Science for Communities

Article:

Gittelson, S., Kalafut, T., Myers, S., Taylor, D., Hicks, T., Taroni, F., Evett, I.W., Bright, J.-A. and Buckleton, J. (2016), **A Practical Guide for the Formulation of Propositions in the Bayesian Approach to DNA Evidence Interpretation in an Adversarial Environment**. *Journal of Forensic Sciences*, 61: 186-195.

INSTITUTE OF ENVIRONMENTAL SCIENCE AND RESEARCH LIMITED
Kenepuru Science Centre: 34 Kenepuru Drive, Kenepuru, Porirua 5022 | PO Box 50348, Porirua 5240, New Zealand
T: +64 4 914 0700 **F:** +64 4 914 0770
www.esr.cri.nz

A Practical Guide for the Formulation of Propositions in the Bayesian Approach to DNA Evidence Interpretation in an Adversarial Environment

Simone Gittelson[1*], Tim Kalafut[2], Steven Myers[3], Duncan Taylor[4], Tacha Hicks[5&6], Franco Taroni[5], Ian W. Evett[7], Jo-Anne Bright[8] and John Buckleton[8].

[1] *Department of Biostatistics, University of Washington, Seattle, WA 98195*

[2] *U.S. Army Criminal Investigation Laboratory, 4930 North 31st Street, Forest Park, GA 30297*

[3] *California Department of Justice, Jan Bashinski DNA Laboratory, Richmond, CA 94804*

[4] *Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia*

[5] *School of Criminal Justice, University of Lausanne, 1015 Lausanne, Switzerland*

[6] *Foundation for Continuing Education UNIL-EPFL, EPFL Innovation Park, 1015 Lausanne, Switzerland*

[7] *Principal Forensic Services Ltd, 34 Southborough Road, Bromley BR1 2EB, UK*

[8] *Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland 1142, New Zealand*

*Corresponding author at: Department of Biostatistics, University of Washington, Seattle, WA, 98195. Email address: simone.gittelson@gmail.com

Abstract

The interpretation of complex DNA profiles is facilitated by a Bayesian approach. This approach requires the development of a pair of propositions: one aligned to the prosecution case and one to the defense case. This note explores the issue of proposition setting in an adversarial environment by a series of examples. A set of guidelines generalize how to formulate propositions when there is a single person of interest and when there are multiple individuals of interest. Additional explanations cover how to handle multiple defense propositions, relatives, and the transition from subsource level to activity level propositions. The propositions depend on case information and the allegations of each of the parties. The prosecution proposition is usually known. The authors suggest that a sensible proposition is selected for the defense that is consistent with their stance, if available, and consistent with a realistic defense if their position is not known.

Keywords: forensic science, forensic DNA, likelihood ratio, propositions, Bayesian approach, DNA mixtures

Introduction

There is a reappraisal of DNA interpretation methods underway internationally, and this has been given strong support from several standard-setting organizations such as the Scientific Working Group on DNA Analysis Methods, (SWGDAM) (1), the European Network of Forensic Science Institutes (ENFSI), the Association of Forensic Science Providers (AFSP) (2) and the International Society for Forensic Genetics (ISFG) (3, 4). Europe and the UK have a long association with the Bayesian approach but North America has, for many years, deployed other methods such as the cumulative probability of inclusion (CPI) (5) and the random match probability (RMP) (6). The emerging methods that enable interpretation of low template and mixed samples, however, are based on the assignment of a likelihood ratio (*LR*):

$$LR = \frac{Pr(E|H_p,I)}{Pr(E|H_d,I)} \quad (1)$$

where E stands for the findings (e.g., DNA profile(s), presumptive test results, observations), H_p and H_d the propositions representing the standpoints of prosecution and defense, respectively, and I the relevant case information. From our teaching activities, especially in the US, we have noticed that the assignment of propositions is one of the areas presenting difficulty. For this reason it is timely to revisit the principles underlying the practical implementation of the Bayesian approach, notably the formulation of the propositions.

The formulation of propositions is crucial to the interpretation of a set of observations. This concept is founded strongly in the scientific method where it is vital to consider your hypotheses before acquiring the data. Forensic science is no different (except that we will speak of *propositions* instead of *hypotheses*). Note that in science *hypotheses* refer to statements that can be tested by performing a scientific experiment, whereas a *proposition* is not necessarily testable. In forensic science, a statement such as "The defendant murdered the victim" cannot be tested in a lab, so we prefer to call this a *proposition*, and to generally refer to the statements corresponding to H_p and H_d as *propositions*.

The value of the *LR* for the DNA typing results will be different for different pairs of propositions. Choosing the appropriate pair of propositions is therefore just as essential as the DNA lab work and assigning the probabilities for the numerator and denominator of the *LR* formula.

The formulation of propositions needs to follow a number of principles. First, propositions need to be mutually exclusive and exhaustive in the context of the case (i.e., one should not consider all propositions as default, but only those that are of interest to the Court). The second principle is that the propositions should help the Court address the issue of interest. Usually, the ultimate question of interest is the question of guilt, but forensic scientists will rarely be able to add value in such a context. They might however be able to help the Court discriminate between alleged activities and alleged DNA sources.

The propositions in any given case depend on the case circumstances at the time the interpretation is being carried out. Relevant case information may include, but is not limited to, the following elements:

1. The activities alleged by the prosecution and defense (e.g., the defendant having intercourse with the victim, the defendant biting the victim's breast, or the defendant giving the victim a hug),
2. The time at which these activities supposedly took place,
3. Where the activities supposedly occurred and what other activities took place at this location before and after the alleged activities (e.g., in the defendant's bed where the

defendant had had consensual sex with his girlfriend before the alleged activities with the complainant),

4. The nature of the population of alternative sources for the DNA trace (e.g., close relatives of the defendant or complainant, members of a particular ethnic group, or individuals considered unrelated to the defendant and the complainant).

When we talk about the case information, we need to emphasize that we do not consider information such as prior conviction, motive, presence of other types of evidence, or a confession as relevant forensic information: these are only the concern of the court and forensic scientists do not need, nor wish to have, this type of information.

In this note we attempt to address the challenge of helping the parties formulate an adequate pair of propositions for DNA typing results by presenting guidelines, explanations and examples on how one can handle common situations. We discuss the following topics: the hierarchy of propositions; cases with one person of interest; cases with multiple individuals of interest; cases where the number of unknown contributors under the defense proposition is not equal to one plus the number of unknown contributors under the prosecution proposition; relatives; and the transition from the sub-source level to the activity level in the hierarchy of propositions. Although science is international, many of our views reflect our experience in an adversarial environment where the defense is under no requirement to offer a proposition and often do not do so. In such a case a principle is that every accused is entitled to the best defense.

The hierarchy of propositions

Cook et al. (7) and Evett et al. (8, 9) classified propositions into four levels: from top to bottom these are offense, activity, source, and sub-source (Figure 1). Please note that forensic scientists must not address propositions, they evaluate their results given these propositions. In order to rise in the hierarchy, forensic scientists ought to add value using their expertise (i.e., forensic knowledge).

Offense level

The highest level is the offense level. The propositions at this level address the issue of whether a person of interest (POI), such as the defendant, is guilty or innocent of an offense. For example, the pair of propositions “The defendant raped the complainant” and “The defendant had consensual sex with the complainant” (Figure 1, row 1) consider whether or not the defendant is guilty of the offense ‘rape’. Here the issue of interest is not whether the defendant and the complainant had intercourse, but whether the complainant consented to this sexual act. In such a case the forensic scientist is unlikely to have information to add value. Possible information that may inform the issue of consent might be clothing tears or physical injuries such as bruising.

Activity level

One level below the offense is the activity level. The issue of interest for a pair of activity level propositions is what activity was performed. This pair is different from offense level propositions because the activity specified in the prosecution proposition does not necessarily qualify as an offense. For example, the pair of propositions “The defendant had intercourse with the complainant” and “The defendant and the complainant only engaged in foreplay” (Figure 1, row 2) are activity level propositions because they consider the activity ‘intercourse’ (versus ‘foreplay’). Alone, this activity does not qualify as an offense (e.g., ‘rape’) and thus marks a clear distinction between the activity and the offense levels. However, note that there are cases where the activity implies an offense: "The suspect fired the gun…", "The suspect

forged the signature…". In these cases activity and offense level propositions are indistinguishable.

Another point to emphasize in the example with the propositions "The defendant had intercourse with the complainant" and "The defendant and the complainant only engaged in foreplay" is that the source of the traces is not in dispute: both parties agree that the semen (source level) and the DNA (sub-source level) came from the defendant. In this case forensic scientists should help the court assess activity level propositions. That is if the evaluation of these findings requires forensic knowledge, they will assess the findings (e.g., where the trace was found) given the different alleged activities.

Source level

The source level is one level below the activity level. Source level propositions consider the source of the crime stain, that is, from whom the crime stain comes. In source level propositions, the crime stain is clearly defined as a particular body fluid, such as blood, saliva or semen, or as a particular cell type, such as epithelial cells. An example of a pair of source level propositions is "The semen recovered on the bed sheet came from the defendant" and "The semen recovered on the bed sheet came from an unknown person" (Figure 1, row 3). In addition to considering the probability of the DNA profile given each of the possible donors, one needs to consider here the probability that this DNA profile came from the trace of interest (i.e., the semen in this example) as opposed to coming from background DNA on the surface or some other source.

Sub-source level

Finally, advances in DNA typing technology have made it possible to obtain DNA profiles for latent traces (i.e., invisible to the naked eye), which means that it is no longer always possible to infer from what body fluid or cell type the DNA came. As a result, the notion of "sub-source level" propositions was proposed (9). These propositions do not imply any inference on the body fluid or cell type of the biological trace. For example, the sub-source equivalent of the pair of source level propositions stated in the previous paragraph is "The DNA recovered from the bed sheet came from the defendant" and "The DNA recovered from the bed sheet came from an unknown person" (Figure 1, row 4). These propositions are at the sub-source level because they do not imply any inference on whether the DNA comes from semen: it could just as well have come from saliva or epithelial cells.

An essential question is how to choose (or help the prosecution or defense choose) the appropriate level. We find that the notion of hierarchy of propositions is useful in the sense that it provides a framework that helps scientists think how they can help and what are the findings that can be helpful for the court. In the absence of formal propositions given by the parties, the appropriate level would thus depend on the issue of interest to the court that forensic science can help with. That is, the pair of propositions that the scientific results can help distinguish. As a guide, one could suggest, as stated in ((10) @ p. 259):

"The further down the hierarchy the scientist operates the more the responsibility for interpreting the evidence is transferred to the court or to other experts. It is therefore important that, if the assessment of the results demands forensic knowledge (e.g., factors such as transfer, persistence, presence of material for reasons unconnected to the alleged offence), forensic scientists help the court to the best of their abilities and that they explain clearly what the results mean."

Much of the time DNA analysts will be operating at the sub-source level, because they do not have enough information about the case circumstances (e.g., observations made on the crime

scene, object or individual where the crime sample was collected) to help address propositions at a higher level in the hierarchy. Also one has to consider sub-source level propositions to assess the value of the DNA profiles whatever the level is in the hierarchy. For these reasons we will use examples at the sub-source level. In the last section we will discuss two cases where the sub-source level propositions given by the prosecution and defense both accept the source of the DNA. In those cases, the parties disagree on the activity that led to the DNA's presence, and to interpret the findings, one has to consider activity level propositions.

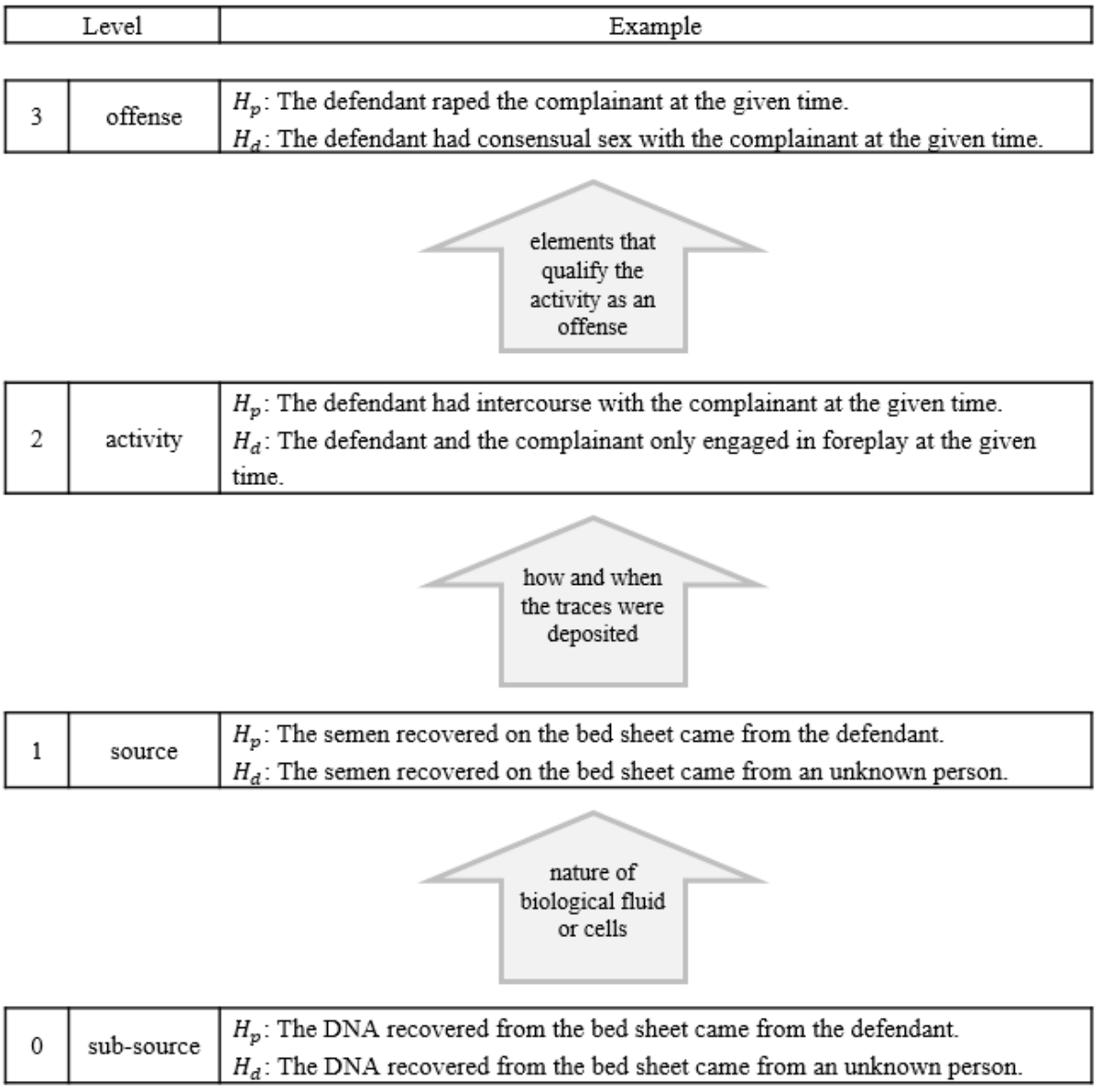

Figure 1: The four levels of the hierarchy of propositions. The arrows indicate the additional information needed to rise from one level to the next.

Should propositions be assigned before the evidence is observed?

Considerable concerns have been voiced about the biasing effect of seeing the analytical results before formulating the propositions. It is therefore highly beneficial to think through propositions before a comparison to the complainant or POI, whichever is probative.

For example, imagine that a high vaginal swab is examined and found to be a two-person mixture. The presence of the complainant would be expected and unsurprising. No harm will accrue by examining the profile of the complainant because we assume the contribution of the complainant under both the prosecution and the defense proposition. Let us imagine that the profile supports the presence of the complainant. This only confirms our preliminary assumption. The scientist would now be in a position to form the pair of competing propositions and as yet has not examined the profile of the POI.

A useful distinction is drawn between investigative and evaluative analysis (8). During the investigative phase propositions are explored that are suggested by the observations. During the evaluative phase the value of the observations is assigned with regard to a given pair of propositions.

Assigning propositions when there is one person of interest

The case of a single source DNA trace is straightforward, and we refer the reader to the sub-source pair of propositions in Figure 1 for an example. The case of mixtures is less straightforward. A typical example of a mixture with one person of interest is a two-person mixture where the prosecution's proposition is:

H_p: The DNA came from the complainant C and the person of interest POI.

The assignment of a prosecution proposition is often not a problem, but it is most often the case that the scientist does not know the position that defense will take at court. In this case the defense's proposition could be any of the following:

H_{d1}: The DNA came from C and an unknown person U
H_{d2}: The DNA came from POI and an unknown person U
H_{d3}: The DNA came from two unknown people U_1 and U_2.

To choose which of these propositions to use as an alternative, we need to consider the circumstances of the case. Relevant case information includes answers to the following questions:

1. From where was the crime stain collected?
2. If it was collected on an object, does this object belong to either C or POI?
3. Should we expect to find the DNA of either C or POI in this trace?

The first defense proposition, H_{d1}, is appropriate whenever the crime stain is taken from the complainant C, or from an object (or in a location) associated with C so that we expect to find C's DNA on it regardless of whether the prosecution's or the defense's position is true. The second defense proposition, H_{d2}, is appropriate whenever the crime stain is taken from the person of interest POI, or from an object (or in a location) associated with POI so that we expect to find POI's DNA on it regardless of whether the prosecution's or the defense's position is true.

Assigning propositions when there is ambiguity in the defense proposition

When the crime stain is recovered on an object or in a location that is associated with neither C nor POI, exonerations may be implied by any of H_{d1}…H_{d3}. It may be very difficult to

determine which proposition to use on behalf of the defense. This is one of the situations that may cause the most difficulty. In an adversarial environment the defense regularly decline to state a position and even more rarely do so in advance. The full assessment of the weight of evidence takes into account all possible defense propositions (11):

$$\frac{Pr(E|H_p,I)}{\sum_{i=1}^{3} Pr(E|H_{di})Pr(H_{di}|H_d,I)} \quad (2)$$

where $Pr(H_{di}|H_d,I)$ is the weight of the i'th defense proposition given that the defense's position is true. The terms $Pr(H_{di}|H_d,I)$ are unlikely to be available.

To avoid this development of the weight of evidence (Eq. 2), it is the widespread practice (though often unconsciously so) to adopt a "proxy" proposition that is no more than a provisional representation of the defense position. If the scientist chooses this approach, he/she should always make this clear and state a willingness to consider whatever alternative position defense may take at trial.

Further, it can be shown that for any defense proposition H_{di}:

$$LR \geq \frac{Pr(E|H_p)}{max_i Pr(E|H_{di})} \quad (3)$$

This states that the *LR* is minimized when the defense proposition is chosen that maximizes the probability of the DNA results. A suggested phrasing in the statement for this might be:

"I have considered the propositions that the DNA came from C and POI against the propositions that either C is not, or POI is not, or neither are donors to the DNA. The *LR* is at least *x* no matter which of these propositions is chosen for the defense. Should any of the information be incorrect or any further information become available (in particular if there is a change in propositions), it will be necessary for me to reconsider my interpretation."

In the case of multiple possibilities for the defense proposition the defense is unlikely to have the scientific knowledge to inform themselves. A way forward would be to spell out the consequences of each choice. Since it is rare that this will be done before the issuance of a statement we would suggest that this could be done by providing an appendix in the statement that outlined the *LR* for the different choices. These multiple *LR*s would be investigative. Once one has been selected by the defense, in all probability at trial, then this *LR* becomes evaluative.

The provision of three *LR*s for three defense propositions invites "proposition shopping" from the defense. From an adversarial standpoint this is acceptable. The defense are entitled to choose their best defense and those scientists appearing for the prosecution see it as their duty to inform the defense of scientific matters that may be in their interests.

Table 1 gives a summary of these ideas, and the four examples below show how one can apply them to casework. We start with some simple examples. It is often suggested, quite correctly, that casework is not always simple. It is certainly possible to find situations where the principles exposed by simple examples confront each other and selection is not obvious. We explore some of these in examples 3 and 4.

Table 1: The formulation of a pair of sub-source level propositions in a two-person mixture case with one person of interest POI.

"The DNA came from…"		
H_p	H_d	
"…the complainant C and the person of interest POI."	*if the crime stain was recovered from C:*	"…C and an unknown person U."
	if the crime stain was recovered from POI:	"…POI and an unknown person U."
	if the crime stain was not recovered from C, nor POI, and there is no information attributing the stain to C or POI:	"…C and an unknown person U." "…POI and an unknown person U." "…two unknown people U_1 and U_2." (If H_d is not specified by the defense, either assign an *LR* that weighs the different possibilities in its denominator (eq 2 or 3), or choose the most plausible one and clearly state the assumption made.)

Example 1. Imagine that the sample is a vaginal swab taken from the complainant, C. The profile is a two-person mixture and the complainant's DNA is supported as being present.

It seems a reasonable assumption that any DNA components matching C have indeed come from her. That assumption should be stated in the scientist's report and is rightfully open to challenge down the line. However there is no advantage to the defense to debate the presence of C: it is the presence of POI which is under question. Therefore the propositions "The DNA came from C and POI" and "The DNA came from C and an unknown person U" are perfectly reasonable for the prosecution and defense propositions, respectively.

Example 2. Next, imagine that the DNA sample was taken from a shirt belonging to the person of interest, POI. The profile is a two-person mixture and DNA from the POI is supported as being present.

Again it seems a reasonable assumption that the DNA components matching POI have indeed come from the person of interest. Therefore the propositions "The DNA came from C and POI" and the "DNA came from POI and U" are perfectly reasonable.

Example 3. Let us consider that the sample is from an object, for example a cloth, neither associated with the complainant nor the person of interest. Further let us assume that there is absolutely no other information that could guide us such as evidence that the cloth was taken from the complainant's house. We assume we are examining this cloth for some reason and the only logical one is that the prosecution suspects it is involved with the crime. Imagine that it is a two-person mixture.

It seems likely the prosecution will assert that the DNA is from C and POI. Exoneration could follow if any of the propositions below were true:

H_{d1}: The DNA came from C and an unknown person U,
H_{d2}: The DNA came from POI and an unknown person U,
H_{d3}: The DNA came from two unknown people U_1 and U_2.

The defense therefore has a right to any of $H_{d1}…H_{d3}$. A pragmatic approach to the uncertainty might be to calculate the *LR* for all three and report the lowest, and in the absence of any

emergent evidence to guide the decision that is our recommendation. The lowest will almost never be for H_{d3}. This approach is consistent with the inequality in equation 3.

Example 4. The complainant states that she was assaulted while jogging along a path and raped. She states that the assailant used a scrap of fabric with which he cleaned up. A scrap of fabric was found at the scene that is semen positive. The complainant identifies POI as her assailant. DNA profiling produces a two-person mixture. A person of interest is asked to provide a DNA sample. He makes no comment when interviewed.

The prosecution proposition is quite straightforward.

H_p: The DNA came from C and POI.

The difficulty in this case comes, as usual, with H_d. The defense appears to have three options.

H_{d1}: The DNA came from C and U (leading to LR_1),
H_{d2}: The DNA came from POI and U (leading to LR_2), and
H_{d3}: The DNA came from U_1 and U_2 (leading to LR_3).

This is an example where we can find a point where the principles become very hard to implement. It is the decision of which proposition most suits the defense but is still consistent with innocence that is difficult. It would be best, but unlikely, for the defense to choose one. In our experience the more likely response is that nothing is conceded: the defense position is simply that the complainant is wrong. This is, of course, their right.

Of the three options, we feel that LR_2 does the defense the least good. It carries with it an admission that POI was in the area. LR_1 admits at least part of the complainant's story (i.e., that the complainant's DNA is on the cloth). However this is close to admitting the already proven. At this point we have her testimony and a DNA profile supporting the statement that a cloth potentially bearing her and the assailant's DNA was deposited here. We are confident that rational scientists might split between LR_1 and LR_3. If this is so then defense attorneys might also split between these.

We find this a good place to reproduce a triangle attributable to John Buckleton that highlights the compromises present in court writing and testimony.

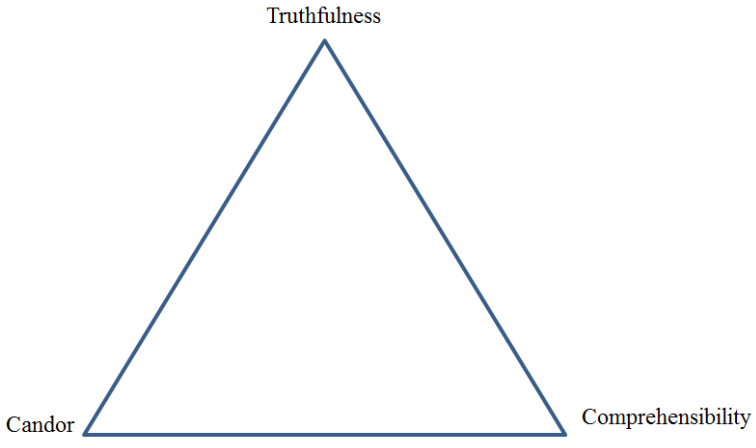

Figure 2: A triangle attributable to John Buckleton illustrating the goals and compromises made when communicating with the court. The three goals are truthfulness, candor and comprehensibility. The triangle shows that pursuing one of them requires sacrificing the others.

It suggests that the goals of any communication with the court should be truthfulness, candor and comprehensibility. However the pursuit of candor often requires the sacrifice of comprehensibility and vice versa. Example 4 highlights this tension. That full disclosure of every detail reduces comprehensibility is obvious but the recognition of this was valuable to us when considering how to phrase some matters. We find that most differences of opinion within or between laboratories arise from different decisions about these compromises. The inclusion of all three *LR*s in a statement with their associated sets of propositions will lengthen the statement and lower the chance of comprehension. Picking one over the others means the scientists are making decisions on behalf of the defense that are not straightforward.

In the case under discussion we suggest allowing the scientist's best judgment between LR_1 and LR_3. To encourage discussion with the defense we suggest that all three *LR*s and the associated propositions be placed in an appendix accepting the fact that this lengthens the report.

Again in our experience the defense will like the smaller number no matter what concessions this involves.

Assigning propositions when there are multiple individuals of interest

Another situation where setting propositions needs considerable thought is when there are multiple individuals of interest. We distinguish between two situations:

1. The prosecution considers each of the multiple individuals of interest separately as a contributor,

   or

2. The prosecution considers the multiple individuals of interest together as the contributors.

We discuss each of these below and give a summary in Table 2.

Considering the individuals of interest separately (POI_1 or POI_2)

This is the case of a two-contributor stain where the prosecution allegation is that either POI_1 or POI_2 is a contributor, so that we have two prosecution propositions:

H_{p1}: The DNA came from POI_1 and an unknown person U,
H_{p2}: The DNA came from POI_2 and an unknown person U,

and one defense proposition:

H_d: The DNA came from two unknown people U_1 and U_2.

Let H_{p1} lead to LR_1 and H_{p2} lead to LR_2. We would suggest that both of these be reported. However, in the event of both LR_1 and LR_2 being large, we suggest that the analysis should be repeated using the "POI_1 and POI_2" pairs of propositions described below.

Considering the individuals of interest together (POI_1 and POI_2)

This is the case of a two-contributor stain where the prosecution allegation is that both person of interest POI_1 and person of interest POI_2 are contributors:

H_p: The DNA came from POI_1 and POI_2.

In this case, any issue arises when we consider H_d. There are three possibilities:

H_{d1}: The DNA came from POI_2 and an unknown person U (leading to LR_3),
H_{d2}: The DNA came from POI_1 and an unknown person U (leading to LR_4), and
H_{d3}: The DNA came from two unknown people U_1 and U_2 (leading to LR_5).

We suggest that all three will be needed (but not necessarily reported). Here LR_3 measures the weight of the evidence in support of person of interest POI_1 being a contributor, given the assumption that POI_2 is the other contributor, and LR_4 the weight of the evidence in support of person of interest POI_2 being a contributor, given the assumption that POI_1 is a contributor. These are the values we would suggest might be reported as the individual weights of the evidence in support of the contribution of each person of interest if the case circumstances and the defense's position render H_{d1} and H_{d2} plausible.

However, we could imagine that in some cases the circumstances may make it impossible for POI_1 to concede the presence of POI_2 and vice versa. For example if the defense proposition is that POI_1 and POI_2 were at a football game together at the time of the crime. In such a case the propositions should be H_p and H_{d3}, leading to LR_5.

Table 2: The formulation of a pair of sub-source level propositions in a two-person mixture case with two individuals of interest POI_1 and POI_2.

<table>
<tr><td colspan="4">"The DNA came from…"</td></tr>
<tr><td></td><td>H_p</td><td>H_d</td><td></td></tr>
<tr><td rowspan="2">OR</td><td>"…the person of interest POI_1 and an unknown person U."</td><td rowspan="2">"…two unknown people U_1 and U_2."</td><td>LR_1 is a measure of the weight of evidence in support of POI_1 being one of the contributors.</td></tr>
<tr><td>"…the person of interest POI_2 and an unknown person U."</td><td>LR_2 is a measure of the weight of evidence in support of POI_2 being one of the contributors.</td></tr>
<tr><td rowspan="3">AND</td><td rowspan="3">"…the individuals of interest POI_1 and POI_2."</td><td>"…POI_2 and an unknown person U."</td><td>LR_3 is a measure of the weight of evidence in support of POI_1 being one of the contributors, assuming POI_2 is the other contributor.</td></tr>
<tr><td>"…POI_1 and an unknown person U."</td><td>LR_4 is a measure of the weight of evidence in support of POI_2 being one of the contributors, assuming POI_1 is the other contributor.</td></tr>
<tr><td>"…two unknown people U_1 and U_2."</td><td>LR_5 is a measure of the weight of evidence in support of the pair of individuals POI_1 and POI_2 being the two contributors.</td></tr>
</table>

Under either of the situations given above, it is possible that the defense nominate H_d. For example, ". . .person X and an unknown person U" could replace ". . .two unknown people U_1 and U_2".

Practical implementation
When considering "POI_1 and POI_2" situations, we have noted a tendency to use the sets:

H_{p1}: The DNA came from POI_1 and an unknown person U,
H_{d3}: The DNA came from two unknown people U_1 and U_2 (leading to LR_1),

and

H_{p2}: The DNA came from POI_2 and an unknown person U,
H_{d3}: The DNA came from two unknown people U_1 and U_2 (leading to LR_2),

instead of the set:

H_p: The DNA came from POI_1 and POI_2,
H_{d3}: The DNA came from two unknown people U_1 and U_2 (leading to LR_5).

In our opinion, the first two sets (LR_1 and LR_2) are useful for a preliminary examination in this case. Yet the prosecution proposition in this case is clear: it is that both POI_1 *and* POI_2 are donors. So this proposition must be tested prior to court by obtaining a value for LR_5.

In this type of case, we could envisage some significant risk in presenting only LR_1 and LR_2. It is possible that LR_1 and LR_2 each have a high value, and yet we obtain a low value or 0 for LR_5. This would suggest that the DNA typing results of the mixture support *either* POI_1 *or* POI_2 being a donor, but do not support POI_1 *and* POI_2 being the two donors together. This would obviously introduce questions about the validity of H_p. It is likely that this is vital information for a court and the scientist should report it.

Again, we recommend that the reported *LR* should be the one for the propositions that best describe the standpoints of prosecution and defense. We also note that the more known contributors there are, the fewer possibilities there are for the genotypes of the unknown contributor(s). This leads to two consequences: (i) fewer false inclusions of possible contributors to the mixture (i.e., the candidates not having a genotype in this restrained set of possibilities), and (ii) higher *LR* values for those candidates having a genotype in this restrained set of possibilities.

Should the number of unknowns under H_d always be one more than under H_p?

There is no need for the number of unknowns under the defense proposition to be one more than under the prosecution proposition. Consider the following examples.

Example 5. In a multiple homicide four people, D_1, …, D_4, are killed. A blood stain is recovered from the address of the person of interest. This stain appears to arise from four contributors. The prosecution proposition would be:

H_p: The stain arises from D_1, …, D_4.

It is likely that defense cannot meaningfully accept the presence of any of the deceased in the stain. Hence we would suggest:

H_d: The stain arises from U_1, …, U_4.

Clearly the difference in the number of unknowns between the two propositions is four.

Example 6. A woman, C, attends a party at a gang house which is next door to her own house. She states that she was raped by POI in her own bed the following morning. POI states that he had consensual sex with her on his bed and then more consensual sex with her in her bed the next morning. A stain is analyzed from the sheet of POI's bed. This can be explained as a mixture of three contributors. We know that A is the usual partner of POI. This would suggest the following pair of propositions:

H_p: The DNA on the sheet of POI's bed comes from POI, A, and an unknown person,
H_d: The DNA on the sheet of POI's bed comes from POI, A, and C.

In this example there is actually one fewer unknown under H_d, because the defense claims the presence of C's DNA in the mixture, whereas the prosecution denies the presence of C's DNA in the mixture.

Relatives

The Bayesian approach is suitable even if the individuals of interest or the person of interest and complainant are related. We could envisage two situations:

1. The known individuals are related and have been genotyped, or
2. One, or more, of the postulated unknown individuals is related to the genotyped individuals.

In either case the Bayesian approach is suitable and has considerable advantages.

The issue of relatedness applies only to the probability of the genotype of an unknown person. If the individuals are genotyped then the probability of their genotype is 1 no matter to whom they are related. Therefore if the known individuals are genotyped the issue of relatedness has no effect. Consider the following examples.

Example 7. The complainant, C, states that she was raped by her uncle, POI. Both are genotyped. A high vaginal swab was obtained and it produced a two-person mixture that may be explained by C and one other person.

Suitable propositions are:

H_p: The DNA came from C and POI,
H_d: The DNA came from C and an unknown person U.

The only genotype probability we will need is the probability that the unknown person has genotype G_U given the information about the genotypes of POI and C, G_{POI} and G_C: $Pr(G_U|G_{POI},G_C, H_d)$. If U is unrelated to C and POI then there is no effect of relatedness although there may be a minor effect via co-ancestry if it is postulated that U, C, and POI, although unrelated, come from the same subpopulation (12).

Example 8. The complainant, C, states that she was raped by a family friend, POI, at a gathering. Both are genotyped. A high vaginal swab was obtained and it produced a two-person mixture that may be explained by C and one other person. The defense suggests that it was not POI but his natural brother B, who was also at the gathering, who raped POI. B declined to be genotyped.

Suitable propositions are:

H_p: The DNA came from C and POI,
H_d: The DNA came from C and B.

The genotype probability we will need is the probability that B has genotype G_B given the information about the genotypes of POI and C, G_{POI} and G_C: $Pr(G_B|G_{POI},G_C, H_d)$. There are readily available formulae for this problem (5, 13, 14) and the fact of relatedness between B and POI significantly affects the calculation. However it is easily incorporated into the Bayesian framework (15).

The transition from sub-source to activity

In routine DNA analysis the propositions addressed will typically be at sub-source level. The evaluation of these results will be reported as a match probability or preferably an *LR*. In our experience this may be referred to as "the number" by prosecutor and defense attorney. This practice breaks the connection between the *LR* and the propositions, and this is regrettable. Our experience is that discussion in court then very likely evolves to activity level, yet there is no direct relationship between the *LR* at sub-source level and one at activity level.

In practice the most likely line of questioning concedes the source of the DNA and asks about mechanisms of transfer and persistence. If both parties agree that the DNA came from the accused then the *LR* at sub-source level is no longer relevant. In this case, it is the activity that caused the transfer of the DNA that makes the defense proposition different from the prosecution proposition (we refer the reader to the section *The hierarchy of propositions* for an explanation on activity level propositions). The relevant information in *E* is no longer the trace's intrinsic characteristics (i.e., the alleles present in the DNA trace) but the trace's extrinsic characteristics, such as the quantity, quality and location of the recovered trace material:

1. The quantity of material provides information about the type of activity that caused the transfer. It is useful for differentiating between two activities where one activity transfers a large amount of DNA and the other a small amount. If there is information on the conditions and actions to which the trace was subject, then it may also provide information on the amount of time that has elapsed since the transfer.
2. The quality, for example a fresh stain versus an old stain, may help the court decide on issues such as the time elapsed since the transfer and/or the conditions it was subject to after the transfer. It is useful for differentiating between two activities performed at different times and separated by a large lapse of time.
3. The location provides information about the type of activity performed and can help differentiate between activities that would leave a trace at different locations.

It is important that analysts are prepared for the transition from sub-source to activity level. We would suggest that useful tools are an understanding of the hierarchy of propositions as this allows the analyst to recognize the shift in propositions. A comprehensive study of transfer and persistence of DNA is vital for reporting scientists in order to help address activity level issues. The following two examples illustrate how consideration of a pair of sub-source level propositions can progress to a pair of activity level propositions when the issue of interest to the court becomes one about what activity was performed.

Example 9. The complainant states that she was raped at her home by POI after a drunken party. No statement is available from the defense at the time of analysis. The reporting officer proceeds with the propositions:

$H_{p\ \text{sub-source}}$: The DNA came from POI (or C and POI if it is a mixture),
$H_{d\ \text{sub-source}}$: The DNA came from U (or C and U if it is a mixture).

POI matches the DNA from the complainant's cervical swab. The reporting officer assigns and reports an $LR_{\text{sub-source}}=10^9$.

At trial, the defense said that POI had previously masturbated and cleaned up with the last of the toilet paper on the roll and threw it in a trash can next to the toilet. Defense suggest that C used the toilet, realized there was no toilet paper on the roll, so she must have reached into the trash and used that toilet paper. This now suggests that activity level propositions are being considered. In our experience this is a very common transition. That is, from a report at sub-source to questions at trial at activity level, the pair of propositions becomes:

$H_{p\ \text{activity}}$: POI had intercourse with C,
$H_{d\ \text{activity}}$: POI masturbated, cleaned up with the toilet paper, and C wiped using the toilet paper.

This requires a different body of knowledge. What is required here is an understanding of the probability of transfer of the observed amount of DNA by the mechanisms described: intercourse and transfer from toilet paper. The initial $LR_{\text{sub-source}}$ is now redundant. Both parties are accepting that the DNA came from POI (or POI and C). Now we require:

$$LR_{\text{activity}} = \frac{Pr(E|H_{p\ \text{activity}}, I)}{Pr(E|H_{d\ \text{activity}}, I)}.$$

It is probably quite difficult to assign numerical probabilities to the terms in the numerator and denominator; however it might be quite reasonable to state that the probability of a strong signal in the DNA profile and the presence of sperm in a cervical swab is much higher given intercourse than given contamination via toilet paper.

Example 10. The complainant states that she was digitally penetrated by an assailant whom she identifies as POI after a night of bar-hopping and clubbing. Subsequently POI put her into a taxi to send her home. POI states that he went back to the clubs. A DNA mixture is obtained from under the fingernails of POI. There are two minor alleles that match POI and a strong major component that may be explained by C. The reporting officer assigns an $LR_{\text{sub-source}} = 10^9$ for the pair of propositions:

$H_{p\ \text{sub-source}}$: The DNA came from POI and C,
$H_{d\ \text{sub-source}}$: The DNA came from POI and U.

At trial the defense claim that the DNA resulted from transfer whilst POI walked C out to the taxi with his hands on her bare arms and shoulders to keep her from falling. This suggests the following propositions at the activity level:

$H_{p\ \text{activity}}$: POI digitally penetrated C, and held C's bare arms and shoulders with his hands while walking C to a taxi,
$H_{d\ \text{activity}}$: POI held C's bare arms and shoulders with his hands while walking C to a taxi.

Again we require $\frac{\Pr(E|H_{p\ activity}, I)}{\Pr(E|H_{d\ activity}, I)}$. As in Example 9 above, it may be difficult to quantify this *LR*. Relevant considerations would include the observation that the DNA was a two person mixture that may be explained by C and POI. There was no other DNA present. C is the very strong major. Knowledge of transfer and persistence of DNA by casual contact and by digital penetration is the key to assigning the probabilities. We could be confident that this evidence is more probable if digital penetration had occurred than if casual contact alone had occurred.

Please note that even when stating a non-numerical *LR* such as this, it is important to state the probability of the findings given the two propositions, rather than the other way around. The statement "It is more probable that digital penetration occurred rather than casual contact alone given these DNA results" is known as the fallacy of the transposed conditional (16) and should be avoided.
We add a cautionary note about uninformed speculation. It is important when a non-numerical assignment of probability is given that we ensure that it has some sound scientific foundation. The risks of this have been elegantly debated by Gill (17).

Conclusions

In this note we provide a set of guidelines, explanations and examples for helping to formulate pair of propositions required for assigning an *LR* for DNA typing results. We have shown how the choice of these propositions depends on the case information and the allegations of each of the parties. This dependence is unavoidable for the forensic scientist to be able to accomplish his/her duty of presenting what the DNA results mean with regard to the issue of interest to the court.

We realize that choosing the most adequate pair of propositions can be challenging for the forensic scientist because it requires many pieces of information, and the scientist may not have all of these pieces. We suggest (as we have done in some of the examples in this note) that a reasonable way for handling these cases is to: (i) open a discussion with the defense, likely to actually take place in court, by assigning an *LR* given plausible propositions and underline one's willingness to consider other propositions should the ones considered be inadequate, or (ii) state, in the body of the report, the smallest of these *LR* values if based on the available information where multiple pairs of propositions are plausible. The most important factor is for the scientist to remain transparent and clear about what propositions and assumptions lie behind the presented *LR* value(s).

Finally, we also recognize that real cases can involve more complex issues than those described in this note (e.g., being uncertain about the number of contributors, not knowing how to combine the results of several traces, or not knowing how additional variables, such as probabilities of transfer and persistence, will affect the value of the *LR*). We advocate considerable restraint in these areas. There are certain profiles or situations where there is so much uncertainty that attempting an interpretation may be significantly hazardous. For example mixed profiles where the number of contributors is greater than four should, at least at time of writing, be considered very complex. The need to formalize one's propositions to create an *LR* may act as a beneficial restraint. If it is simply not possible to form propositions then maybe the situation is beyond interpretation.
For issues on more advanced topics, we can recommend education (18) and tools such as Bayesian networks as a slightly more advanced topic (18-24).

Acknowledgements

This work was supported in part by grant number P2LAP3-148445 from the Swiss National Science Foundation and grant number 2011-DN-BX-K541 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice or the Department of the Army, and Department of Defense. We sincerely thank Catherine McGovern, Stuart Cooper, Jason LeBlanc, and two anonymous referees for comments that have greatly improved this paper.

References

1. Scientific Working Group on DNA Analysis Methods (SWGDAM). SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. 2010 [updated 2010 14 January 2010; cited 1 June 2018]; Available from: http://www.forensicdna.com/assets/swgdam_2010.pdf.
2. Association of Forensic Science Providers. Standards for the formulation of evaluative forensic science expert opinion. Sci Justice. 2009;49(3):161-4.
3. Gill P, Gusmão L, Haned H, Mayr WR, Morling N, Parson W, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. Forensic Sci Int Genet. 2012;6(6):679-88.
4. Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Sci Int. 2006;160:90-101.
5. Buckleton JS, Triggs CM, Walsh SJ. Forensic DNA Evidence Interpretation. Boca Raton, Florida: CRC Press, 2004.
6. Bille T, Bright J-A, Buckleton J. Application of Random Match Probability Calculations to Mixed STR Profiles. J Forensic Sci. 2013;58(2):474-85.
7. Cook R, Evett IW, Jackson G, Jones PP, Lambert JA. A model for case assessment and interpretation. Sci Justice. 1998;38(3):151-6.
8. Evett IW, Jackson G, Lambert JA. More on the hierarchy of propositions: exploring the distinction between explanations and propositions. Sci Justice. 2000;40(1):3-10.
9. Evett IW, Gill PD, Jackson G, Whitaker J, Champod C. Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. J Forensic Sci. 2002;47(3):520 - 30.
10. Buckleton JS, Bright J-A, Taylor D, Evett IW, Hicks T, Jackson G, et al. Helping formulate propositions in forensic DNA analysis. Science and Justice. 2014;54(4):258-61.
11. Buckleton J, Triggs C, Champod C. An extended likelihood ratio framework for interpreting evidence. Sci Justice. 2006;46:69-78.
12. Balding DJ, Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int. 1994;64:125-40.

13. NRC_II. National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence. Washington, D.C.: National Academy Press, 1996.

14. Evett IW, Weir BS. Interpreting DNA Evidence – Statistical Genetics for Forensic Scientists. Sunderland: Sinauer Associates, Inc., 1998.

15. Evett IW. Evaluating DNA Profiles in a Case Where the Defence is "it was my Brother". J Forensic Sci Soc. 1992;32(1):5-14.

16. Thompson WC, Schumann EL. Interpretation of statistical evidence in criminal trials - The prosecutors fallacy and the defence attorneys fallacy. Law Hum Behav. 1987;11:167-87.

17. Gill P. Misleading DNA Evidence. San Diego: Academic Press, 2014.

18. Biedermann A, Hicks T, Voisard R, Taroni F, Champod C, Aitken C, et al. E-learning initiatives in forensic interpretation: report on experiences from current projects and outlook. Forensic Sci Int. 2013;230:2-7.

19. Gittelson S, Biedermann A, Bozza S, Taroni F. Modeling the forensic two-trace problem with Bayesian networks. Artificial Intelligence and the Law. 2013;21(2):221-52.

20. Biedermann A, Bozza S, Konis K, Taroni F. Inference about the number of contributors to a DNA mixture: Comparative analyses of a Bayesian network approach and the maximum allele count method. Forensic Sci Int Genet. 2012;6(6):689-96.

21. Gittelson S, Biedermann A, Bozza S, Taroni F. Bayesian Networks and the Value of the Evidence for the Forensic Two-Trace Transfer Problem. Journal of Forensic Sciences. 2012;57(5):1199-216.

22. Taroni F, Aitken C, Garbolino P, Biedermann A. Bayesian Networks and Probabilistic Inference in Forensic Science. Chichester: John Wiley & Sons, Ltd., 2006.

23. Garbolino P, Taroni F. Evaluation of scientific evidence using Bayesian networks. Forensic Sci Int. 2002;125:149-55.

24. Mortera J. Analysis of DNA mixtures using Bayesian networks. In: Green P, Hjort NL, Richardson S, editors. Highly Structured Stochastic Systems. Oxford: Oxford University Press; 2002.