jack's blood sample and crime stain share the same DNA profile.  
*F*: Jack was in a certain place *f* near the house where Lulu lived shortly after the time the crime was committed.  
*W*: John says that Jack was in place *f* shortly after the time when the crime was committed.  
*J*: Jack loved Lulu.  
*D*: John was jealous of Jack.



# Bayes nets

Visualisation of complex causal (but not necessarily deterministic) relations

Graphical and computational tool for correct reasoning with uncertainty

Rapid computation of evidential value in situations with complex causal relations between different components of the evidence in question.

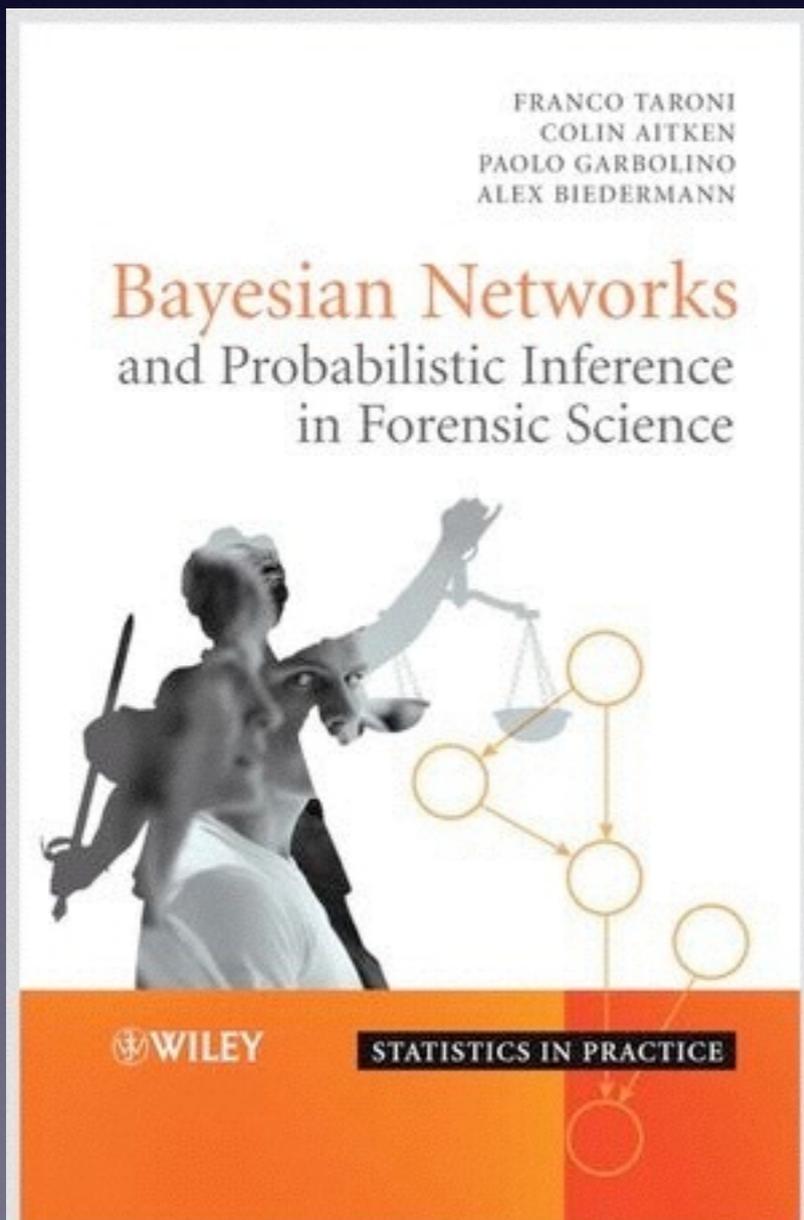


Table 2.12 Probability distributions for Figure 2.9.

<i>A</i> : $Pr_0(A) = 0.01$	<i>J</i> : $Pr_0(J) = 0.9$
<i>B</i> : $Pr_0(B) = 0.95$	<i>F</i> : $Pr_0(F   A, J) = 1$
<i>C</i> : $Pr_0(C   A, B, F) = 1$	$Pr_0(F   A, \bar{J}) = 1$
$Pr_0(C   A, B, \bar{F}) = 1$	$Pr_0(F   \bar{A}, J) = 0.1$
$Pr_0(C   A, \bar{B}, F) = 0$	$Pr_0(F   \bar{A}, \bar{J}) = 0.01$
$Pr_0(C   A, \bar{B}, \bar{F}) = 0$	<i>D</i> : $Pr_0(D   J) = 1$
$Pr_0(C   \bar{A}, B, F) = 0$	$Pr_0(D   \bar{J}) = 0.6$
$Pr_0(C   \bar{A}, B, \bar{F}) = 0$	<i>W</i> : $Pr_0(W   F, D) = 0.99$
$Pr_0(C   \bar{A}, \bar{B}, F) = 0.1$	$Pr_0(W   F, \bar{D}) = 0.99$
$Pr_0(C   \bar{A}, \bar{B}, \bar{F}) = 0.01$	$Pr_0(W   \bar{F}, D) = 0.8$
<i>E</i> : $Pr_0(E   C) = 1$	$Pr_0(W   \bar{F}, \bar{D}) = 0.01$
$Pr_0(E   \bar{C}) = 0.001$	

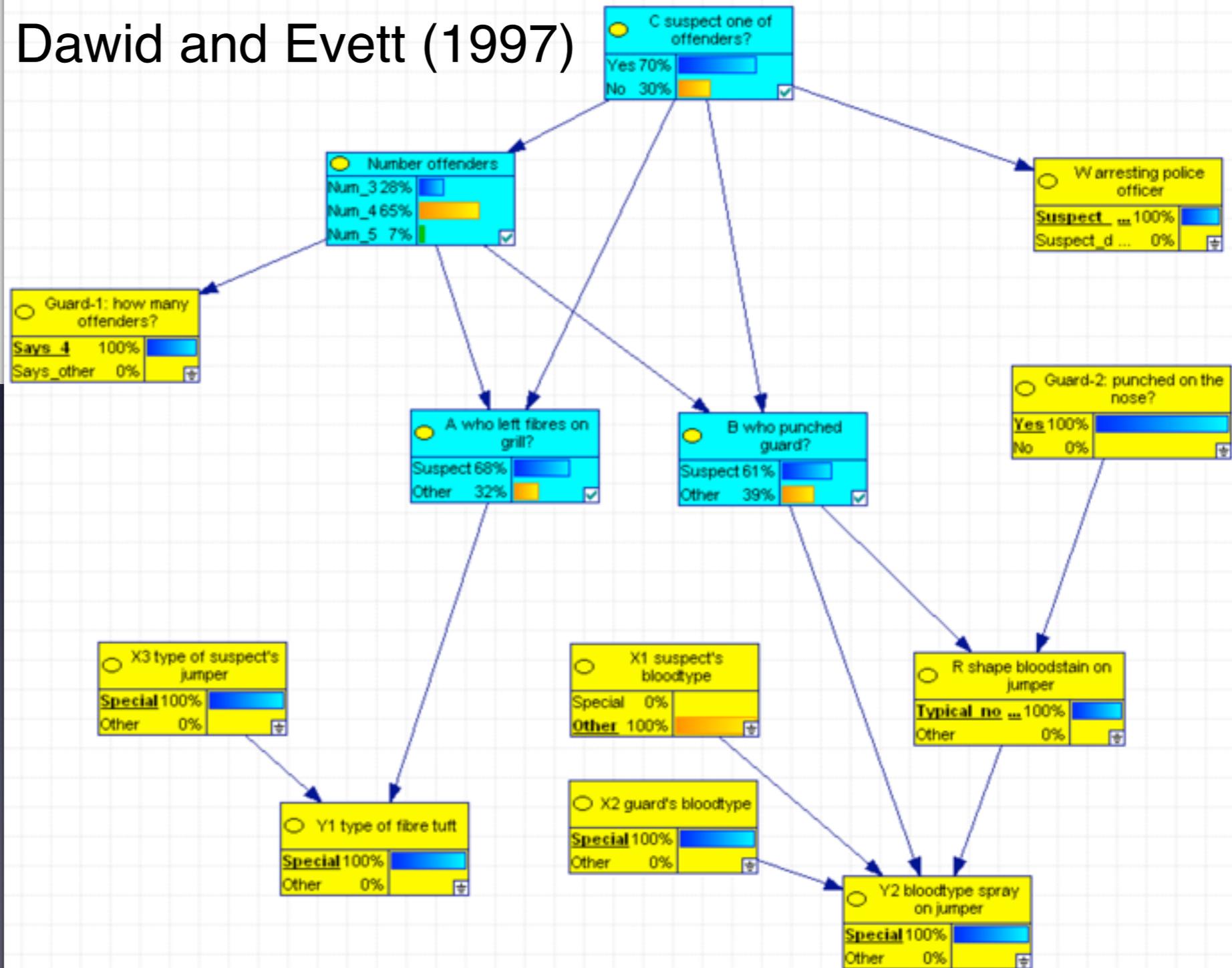
# Using a Graphical Method to Assist the Evaluation of Complicated Patterns of Evidence

**REFERENCE:** Dawid AP, Evett IW. Using a graphical method to assist the evaluation of complicated patterns of evidence. *J Forensic Sci* 1997;42(2):226-231.

**ABSTRACT:** The forensic scientist often faces the task of interpreting patterns of evidence which involve many variables. Combining different items of evidence within a complex framework of circumstances requires logical powers of reasoning and this can be assisted by formal methods. We discuss one such method which, as has already been pointed out by Aitken and Gammerman (1), offers considerable potential for creating probabilistic expert systems to assist in evidence interpretation. In particular, we show how the method, which is based on a directed acyclic graph, enables dependencies between different aspects of the evidence to be considered. The discussion is based on an imaginary case example.

**KEYWORDS:** forensic science, interpretation, expert systems, probability, Bayesian, fibers, trace evidence

## Dawid and Evett (1997)



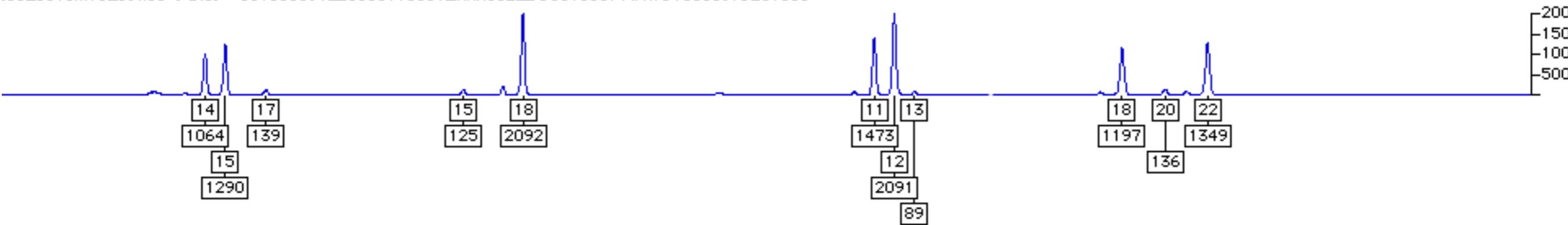
A classical painting of a woman, likely a personification of Justice or Equity, holding a scale of justice. She is depicted from the waist up, wearing a green and blue robe with a red sash. Her right hand holds the scale's beam, and her left hand holds a scale pan. The background is dark and indistinct.

# DNA profiles (genetics and biochemistry)

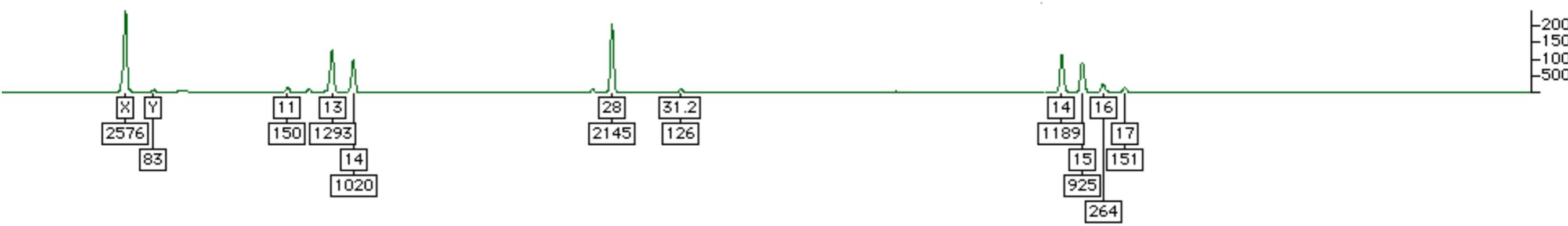
Peak: Scan 5520 Size 286.83 Height 68 Area 638 Category: D18S51:13

30 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380

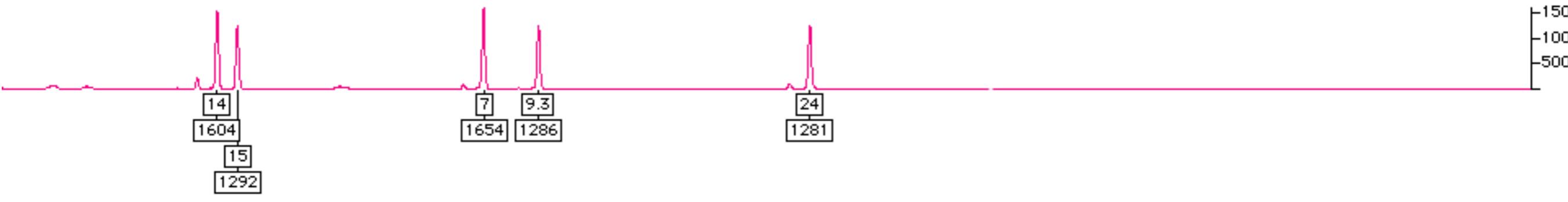
:03\_3015...15\_09.fsa 9 Blue 301583097\_20060713001\_AHH352\_2 D3S1358/VWA/D16S539/D2S1338



:03\_3015...15\_09.fsa 9 Green 301583097\_20060713001\_AHH352\_2 AMEL/D8S1179/D21S11/D18S51



:03\_3015...15\_09.fsa 9 Yellow 301583097\_20060713001\_AHH352\_2 D19S433/TH01/FGA(-\*P06-971/019)



- Each group of peaks corresponds to one locus on a different chromosome
- Genotype of **one person** at **one locus** is pair  $(m,n)$ ,  $m \leq n$  (numbers of repeats in two STR alleles), e.g. (7,9) or (18,18)
- *Relative size of peak is roughly proportional to sum over contributors of:*
  - 0, 1, or 2 depending on # alleles contributed
  - **x** proportion of contributor to mixture
- *Absolute size of peak is (almost) irrelevant*
- The peak sizes are definitely *random*, small peaks much more than large peaks

# Randomness

- Mendel's laws and relative frequencies of alleles in general population give us a fairly well understood model for the genotype of a random unknown person
- PCR procedure generates randomly sized peaks and suffers from "artefacts":  
stutter, dropout, silent alleles, mozaicism ...

A classical painting of a woman, likely representing Justice or a similar allegorical figure. She is shown from the waist up, wearing a green and blue draped garment. She holds a pair of golden scales of justice in her right hand, which is raised. Her left hand is resting on her lap. The background is dark and indistinct.

# Application to Tamara Wolvers case



Research articles

# Probabilistic modelling for DNA mixture analysis

R.G. Cowell<sup>a,\*</sup>, S.L. Lauritzen<sup>b</sup>, J. Mortera<sup>c</sup>

<sup>a</sup> *Cass Business School, 106 Bunhill Row, London EC1Y 8TZ, UK*

<sup>b</sup> *Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK*

<sup>c</sup> *Dipartimento di Economia, Università Roma Tre, Via Silvio D'Amico, 77, 00145 Roma, Italy*

Received 13 August 2007; accepted 7 October 2007

## Abstract

Taking peak area information into account when analysing STR DNA mixtures is acknowledged to be a difficult task. There have been a number of non-probabilistic approaches proposed in the literature, and some have been incorporated into computer systems, but comparatively little has been published from a probabilistic perspective. Here we briefly review our previous work on using Bayesian networks to analyse two-person mixtures within a probabilistic framework, and present preliminary results obtained for analysing two-person and three-person mixtures that combine peak area information from multiple independent samples.

© 2008 Elsevier Ireland Ltd. All rights reserved.

*Keywords:* DNA mixtures; Bayesian networks; Multiple traces; Multiple contributors

## 1. Introduction

In a recent series of papers [1–3] we have presented a probabilistic methodology for analysing peak area information from DNA mixtures based on Bayesian networks. A representative fragment of these networks is shown in Fig. 1

apparatus *after* amplification of the mixture sample. We model the stochastic variations in these areas by Gamma distributions, where the Gamma distribution of the area for allele  $a$  depends on the mean  $\mu_a$  and has expectation proportional to  $\mu_a$ ; similarly for alleles  $b$  and  $c$ . For further details of the Gamma model and Bayesian networks and how the probability

- 2010: first ever court application of Cowell, Lauritzen, Mortera (2008) “graphical model” (Bayes’ nets) for forensic DNA mixture profiles
- RDG with Cowell, Lauritzen, Mortera and master students  
Jasper van Wamelen, Giulia Cereda

# Basis model: one locus, three adjacent STR numbers

*R.G. Cowell et al. / Forensic Science International: Genetics Supplement Series 1 (2008) 640–642*

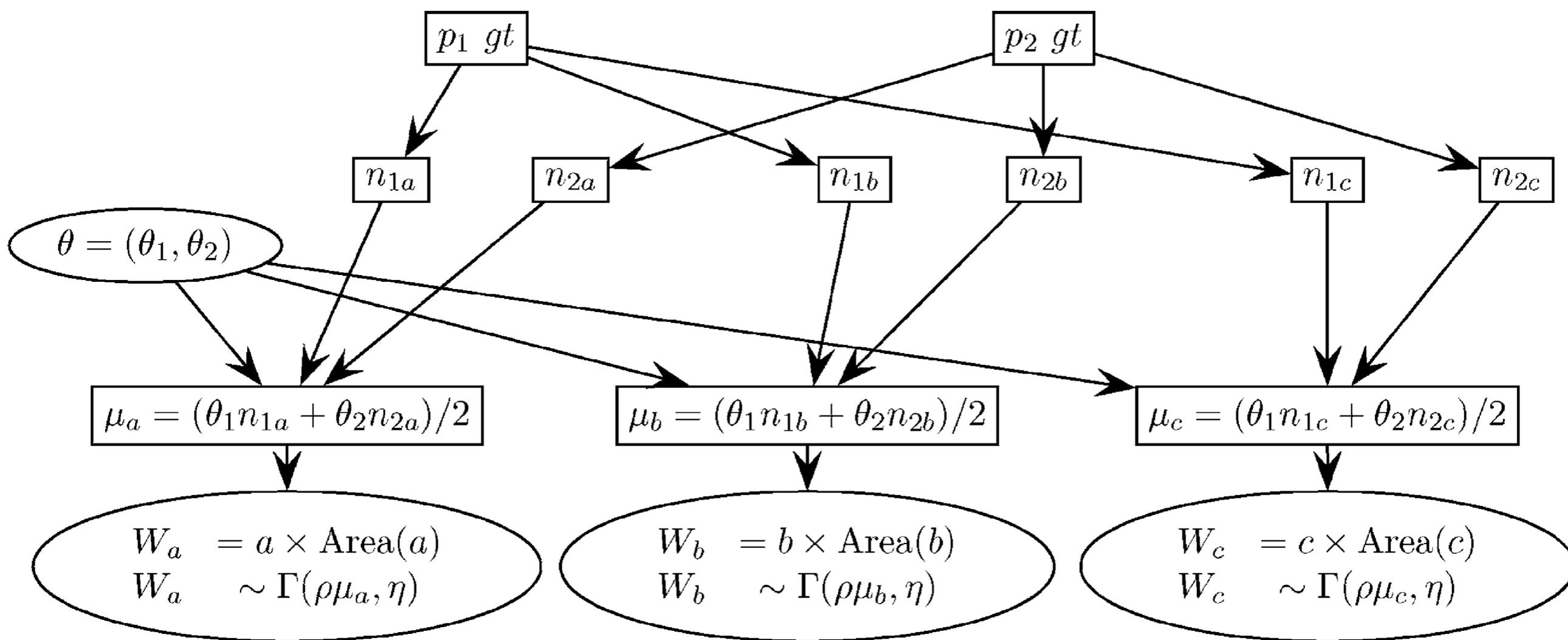
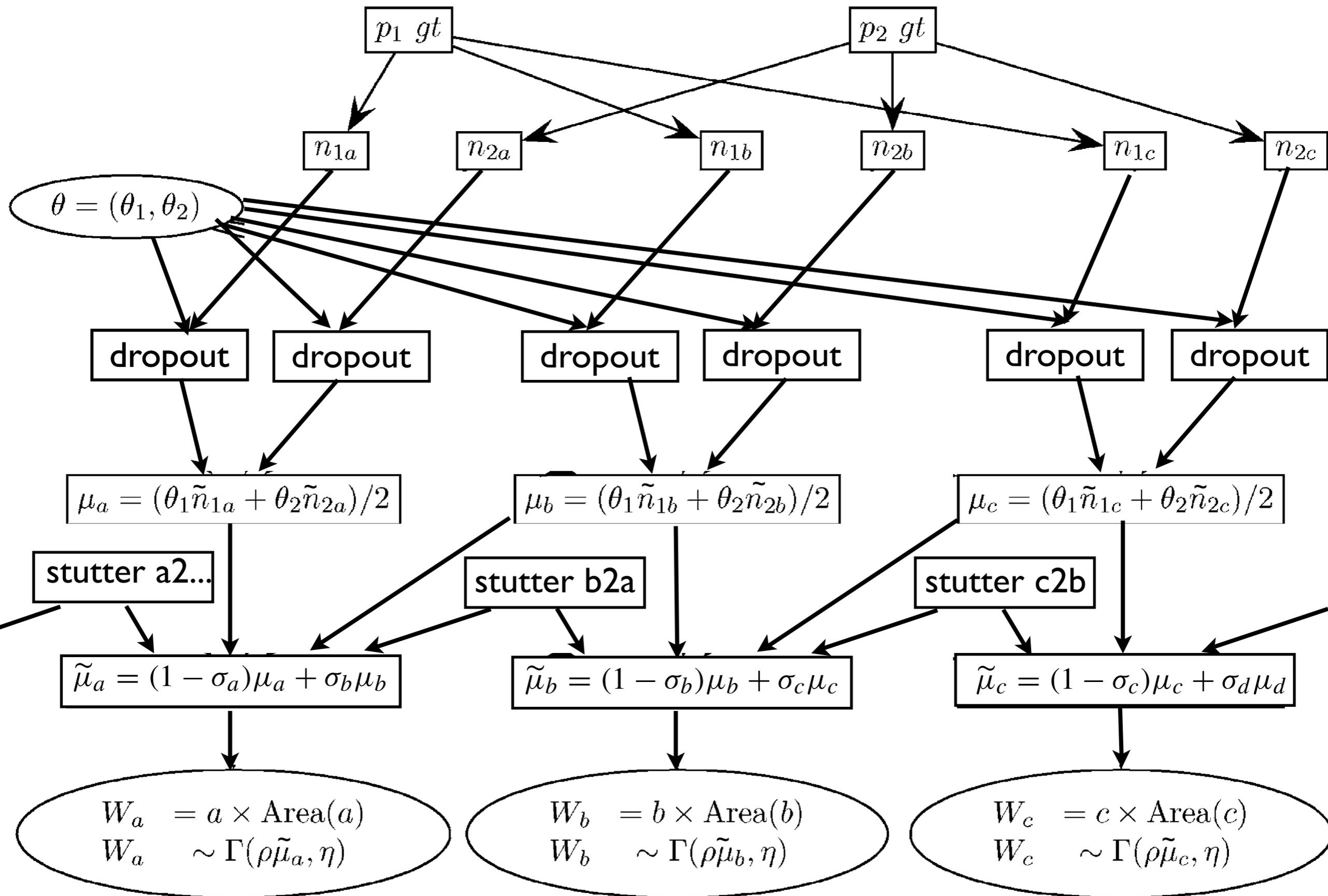
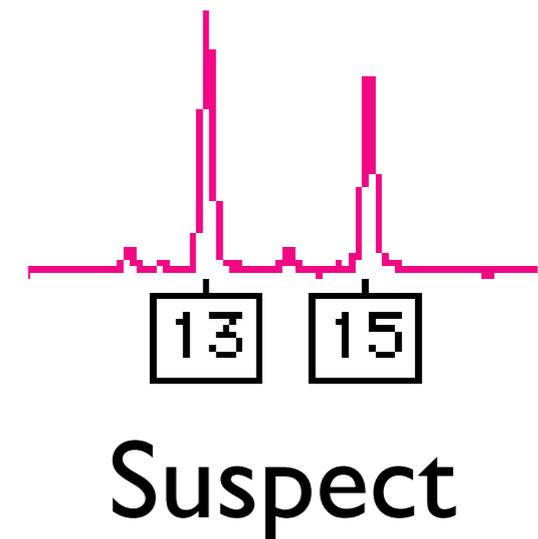
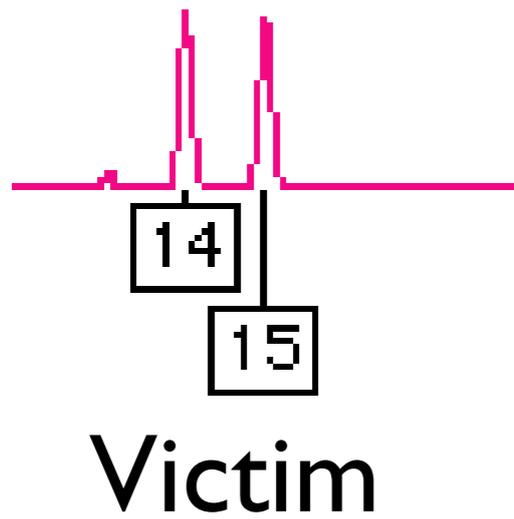
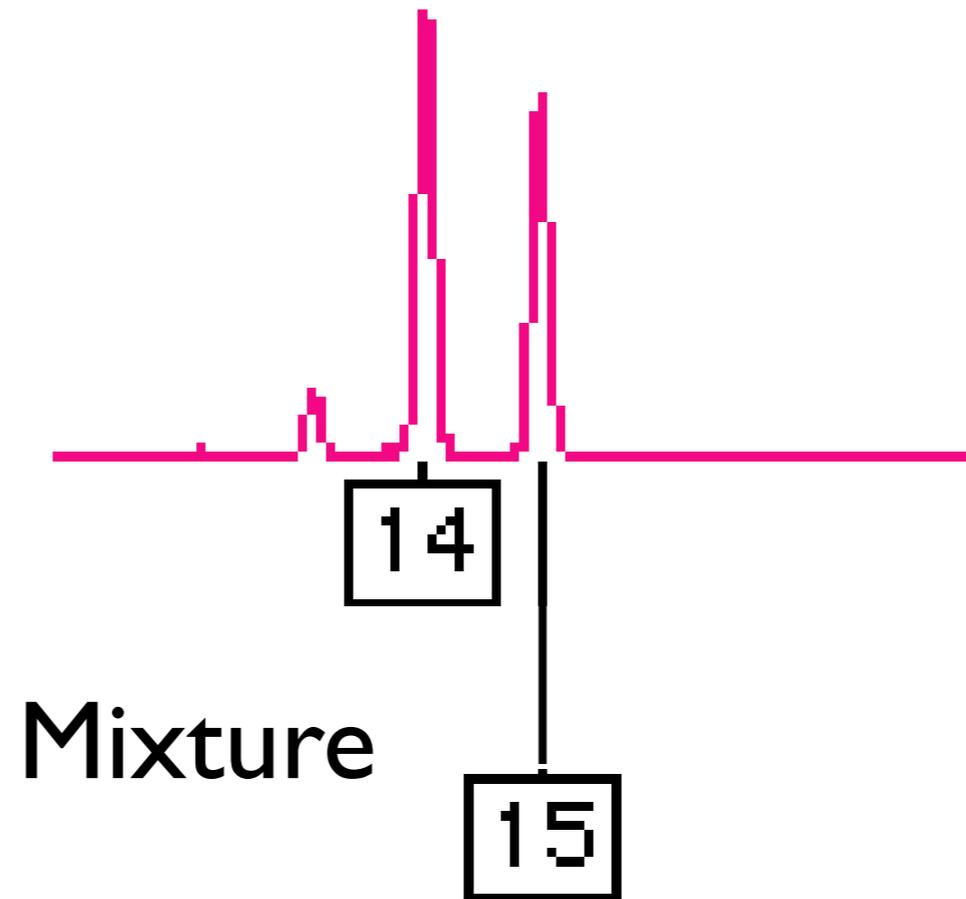


Fig. 1. Bayesian network fragment for modelling peak areas in a mixture.

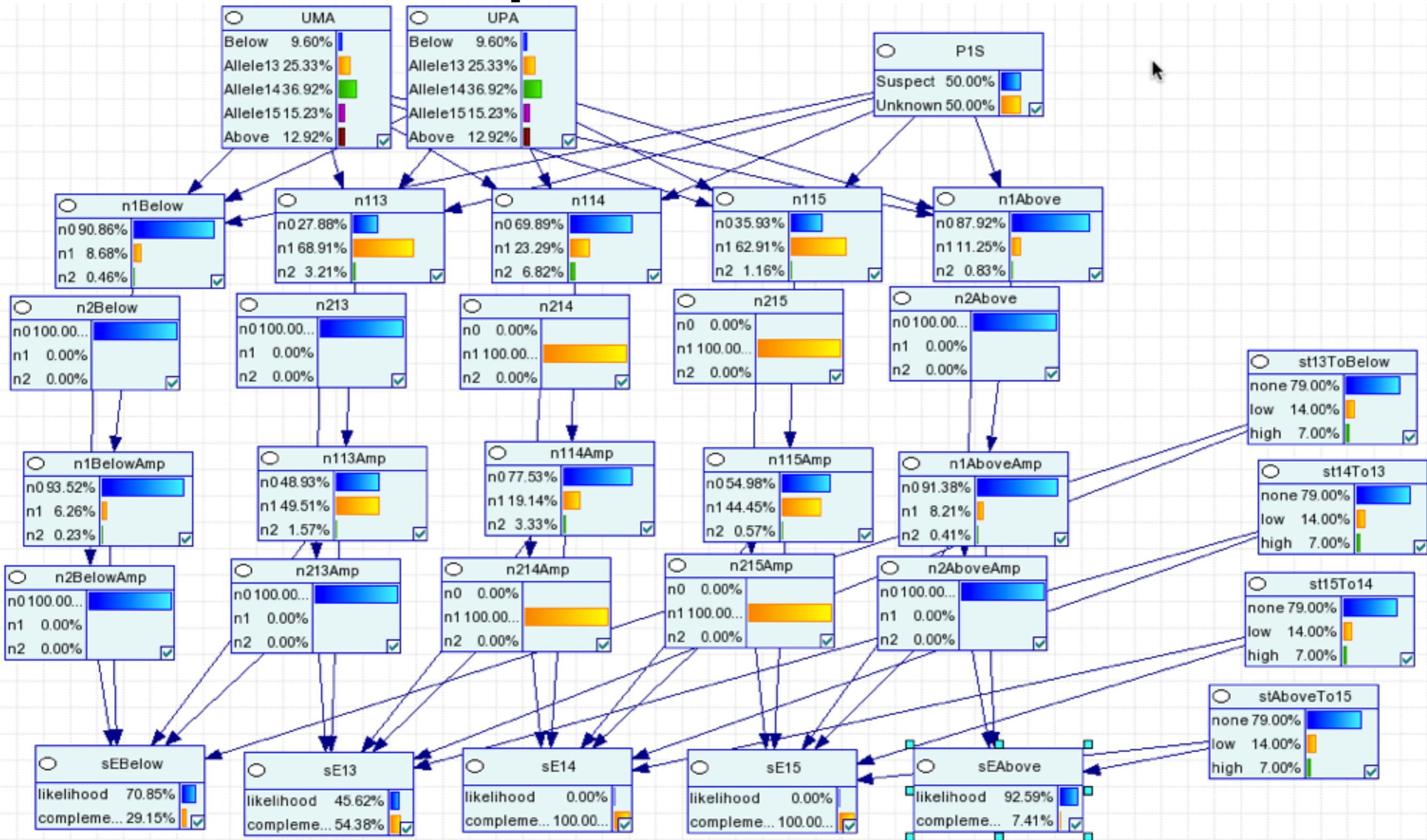
# Extended model (with dropout and stutter)



# Locus D19 (not problematic...?)

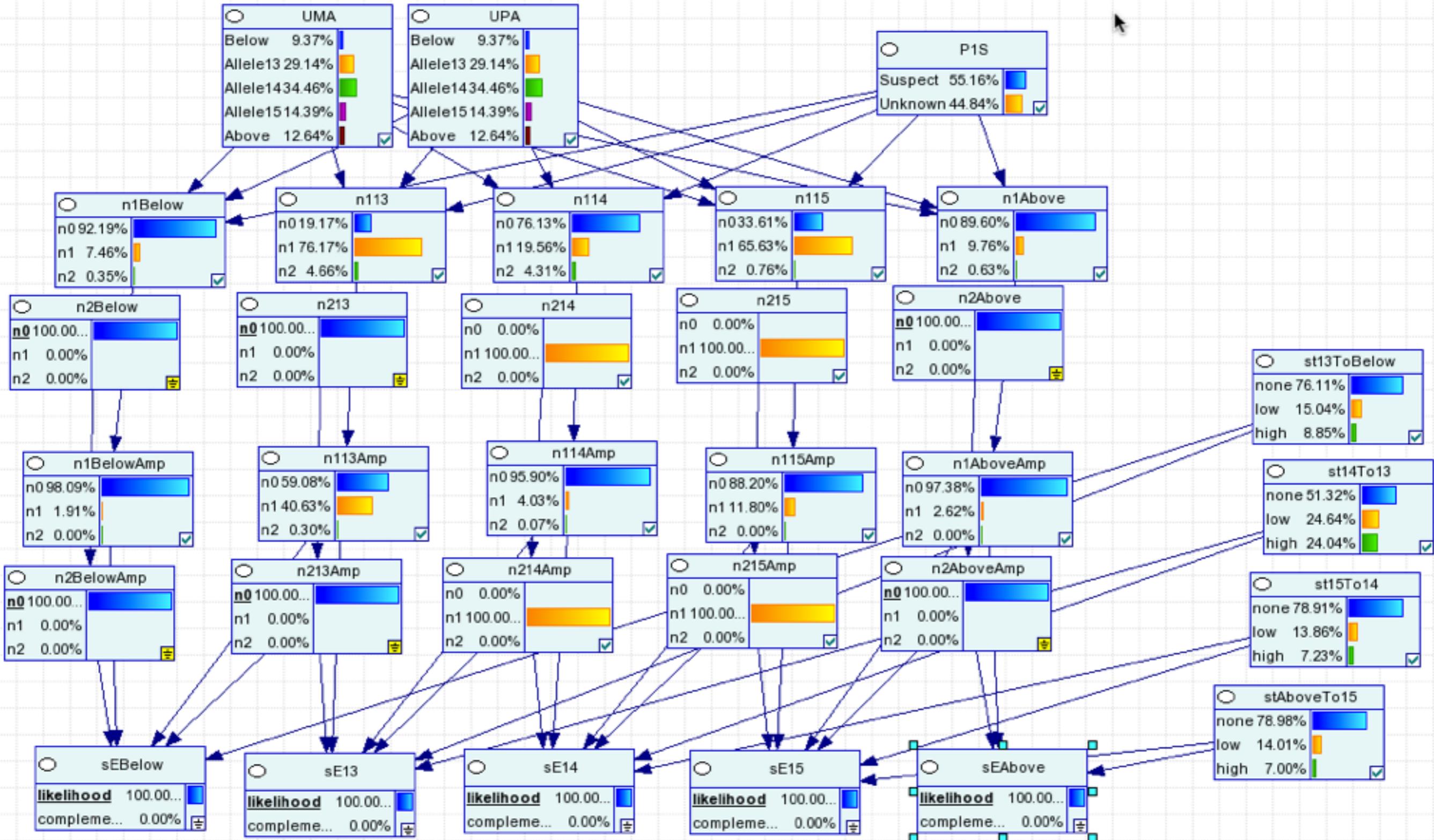


# Locus D19: prior



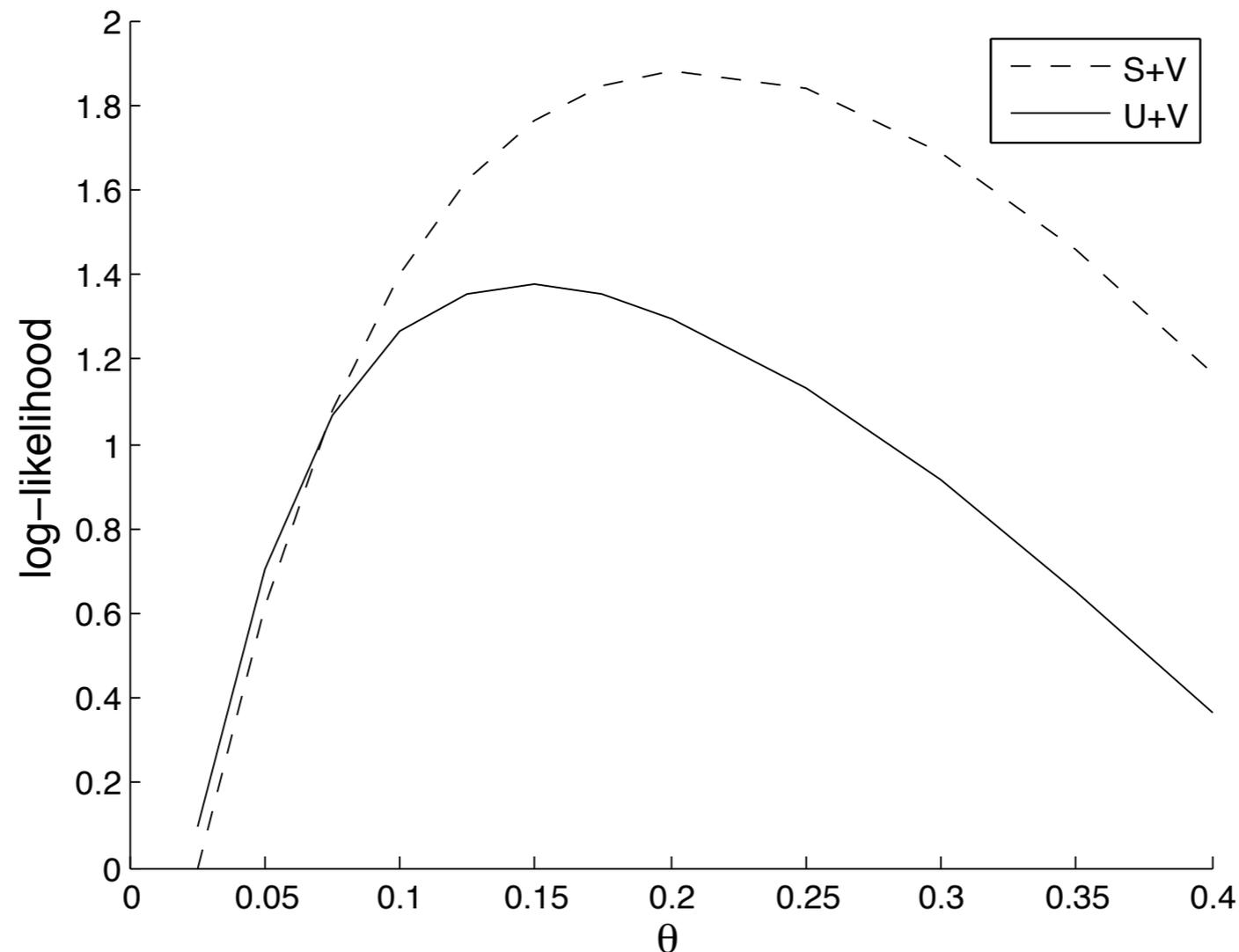
Assume 90:10 mixture

# Locus D19: posterior



Assume 90:10 mixture

# What does D19 say about the mixture proportion?



Defence would like  $\theta < 0.08$ , prosecution  $\theta > 0.25$

Combining three replicate mixture profiles:

*Evidential value for*

**victim+suspect+mosaic** versus **victim+unknown**  
is around 10 000 : 1 (“very strong evidence”)

*Evidential value for*

**victim+suspect+unknown** versus **victim+unknown**  
**+unknown**

is also around 10 000 : 1 (“very strong evidence”)

Free tools: GeNle, octave, R

GeNle: <http://genie.sis.pitt.edu/>

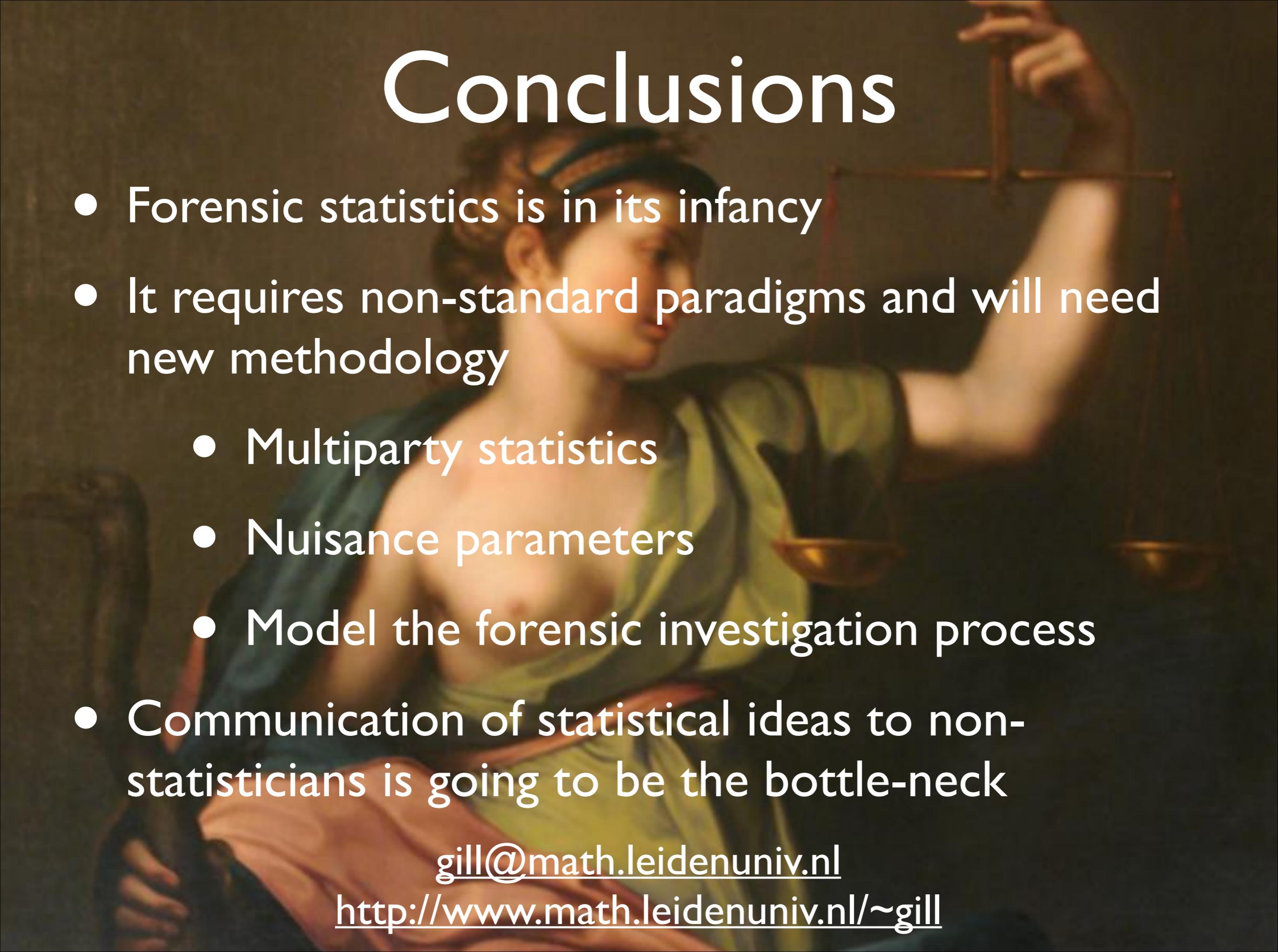
# Experience

- Important to over-estimate “noise parameters” in model to compensate for misspecification – otherwise you will overestimate weight of evidence
- Once you have done that, our results were pretty robust to unknown nuisance parameters
- The courts, and the DNA specialists, are not ready for this

# New technologies

- New technologies are generating new forensic data of frightening dimension, little scientific understanding
- Plug-in methods (fit models using training data, then “estimate” likelihood ratio) tends to grossly overestimates weight of evidence
- Present research plans: tune fitting to task
- Courts won't be ready for this for many years to come (and shouldn't be)

# Conclusions

A background image of a woman, likely a personification of Justice, holding a pair of scales. She is wearing a green and blue robe and has a serious expression. The scales are golden and she is holding them with both hands. The background is dark and slightly blurred.

- Forensic statistics is in its infancy
- It requires non-standard paradigms and will need new methodology
  - Multiparty statistics
  - Nuisance parameters
  - Model the forensic investigation process
- Communication of statistical ideas to non-statisticians is going to be the bottle-neck

[gill@math.leidenuniv.nl](mailto:gill@math.leidenuniv.nl)

<http://www.math.leidenuniv.nl/~gill>