TITLE:

The importance of considering common sources of unknown DNA when evaluating findings given activity level propositions

AUTHORS:

Duncan Taylor[1, 2], Luke Volgin[2], Bas Kokshoorn[3], Christophe Champod[4]

1. College of Science and Engineering, Flinders University, GPO Box 2100 Adelaide SA, Australia 5001
2. Forensic Science SA, GPO Box 2790, Adelaide, South Australia 5001
3. Netherlands Forensic Institute, P.O. Box 24044 NL-2490AA, The Hague, The Netherlands
4. Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice, University of Lausanne, CH-1015 Lausanne-Dorigny, Switzerland

CORRESPONDING AUTHOR:

Duncan A. Taylor, PhD
Forensic Science SA
GPO Box 2790
Adelaide
South Australia 5001
Phone: +61-8 8226 7700
Fax: +61-8 8226 7777
Email: Duncan.Taylor@sa.gov.au

KEY WORDS:

Background DNA; activity level; evidence evaluation; Bayesian networks.

HIGHLIGHTS:

- We consider the evaluation of findings from multiple similar items
- Bayesian Networks are used to construct evaluative frameworks
- We demonstrate the importance of considering multiple sources of unknown DNA
- We apply the theory presented to a real case scenario from South Australia

ABSTRACT:
Evaluating forensic biological evidence considering activity level propositions is becoming more prominent around the world. In such evaluations it is common to combine results from multiple items associated with the alleged activities. The results from these items may not be conditionally independent, depending on the mechanism of cell/DNA transfer being considered and it is important that the evaluation takes these dependencies into account. Part of this consideration is to incorporate our understanding of prevalent DNA and of background DNA on objects and people, and how activities can lead to common sources of unknown DNA being deposited on items. We demonstrate a framework for evaluation of DNA evidence in such a scenario using Object-Oriented Bayesian Networks and apply it to a motivating case from South Australia.

INTRODUCTION:

Evaluating forensic DNA results considering activity level propositions is an important task for the forensic scientist. It puts the DNA results in a case context that is not achieved with the evaluation of the same results given (sub-) source level propositions. There are guidelines and publications recommending the evaluation of results given activity level propositions [1, 2]. There are a number of published examples of evaluations using activity level propositions available [2-6]. Depending on the case circumstances, the propositions of prosecution and defence may dispute the actor involved in the alleged crime, or the activity of the alleged crime [7]. If the actor is in dispute, then it is common to consider an alternate offender (often designated as AO), who is currently unidentified and who may have donated DNA that will present itself as originating from an unknown individual in the context of the forensic work being conducted. Another source of unknown DNA is the background DNA present almost ubiquitously on all items. The level to which the presence or absence of unknown DNA supports one proposition over the other then becomes a balance between the probabilities of the results given a DNA transfer from an AO compared to the presence of background DNA. However, as well as the fact that there is a presence of DNA from an unknown source on multiple items, there is additional information that can assist evaluations when considering if the same unknown contributor is present on these items.

One scenario that has not been explored extensively is how an evaluation should proceed when there are multiple similar items in a case. In this case there are two aspects to consider that will affect the evaluation, and the results of the evaluation:

1) The choice of whether to treat transfers to the similar items as one or multiple events.
2) The probability of a common unknown contributor having donated DNA to the samples.

The first point is related to the issue commonly referred to as the ‘two-trace problem’, which was originally discussed by Evett [8], and considers when there are two or more stains and two offenders. This was later extended to generality with respect to the number and type of traces and the number of offenders by Triggs et al. [9] and tackled using Bayesian networks by Gittelson et al. [10]. A general explanation on the combination of dependant pieces of evidence was given by Juchli et al. [11]. We draw upon the concepts of Gittelson et al. [10] and Juchli

et al. [11] to show how propositions can dictate whether DNA transfer to multiple items are considered one, or multiple events, and how this impacts the evaluation.

We also demonstrate the importance of considering the various pathways that unknown DNA can be deposited on to similar items, particularly if it is being suggested that they have been handled in a similar way, and by the same person (whether that be the defendant or an alternate offender). The classic treatment of common unknowns is to consider them as occurring either because they have come from a single unknown offender, or because they come from different individuals (as background), whose DNA profiles match. The probability associated with matching background DNA profile is typically set quite low, due to the discrimination power of DNA profiling systems. The value can be assigned based on a match probability if the unknowns are able to be interpreted, or could be based on mixture to mixture comparisons [12, 13] if the unknowns are not resolvable. When common unknown donors are found to be present, the evaluation will show that the findings provide strong support for the presence of the alternate offender, as the probability of transfer, persistence and recovery of the DNA of that donor is much higher than matching profiles from different sources present as background. However, given our knowledge of the level of background DNA on items from being in proximity to a person (such as in their home [14], car [15] or workplace [16]), or the presence of an unknown person's DNA on the hands of the person who touched the item (e.g. such as a cohabitant [14]) there are other explanations for the presence of common unknowns on multiple similar items. It becomes quite important to consider these alternate routes for common unknowns within an evaluation in order to obtain a sensible result.

Providing even more power to help to discriminate between propositions is to have the reference samples of people associated with the persons of interest in a case. Having profiles from these secondary associated individuals can eliminate, or confirm, certain common donor DNA transfer mechanisms, and is preferable to dealing with the uncertainty statistically. While we focus on a case example that has biological evidence, there are natural extensions of this thinking that can be applied to combine evidence across disciplines. We point the reader to the recent work of de Koeijer et al. [17] for discussions in this area.

We work through the DNA results for the case R v QUIST heard in South Australia in 2016. The defendant was convicted, but this conviction was appealed and overturned in 2017 [18] (not on the basis of DNA evidence, but in the way the trial judge instructed the jury in various non-scientific matters). At the time of this paper being written a retrial has not occurred but is scheduled. The goal of this work is to demonstrate a Bayesian Network (BN) based method for evaluating DNA results from multiple similar items, and in particular we show the importance to the evaluation of considering difference sources of common unknown profiles on the similar items.

We detail the case circumstances and forensic work in sections 2 and 3, and then elaborate on the different ways that the DNA results could be interpreted or evaluated in section 4.

2.0 CASE SCENARIO:

2.1 - Background information

At 5:15pm on the 23rd of December 2013 a fire started in the disabled toilets of a shopping centre in Parafield Gardens in South Australia. The fire had been started by igniting an open plastic bottle filled with petrol. The fire was extinguished, and during the processing of the scene (at approximately 7pm), six additional bottles (all filled with petrol) were found hidden in the ceiling space above the toilet. These bottles were taken out of the toilet ceiling (without touching the lid, and wearing appropriate protective clothing to minimise contamination), out of the toilet block and placed on the ground where the lids were immediately swabbed. There was no fingermark detection work carried out on the body or lid of the bottles. The defendant in this matter was seen around the area at 5:15pm and leaving shortly after with burns to her body.

2.2 - DNA results

In this case there are seven plastic soft drink bottles involved. One bottle was set alight and was considered not suitable for DNA sampling. Six bottles were from the ceiling space and not burnt. South Australian Police took swabs of the lids of these six bottles and submitted them to Forensic Science SA. Of the six swabs of the lids of the hidden bottles, five of these were accepted for DNA profiling (it is unclear why one was not accepted). The results of DNA testing were that three of the five bottles had no DNA detected (and so did not proceed to DNA profiling) and the other two had approximately 0.8 and 0.7ng of DNA detected. The first of these (possessing 0.8 ng) yielded a three person mixture, that using STRmix™ V2.6 [19], were in proportions 82%, 13% and 6%. A DNA profile corresponding to the DNA profile of the defendant was observed in the major component (with the proportion of 82%) and a likelihood ratio (*LR*) was calculated using the following sub-source level propositions:

- Hp: The DNA originated from the defendant and two unknowns
- Hd: The DNA originated from three unknowns

The *LR* being ~ 4.8 billion in support of the defendant being a DNA donor to the sample rather than not.

The second sample (possessing 0.7 ng of DNA) yielded a two-person mixture, that using STRmix™ V2.6, were in proportions 94% and 6%. A DNA profile corresponding to the DNA profile of the defendant was observed in the major component (with the proportion of 94%) and an *LR* was calculated using propositions:

- Hp: The DNA originated from the defendant and an unknown
- Hd: The DNA originated from two unknowns

The *LR* being ~ 1.3 billion in support of the defendant being a DNA donor to the sample rather than not.

A mixture-to-mixture comparison was carried out (as per the method in [12, 13]) and an *LR* calculated considering the propositions:

- H1: There is a common contributor to the minor components of the mixtures
- H2: There are no common contributors to the minor components of the mixtures

With a $LR \sim 50$ in support of H2 compared to H1. Therefore, the DNA observations on the minor components are more likely if different unknown individuals contributed.

Due to a combination of the heat from the fire, and the very public nature of the crime scene, no other DNA samples that were taken were deemed suitable for analysis.

2.3 - The prosecution scenario

The prosecution alleges that the defendant filled the six hidden plastic bottles with petrol and placed them in the ceiling space of the disabled toilet, and then set fire to a seventh open bottle of petrol on the toilet floor.

2.4 - The defence scenario

The defence alleges that someone other than the defendant (an alternate offender, AO) filled the six hidden bottles with petrol and placed them in the ceiling space, and then lit the seventh open bottle on the floor of the toilet. The fire died down due to lack of oxygen, but then reignited (the 'flashback' effect) when the defendant opened the toilet door, burning her. She then left the scene, and due to her injuries, she shed cellular material in the path she walked through the shopping centre. When the Police recovered the bottles, and walked them out of the bathroom to the point where they were sampled, the action of walking through the same space that had been walked through by the defendant, 90 minutes earlier, 'reinvigorated' the cellular material (presumably meaning that the cellular material was stirred up into the air). The cellular material in the air then settled on the bottle lids.

3.0 THE TRIALS AND APPEAL

In the trials held in 2016 and 2019 there were multiple facets of evidence adduced. These included the DNA evidence, and also evidence from fire experts, chemistry experts (to assist in identifying the contents of the bottles), CCTV footage, and eyewitnesses. We will only concentrate on the DNA evidence component of the case.

During the original trial in 2016, the defence scenario was put to the DNA expert. The expert replied by stating that:

'*…I couldn't exclude it as a possibility, but if there was a time delay in between, then I would lean towards it being an unlikely way for DNA to transfer*'

She also brought up the point that if shed DNA being stirred up into the air and depositing on objects was a good explanation for the transfer then it may have been expected to find highly complex mixtures on the bottle lids (due to the very public nature of the shopping centre where they were swabbed). Finally, it was brought up by the prosecution that if the 'air DNA' scenario was a good explanation for DNA transfer, then we might expect to see a similar result on all the bottle lids. Note that this comment goes directly to the heart of whether we consider the possibility of transfer to the bottle lids as independent events or a single event.

In the 2017 appeal ruling [18] the aspect of the testimony of original DNA expert witness obviously weighed heavily in the minds of the Judges. Evidence by the fact that the description of the defence scenario is given in a section of the ruling entitled:

'A highly far-fetched theory of innocent indirect DNA transfer'

In this paper, we would like to go beyond the intuitively appealing argument (as made by the Judge) suggesting that the findings support the allegations of the prosecution and that they would be unlikely under the activities envisaged by the defence. The evaluation is better placed to be made in a forensic context, rather than as an intuitive opinion. Also, we will show that the complexity of the evaluation renders difficult for the forensic scientist to offer on the spot, at trial, a fully articulated response. These cases need to be fully assessed before trial. Object-oriented Bayesian networks (OOBN) offer the flexibility to handle forensic evaluation problem with multiple pathways of DNA transfer and multiple items. The literature though is often limited to one item or a limited set of results. In the case above, we could be tempted to ignore the three bottles with the low quantity of DNA and concentrate the evaluation on the two bottles that provided DNA results that have found correspondence with the defendant. In doing so, we would ignore part of the results, whereas these results, on the three other bottles, were considered by the reporting officer when questioned in court. We aim at constructing the inference engine in the form an OOBN that could tackle all the forensic DNA results. It will allow us to assess if the defence scenario was that fanciful in the sense that the results could be hardly expected under that view, but more expected under the prosecution view.

We develop an evaluation framework that shows how findings such as those in this case can be considered together, and the importance of considering different sources of unknown DNA, and whether they are from a common source. While we calculate an *LR* under different assumptions by assigning values to the probabilities of events occurring that are important to the evaluation, our goal is not to assign an *LR* that would be suitable for this case specifically (indeed we assign probabilities in some instances with no informative data), but rather to show the mechanism by which various aspects of the evaluation can be incorporated.

4.0 EVALUATION:

Given the case information and the scenario given by prosecution and defence, we consider the following activity level propositions in the evaluation of the findings in the case:

- Hp: The defendant filled the soft drink bottles with petrol and placed them in the ceiling space of the public toilet
- Hd: An unknown offender filled the soft drink bottles with petrol and placed them in the ceiling space of the public toilet

The additional information we use is:

- The defendant was in the toilet when the fire was lit. She then moved through the shopping centre to exit the building.
- The bottles were removed from the ceiling space by Police 90 minutes after the incident and taken out into the shopping centre area (where the defendant had earlier walked through) and the bottle lids were sampled for DNA.

The assumptions we make in the evaluation of the findings in this case are:

- The same person who filled the bottles with petrol, also placed them into the ceiling space of the public toilet. This affects the area that we might expect DNA to be transferred to i.e. to the caps of the bottles from the person who filled them (where the samples were taken from) or to the body of the bottles from the person who placed them in the ceiling
- That the offender (whether this was the defendant or another person) did not wear gloves when handling the bottles
- That when the bottles were taken out of the toilet for swabbing that if DNA has been invigorated into the air, that this is a single event, and not one invigoration event per bottle given the narrow window of time in which this occurred
- That due to the fact that there were no issues reported with quality controls within the laboratory we consider the possibility of contamination to be a separate event for each bottle (i.e. as opposed to a reagent contaminated with the defendant's DNA, which would be a single event). We also apply this same reasoning to extend to the work done by the Police when sampling the bottles
- That the defendant is a donor of DNA to two bottles
- That the order of DNA bottle handling by the offender has not had an effect on the DNA transferred to the bottles
- That there are no common unknown DNA donors to the two bottles (we will however, explore the impact of this assumption on the *LR* in section 4.5)
- Note that we do not model persistence in our evaluation and so knowledge of the timeframe is not strictly necessary here. However, by not considering persistence the model assumes the DNA has persisted in the same state as the initial transfer. In making this assumption, we have used the knowledge of a relatively short timeframe between offence and sampling (90 mins).

We explore the consequences of common unknowns. We will also treat the DNA as being present or absent on an object, rather than dealing with DNA amounts. This choice is, again, made simply to keep the OOBN complexity to a minimum, and allowing focus on the main point of the paper, but note that systems such as that in Taylor et al. [20] could be used. We construct OOBNs using software HUGIN [21] and follow the OOBN construction method of Taylor et al. [3].

Throughout our paper we consider that background DNA is DNA of unknown origin on an item that is not being explicitly modelled as coming from a specific person. By "modelled", we mean that given the sampling location, and the framework of circumstances we account for it presence or absence in the context of some defined event (such as a transfer or prevalence). For example, on AO hands we expect to find AO's DNA, but this is not considered background as we are modelling it as coming from AO (using its expected prevalence). Other DNA of unknown origin on AO hands however is considered background as it is not modelled as coming from a designated person. Hence, we consider unknown DNA to be any DNA (background or modelled) which cannot be accounted for by one of the reference DNA profiles.

In the example of DNA on AO hands, both AO DNA and background DNA on AO hands are considered unknown DNA. As is evident from this example there is some overlap between background and unknown DNA.

4.1 – Transfer to bottles, one event or many?

In order to evaluate the evidence in any case where transfer to multiple items may have occurred, the level of dependence existing between them must be considered. One possible course of action is to employ a simplification to the model so that all objects are considered as one, and transfer is considered to have occurred to the meta-object if it has occurred to any of the objects. This is a similar suggestion to one of the steps that was suggested in Taylor [3], and could be used if the items have a very close relationship. For example, it could be that the handle of a knife was divided into ten parts, which were swabbed separately. These ten swabs could be considered together (just as though a single swab had been used to sample the entire handle).

Such a simplifying assumption may not always be appropriate. In the motivating case the prosecution alleges that the defendant has placed the bottles in the ceiling cavity and hence had contact with each separately (in fact it is of no lesser consequence if the defendant has only placed two of the five bottles in the ceiling). If the defence proposition stipulated that the defendant had never had contact with the bottles, and never been in the area where the bottles were found, we could consider that the only possible mechanism for the presence of DNA matching that of the defendant is if an alternate offender possessed a matching DNA profile (ignoring the possibility of laboratory contamination, or error). In this case the activity level *LR* would take the value of the inverse of obtaining chance matching profile (i.e. the same numerical value as the sub-source *LR*) and the combination of presence or absence of the DNA matching the defendant on combination of bottles will not affect the *LR* as long as a profile matching the defendant is on at least one bottle i.e. the *LR* obtained if one bottle had a profile matching the defendant is the same as if all had matching profiles. In this instance the bottles could be considered as one meta-object.

If, however the defence proposition stipulated that the presence of DNA matching the defendant on any bottle has arisen from contamination in the laboratory, then we may consider each bottle with the presence of the defendant's DNA as an independent contamination event. In this case the specific combination of presence or absence of DNA of the accused on the bottles will have an effect on the *LR* i.e. the *LR* obtained if one bottle had a DNA profile matching the defendant would be much lower than if all had matching profiles. In the contamination scenario we would also have to consider that an alternate offender would have had to place the bottles in the ceiling and so the presence or absence of a common unknown profile would also become important.

In the motivating case the defence scenario is that the bottles were placed in an area where the defendant's DNA was swirled around in the air, which then settled on the bottles. We do not consider that there have been five separate instances of the defendant's DNA being swirled

into the air and landing on one bottle. In this scenario the results of all bottles are linked (or have dependence with each other). Further, we must consider that even though the transfer of DNA is considered one event in this instance, due to slight variations in transfer and recovery of the traces, it is still possible that DNA matching the defendant could be recovered from some bottles and not from others (indeed this is the situation we have in the motivating case). To evaluate the DNA findings in the motivating case we consider a single event in which the defendant's DNA has been swirled back into the air around the bottles. Then the probability that the defendant's DNA has settled on each bottle (given it is swirling around in the air) can be assigned considering the event for each bottle as conditionally independent. Note here we are considering an 'event' as the movement of DNA, i.e. from the ground into the air, or from the air onto a bottle.

As with the scenario previously put forward that considered laboratory contamination as the source of the defendant's DNA on the bottles, the *LR* obtained from the air swirling scenario will be sensitive to the number of bottles that have the defendant's DNA and/or unknown DNA as there is both a parent factor (swirling), and independent settling factors that will be assigned probabilities that are different to the probability of DNA transfer from handling.

4.3 – Creating class networks

In Figure 1 we show the Object-Oriented BN (OOBN) that has been developed to evaluate the evidence in this case. There are five parts, in three layers of class networks. At the highest layer is the main BN (Part A of Fig 1), which is where the user would interact with the BN to instantiate information about the case and get the *LR*. In the main BN, the structure has been added to automatically calculate the *LR* and its inverse, purely for convenience, as described in [3, 22]. In the main network there are DNA profile results nodes for each of the bottles (nodes 20 – 24), but then there is also a series of nodes dedicated to whether common unknown contributors have been observed (nodes 10 – 19). There are a couple of ways that this could have been carried out, one would be to have a single node that considered all possibilities of common unknown configurations, i.e. considering the unknowns on bottles B1 to B5, the node would have states:

- B1≠B2≠ B3≠ B4≠B5
- B1=B2, B1≠ B3≠ B4≠B5
- …
- B1=B2, B3= B4, B1≠ B3≠B5
- …
- B1=B2=B3=B4=B5

The alternate possibility, and the one used in our OOBN construction, is to have one node per possible pairing, each with states of 'Y' or 'N'. We chose the latter option due to the number of options and resulting BN table size required for the first configuration. Because the invigoration of the cells into the air is assumed to be one event, it sits up at the main, outermost BN layer, as do the activity nodes (in blue) and the root nodes (in grey) relating to the presence of background DNA on the hands of the defendant (D) or alternate offender (AO). We explain later the reason for including the presence of unknown background DNA on the hands of an

alternate offender, who themselves also has an unknown profile. Finally, root nodes for the probability of profiles of the AO (or an unknown on the hands of the AO) matching the defendant are at the highest level of the BN as there are global values that apply to all possible matching scenarios.

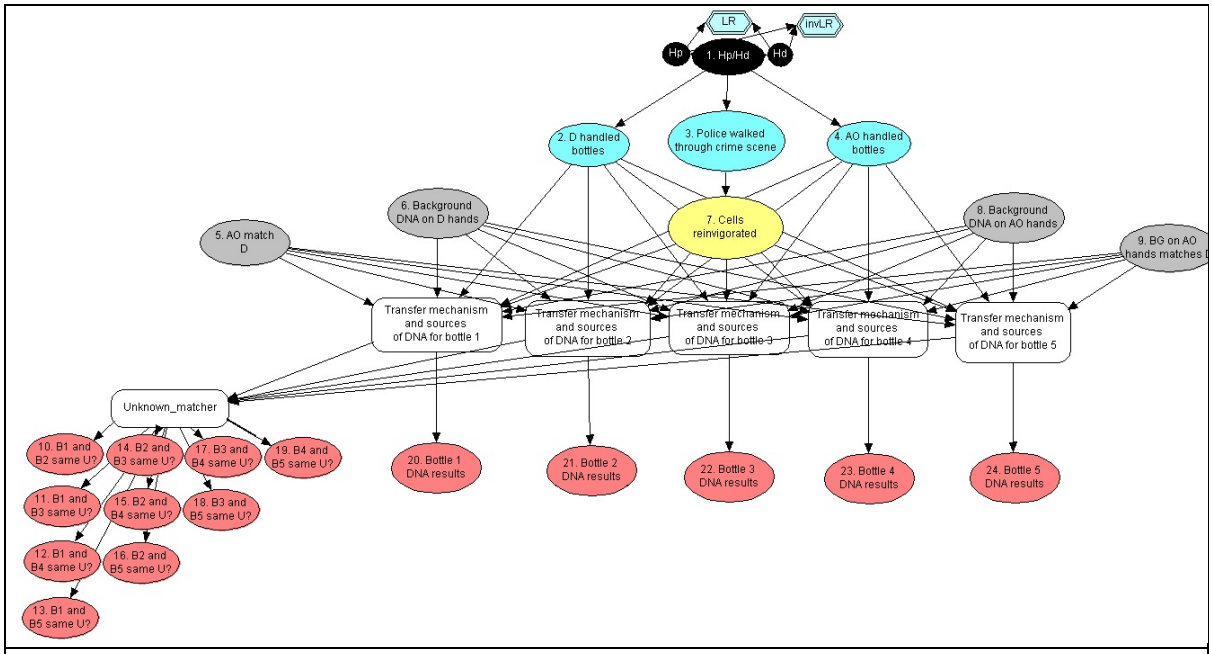
Within the main BN are class objects that relate to each bottle, and also a class network that relates to matching unknowns. The Bottle class BN (Part B of Fig 1) possesses nodes that consider the various DNA transfers that could occur from people that may have handled the bottles, the background DNA on the bottles, the potential for laboratory contamination[1], and the potential for various sources of DNA to match key actors in the evaluation. The Bottle BN also considers the probability of DNA being recovered, as the various bottles samples are taken as separate extracts and may have different extraction efficiencies. The outputs of the Bottle class BN are the DNA results from that bottle (which are then carried back out to the main BN layer so that the user need only interact with the BN on that level) and the presence of specific sources of unknown DNA that will then be used in the unknown DNA matcher. Note that the Bottle class BN also have DNA profile matcher class BNs (Part D of Fig 1) that, at the third layer of class network, deals with the possibility of by-chance matching profiles. As the DNA profile matcher has the matching probability as an input, these must be passed in from the Bottle class network. In some instances, these come from inputs, taking values from the outermost layer of the BN, but in the case of background DNA we pass in separate probabilities.

The main BN class network also possesses an unknown matcher class BN (Part C of Fig 1) that takes the unknown DNA output nodes from the bottle class BN and compares pairs of these for each possible bottle pairing, within a Single-bottle Unknown Matcher class BN (Part E of Fig 1). At this third layer of class BN in the Single-bottle Unknown Matcher the possibility of by-chance matching background DNA is considered as a root node (node 45). This root node could be set up as an input and passed in a value from the unknown matcher class BN for each bottle, if different values were desired (perhaps if relatively neutral *LRs* had been obtained from mixture to mixture matching, e.g. see [12]), however we have made the simplifying architectural decision to use a single value for all unknowns across all bottles.

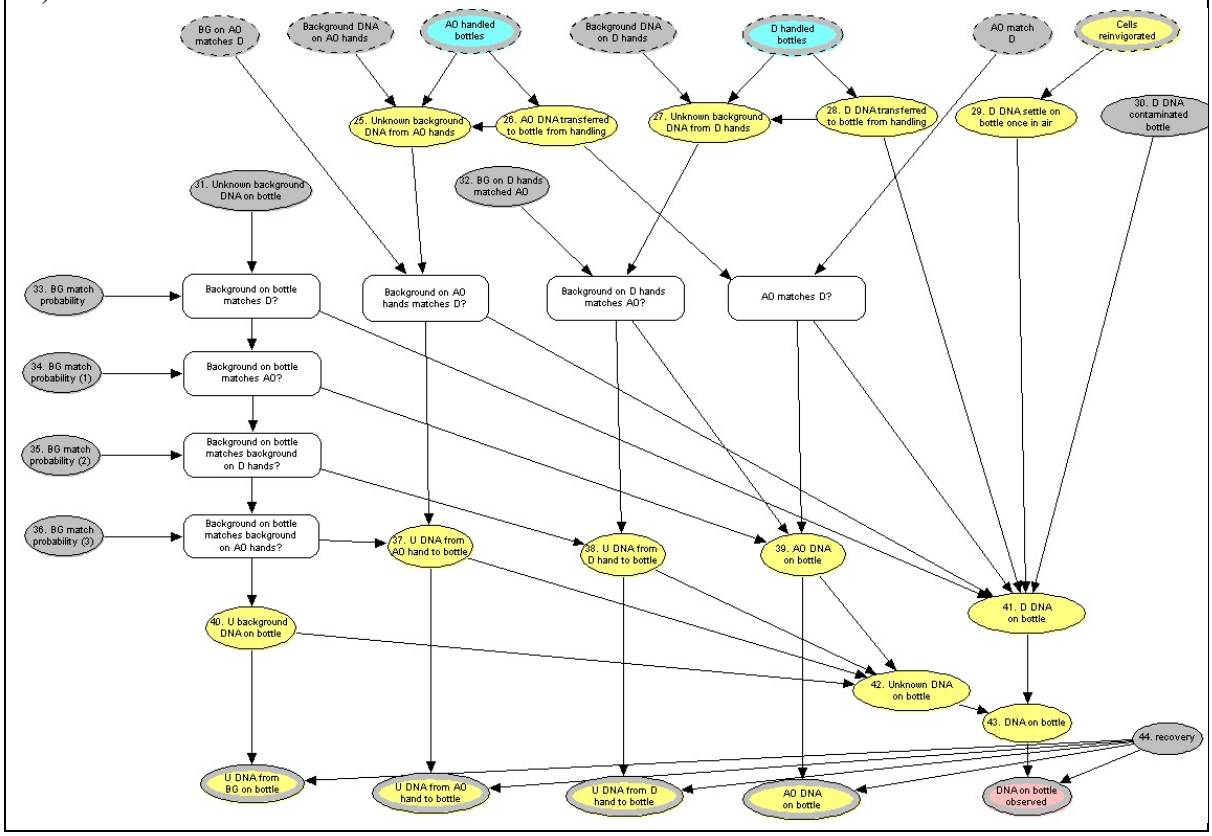
The assignment of probabilities to the nodes within the BN shown in Figure 1 is not the main focus of the paper and so, although important, we do not spend time on explaining the assignments, or carrying out sensitivity analyses. We provide the conditional probability tables as supplementary material (and the OOBN itself) as supplementary material for the interested reader.

A) Main BN

[1] Note that the 'air DNA' mechanism itself should be considered an occurrence of contamination that has occurred as part of the processing of the crime scene. The additional contamination node relates to the probability of other, more conventional, contamination routes e.g. transfer via gloves.



B) Bottle class BN



421

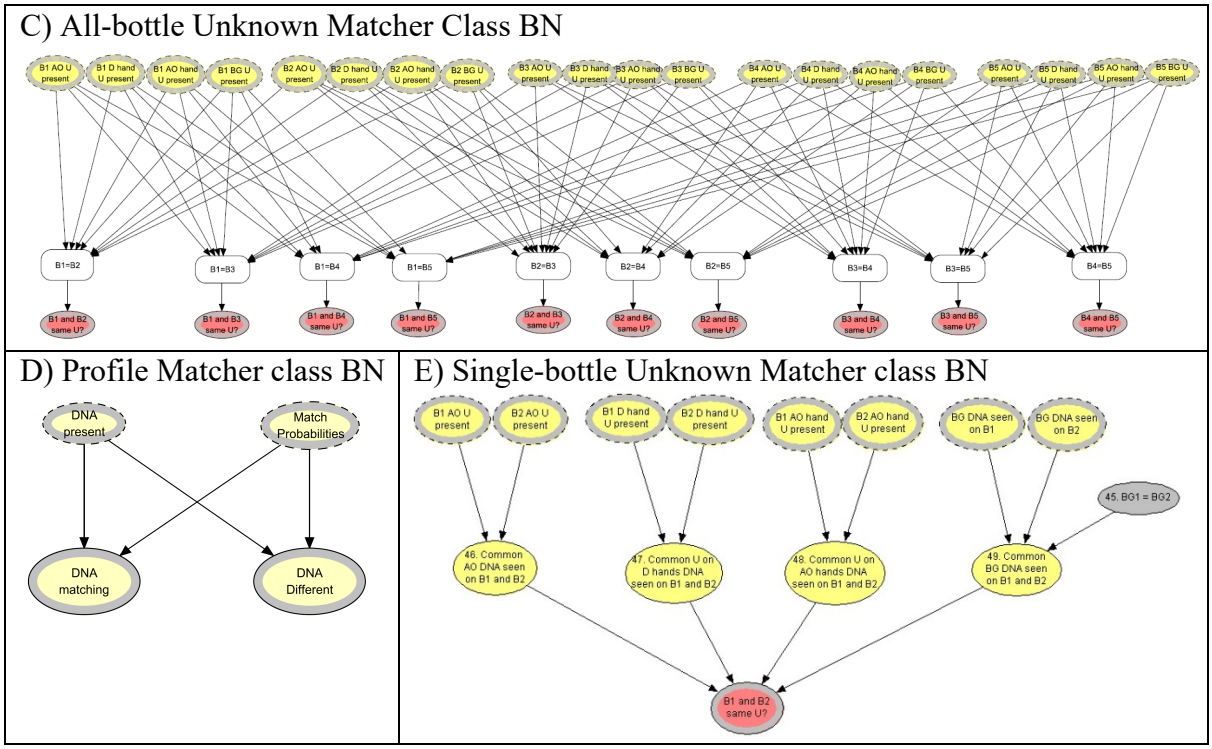


Figure 1: BN considering results for all bottles in the case, and whether common unknowns were present. Panel A shows the over-arching main BN, parts B and C the second layer of class BNs dealing with individual bottles and unknown profile pairwise comparisons and parts D and E show the third layer of class networks dealing with profile and unknown matching. Input nodes are signified by grey border with a dashed outline, and output nodes are signified by a grey border with a solid outline. The function nodes 'LR' and 'Inv LR' allow for calculation of the ratio of the posterior probabilities for each proposition and are not an integral part of the network.

4.4 – Providing the results from the case

Given the mixture-to-mixture result (i.e. an *LR* = 50 in support of the propositions that the bottles did not possess a common donor compared to them possessing a common donor, other than the defendant) we assume that a common unknown donor does not exist. Given this, and the other results previously described, the following case results can be instantiated:

- B1 and B2 same U? – N
- Bottle 1 DNA results – D+U
- Bottle 2 DNA results – D+U
- Bottle 3 DNA results – None
- Bottle 4 DNA results – None
- Bottle 5 DNA results – None

and the *LR* obtained is approximately 107 000 in support of Hp over Hd. We treat 'U' as being the presence of any unknown DNA (from one or multiple sources), and that matching unknowns is matching between any of these 'U's on the bottles being compared. This is a simplification of reality, in that it is possible to have different matching unknowns on different sets of bottles, with one or more bottles having multiple unknowns belonging to multiple sets.

Our BN does not account for such refinement of unknown matching. The instantiated form of the main BN is shown in Figure 2.

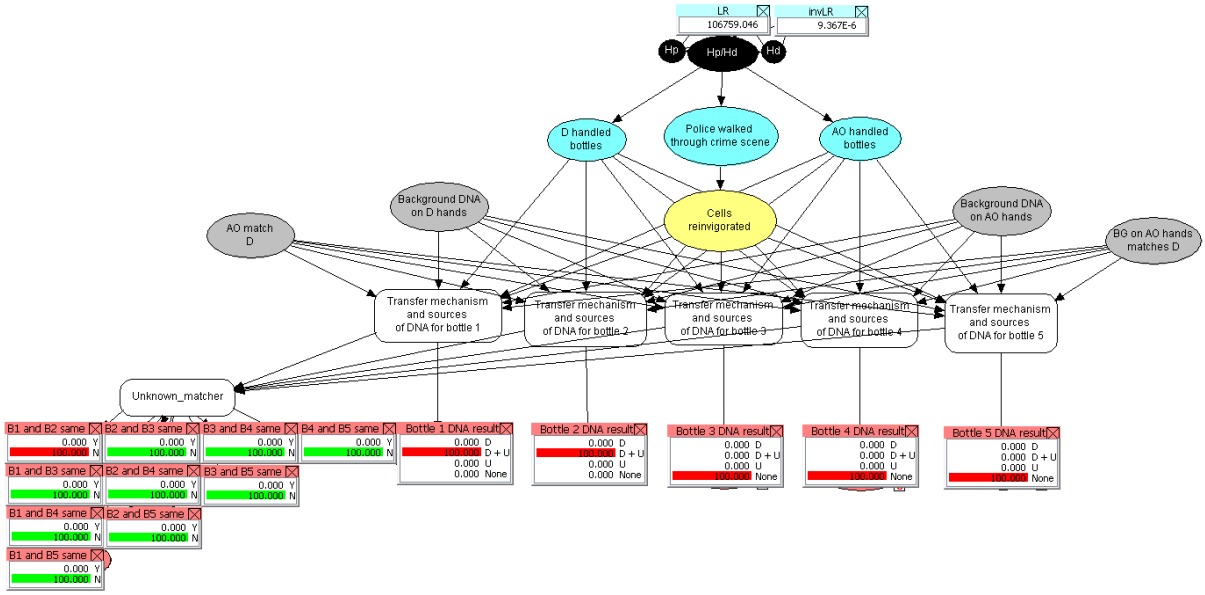


Figure 2: Main BN seen in Part A of Figure 1 with case findings instantiated

In Figure 2 it can been seen when the result of the DNA profiling on bottles are instantiated to 'None' then the unknown matching nodes only have the possibility of 'N' (no unknowns matching) due to the fact that there is no unknown DNA to match.

The *LR* shown in Fig 2 is quite high compared to many evaluations that consider latent DNA samples [23]. The strength comes from the fact that the probability of the results under the defence routes are driven by:

- the air invigoration mechanism, which is an accumulation of an air invigoration (with a probability of 1 in 1000) and two instances of cells landing on the bottle tops (each with a probability of 1 in 100).
- Contamination of the two bottles with the defendant's DNA (each with a probability of 1 in 1000)

Which is quite small compared to the probabilities for the results given the prosecution DNA transfer route, which is two transfer events from direct touching (each with a probability of 0.5). Indeed, given the values we have assigned to the probabilities of DNA air invigoration or DNA (once invigorated) settling back onto the bottles, it is not surprising that the level of support given to Hp over Hd is large. With such assignments we could have intuitively been able to predict such an outcome. It is likely this intuition that led to the appeal judge's description of the scenario as far-fetched or guided the answers of the forensic scientist in court. In our assignments of these probabilities we have not relied on any data, as the assignment itself is not the focus of our work (rather the architecture of the evaluation), however if the case were evaluated for court purposes it would require more rigorous testing, and likely an investigation into the sensitivity of the *LR* to the assignment, which would inform the analyst as to whether the evaluation was robust.

4.5 – Dealing with the presence of unknowns in the evaluation

Having seen the results of the evaluation using the full network, we will now demonstrate why the seemingly extensive consideration of the presence of unknown DNA is important in this evaluation.

In the motivating case, the defence proposition stipulates that if the defendant has not placed the bottles in the ceiling then an AO has. As a specific person has not been suggested, we cannot obtain a reference and their DNA would be interpreted as unknown. If an AO has placed the bottles in the ceiling then we may expect to see a common unknown profile on the bottle top, however there are other mechanisms (both under the prosecution or defence propositions) by which we may also find a common unknown profile on multiple bottles:

- An unknown background DNA source is present on the defendant's hands (UH) and this is transferred to the bottles when they are touched by the defendant.
- There is background DNA (BG) on the bottles from different sources and these have the same alleles (or similar enough to result in support for a common donorship)
- UH is transferred to some bottles and matches the BG on others
- BG is on the bottles and matches the AO on others
- There is unknown background DNA on the AO hands and this is transferred to the bottles when they are touched by the AO
- BG is on the bottles and matches the unknown background on the AO hands

We do not consider aspects of matchings such as the defendant having background DNA on their hands that matches themselves. We also do not consider the possibility of background DNA being present on the first bottle, and then transferred to the hands of the offender and then on to subsequent bottles. This mechanism could be subsumed into the common unknown background DNA on the hands of D or AO being transferred to the bottles, as long as the first bottle is the source of the transferred background. If it is not then we are in a position that we need to consider the order of the bottles being handled, and as this is not known then we need to model uncertainty in the bottle handing order, which adds additional complexity into the evaluation.

The consideration that a common unknown comes from the hands of the defendant as background is important in order to obtain a result that is intuitive with our understanding of DNA in the real world. It is not uncommon for individuals to carry non-self DNA on their hands and transfer this through touching or handling an object [24-26]. Without this consideration, then any presence of matching unknown DNA on the bottles must be described under the prosecution proposition as occurring by the chance matching of alleles (an extremely unlikely event) and will drive the *LR* very strongly towards support for the defence proposition. In the example seen in Figure 2, if we instantiate the node 'B1 and B2 are same U?' to 'Y' then the *LR* increases to approximately 388 000 in favour of Hp over Hd. While this may initially seem counter-intuitive, it is a result of the fact that given D's DNA has been found on the bottles, the most probable source of a common unknown is now unknown BG on D's hands

(hence adding more support to Hp). If the node 'Background DNA on D hands' is instantiated to 'N' (mimicking a BN that did not consider the BG on D's hands as a potential source of common unknown DNA) then the *LR* becomes approximately 4300 in favour of Hd over Hp i.e. if we hadn't considered the possibility of a common unknown source of DNA coming from D's hands in this BN then the effect would have been a shift in the *LR* by a factor of approximately 1 billion. This is because without the common BG on D's hands the only possibility for obtaining a common unknown under Hp is via chance profile matching background occurring (with probabilities of 1 in 1 billion).

If the DNA on the bottles is only unknown (i.e. there is no DNA from the defendant), then the additional information of those unknowns having a common source provides more support to Hd. We show several alternative instantiations of (fictitious) results in Table 1.

Bottles	**Common U DNA**	**Background hands**		***LR* (support for)**
		D hands	**AO hands**	
B1 – D+U B2 – D+U B3 – None B4 – None B5 – None	B1≠B2	Allowed	Allowed	106 759 (Hp over Hd)
		No	Allowed	180 712 (Hp over Hd)
		Allowed	No	105 878 (Hp over Hd)
		No	No	179 220 (Hp over Hd)
B1 – D+U B2 – D+U B3 – None B4 – None B5 – None	B1=B2	Allowed	Allowed	388 193 (Hp over Hd)
		No	Allowed	4301 (Hd over Hp)
		Allowed	No	374 157 (Hp over Hd)
		No	No	4463 (Hd over Hp)
B1 – U B2 – U B3 – U B4 – U B5 – U	B1=B2=B3=B4=B5	Allowed	Allowed	6 400 000 (Hd over Hp)
		No	Allowed	1.6×10^{46} (Hd over Hp)
		Allowed	No	6 400 000 (Hd over Hp)
		No	No	1.6×10^{46} (Hd over Hp)
B1 – U B2 – U B3 – U B4 – U B5 – U	B1≠B2≠B3≠B4≠B5	Allowed	Allowed	11 (Hd over Hp)
		No	Allowed	12 (Hd over Hp)
		Allowed	No	10 (Hd over Hp)
		No	No	11 (Hd over Hp)
B1 – D B2 – D B3 – D B4 – D B5 – D	not relevant as there is no unknown DNA present	Allowed	Allowed	1×10^{9} (Hp over Hd)
		No	Allowed	2×10^{9} (Hp over Hd)
		Allowed	No	5×10^{8} (Hp over Hd)
		No	No	1×10^{9} (Hp over Hd)

Table 1: LR obtained for varying instantiations of results

Note that in our BN construction we have chosen to include the consideration of common unknowns arising from BG DNA on the AO hands. This may seem odd, given that the AO themselves will already appear as unknown DNA. The reason for this is that in order to maintain sensible behaviour in the BN we seek to treat the defendant and the AO equally i.e. whatever the probability of DNA transfer is for one, so too should it be for the other. In somewhat special circumstances, we could have different transfer probabilities for D compared to AO (for example D could be a very poor shedder and we use an average shedder for AO). In the BN shown in Figure 1 the presence of unknown DNA on the AO hands has little effect on the *LR* (as seen in Table 1 in the differences between *LR*s when both unknowns on hands is allowed as background vs when the unknown on AO hands is not allowed) and so could arguably have been left out. However, our own anecdotal experience with workshopping counterintuitive BN behaviour has sometimes arisen from unbalanced treatment of defendant and AO in the architecture of the BN.

DISCUSSION:

The BN we have constructed in Figure 1 takes into account the possibility of contamination of exhibits with the defendant's DNA, and also the possibility for chance matching DNA profiles between different sources of DNA. While the defence do not specifically mention these as part of their proposition or case argument, we do not believe there is a need for them to do so. Indeed, whether or not they mention these occurrences will not have affected the probability of their occurrence. In many evaluations the consideration of these relatively improbable events will not have a significant effect on the *LR*, as the presence of DNA matching the defendant is much more probable given alternate explanations, and so will limit the size of the *LR*. In evaluations such as the one demonstrated, where the scenario being put forward by defence requires one or more quite improbable events to occur, the inclusion of events such as contamination or chance matching DNA profiles become more important to include.

There are various other ways in which the BN in Figure 1 could be extended. We could consider DNA amounts throughout the network rather than simply the presence or absence. This would allow us to consider extraction and sampling efficiency distributions as in [23]. This would also allow a more detailed evaluation with regards to the air DNA mechanism and how the amount of DNA expected to be deposited from such a route compares to the amount of DNA expected from a direct transfer. Incorporating DNA amounts into the BN would also allow us to make use of the fact that the defendant's DNA was the major component of the mixtures observed in the two bottles that had DNA profiles generated. The difficulty with extending the BN to consider DNA amounts comes from the lack of knowledge regarding DNA amounts that relate to different transfer mechanisms in this study, particularly the air-DNA mechanism for which there is little to no relevant literature available. Hence, even if adding DNA amounts may have offered increased discrimination between propositions, adopting a presence/absence strategy is the adequate level of granularity allowing to maintain operational use. As with any evaluation there is always additional factors that could be considered, and there is sometimes a choice as to the level of complexity required to provide meaningful guidance to the court.

Another consideration could be that if the air-DNA mechanism is to resuspend DNA from the surroundings into the air then we could consider the background DNA in the environment, so that in the BN if the defendant's DNA was resuspended, so too might we expect unknown DNA to be also. In this instance the evaluation could have been assisted by taking background DNA swabs (i.e. of the floor or area around where the bottles were swabbed by police) to see if the defendant's DNA was indeed present. We note though that this is not usual practice for most forensic crime scene processing, and the relevance of doing so (i.e. the knowledge of the defence proposition) was not known for three years after the offence. Given that indirect transfer mechanisms are increasingly proposed as 'explanations' for finding DNA on items of interest, anticipating such scenarios by incorporating collection of background samples in routine crime scene procedure may be advisable [27].

Note the importance of including all the bottles in the evaluation, even though no DNA was obtained from some of the extractions (note though that there is no benefit to including the bottle swab that was not analysed in the BN as a complete lack of knowledge of the DNA result provides no power to discrimination between propositions; a lack of knowledge is quite different from a lack of DNA). This consideration would be particularly important if DNA amounts were used in the BN, but even just considering the presence or absence of DNA, there is an effect of knowing that little to no DNA was obtained from three of the bottles. In the case scenario the *LR* when providing all case information was 107 000 in support of Hp over Hd (from section 4.4). If the bottles that did not yield any DNA were not included in the BN (mimicked by not instantiating those nodes) then the *LR* is 104 000. This slight decrease comes about from the fact that as knowledge of no DNA being found is taken away then the probability of the cells being invigorated into the air slightly increases compared to the probability of DNA presence given direct transfer and so the *LR* slightly decreases. While the same probability is assigned under both propositions, the decrease in *LR* comes from the fact that Pr(E|Hp) is the same, but Pr(E|Hd) has slightly increased.

As iteratively more bottles are instantiated to include the defendant's DNA the increase in the *LR* in support of Hp over Hd lessens. This again comes down partly to the effect described above (i.e. an increase in the invigoration mechanism), but as more bottle results are added the dominant defence proposition becomes an AO having a matching profile with the defendant. The effect can be seen in Figure 3, where from zero to five bottles are instantiated with result 'D+U' and common unknown nodes are not instantiated. When the first bottle is instantiated the support for Hp over Hd is approximately 2.6 bans (*LR* expressed in $\log_{10}$), whereas by the time the fifth bottle is instantiated virtually no additional support is gained, and the *LR* plateaus at the probability of the defendant and the AO having matching profiles.

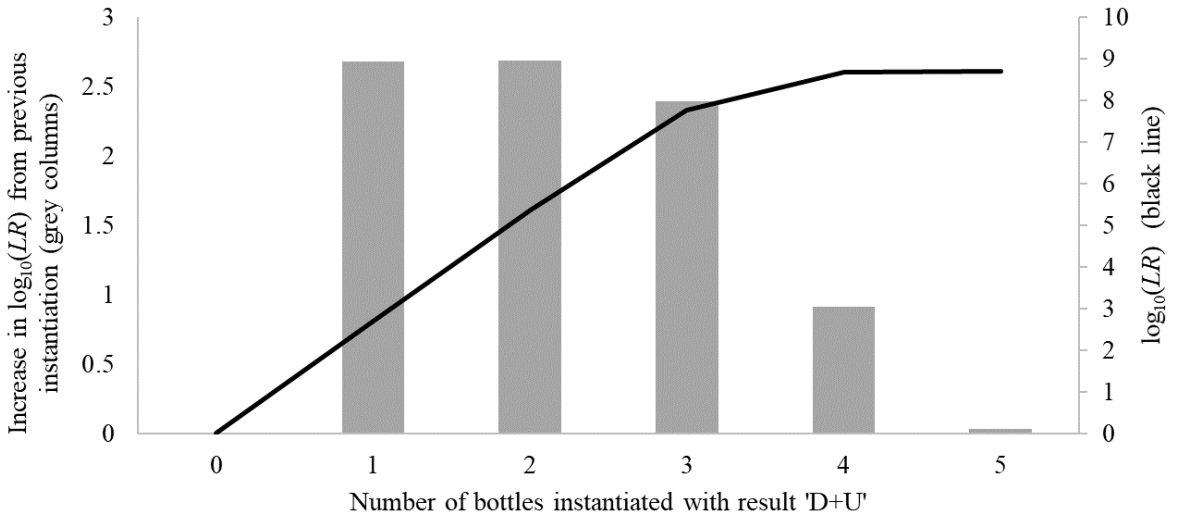


Figure 3: increase in LR in support for Hp over Hd when zero to five bottles are instantiated with result 'D+U' and common unknown nodes are not instantiated. The total LR is shown in black on the graph (and relates to the right-hand axis). The increase in the total $log_{10}(LR)$ with the addition of each bottle is shown with grey bars (and relates to the left axis).

This same effect of the chance matching profiles of the defendant and AO can be seen in the *LR*s in Table 1 all being around the inverse of the match probabilities when all result are instantiated to 'D'. The slight deviations from exactly 1 billion with the different instantiations of background DNA on the hands of the defendant or AO come from the multiple possible routes of matching background DNA within the BN.

In our BN construction we have used the general value of 1 in 1 billion for all profile match probabilities (i.e. between D and AO, but also between unknowns and AO or D). In carrying out this approximation it allows the use of a single class network for the profile matching. However, we could consider different probabilities for each match, which was related to the level of DNA profile information obtained. To do so the profile matching class network would require an additional input node, which would have passed into it a probability of matching, and which then could be set individually per matching type (or per matching type and per bottle if the profile matching probabilities were passed down from the main BN). It depends on the position and type of matching that is being considered but given higher probabilities of matching profiles could lead to the *LR* plateauing at a lower point than shown in Figure 3.

Also note that when there is seemingly unrelated unknown DNA found on all bottles (i.e. only unknown DNA with no indication of any matching unknowns) then the *LR* becomes quite small in comparison to other scenarios in Table 1, only providing slight support for Hd over Hp, which largely comes from probabilities of non-transfer of the defendant's DNA.

CONCLUSION:

We have shown here an evaluation given activity level propositions of the forensic DNA profiling results from five bottles, all of which have been treated in a very similar manner

within the framework of circumstances of the case. Our focus was not on the probability assignment (although this an important aspect of any evaluation), but rather the appropriate treatment of the DNA on these items, and particularly the importance of the correct treatment of unknown DNA. Given the sensitivity of modern DNA profiling techniques, and our ever-increasing knowledge of the prevalence of DNA on items, it is possible to deduce multiple pathways for a common unknown DNA profile to be present on multiple similar items. We have shown the importance of including the key pathways for common unknowns to exist under both prosecution and defence propositions. In our BN, if the presence of unknown background DNA on the hands of the defendant is not considered then the transition from unknowns that are not common, to unknowns that are common between bottles, gives rise to an unrealistically dramatic change in support for the propositions.

In an extension to our BN we could consider DNA amounts rather than the presence or absence of DNA, and we could consider environmental unknown DNA that could have also been resuspended back into the air and these would produce an even more discriminating *LR* given the case circumstances and findings (as long as appropriate data existed to assist with probability assignment).

In the motivating case the defence proposition of an air-DNA transfer is not very well supported by the observed results compared to the prosecution proposition of a direct contact (as illustrated by its description by an appeal judge as *'A highly far-fetched theory of innocent indirect DNA transfer'*) and the evaluation of the case results provides very strong support for the evidence given Hp rather than given Hd, if the probabilities assigned are valid.
Even with much higher probabilities for the DNA swirling into the air, and then landing on bottles (set at some values that an analyst feels may represent the upper bound of a reasonable range, which intuition tells us must be lower than the probability of transfer from a direct contact), the evaluation will still favour Hp over Hd. The detailed probabilistic analysis developed in this paper confirms the broad intuitive assessment of both the reporting scientist and the court. It may lead some to conclude that such a complex analysis is not needed. We disagree with that view. The benefit of the above approach is to be able to actually number the probabilities of the observations given each of the allegations. It enables to qualify what is meant by '*highly far-fetched*' or '*being an unlikely way for DNA to transfer*'.

Supplementary material:

1. The OOBN from Figure 1
2. A description on the setup of each node and the population of the conditional probability tables with data

REFERENCES:

[1] P. Gill, T. Hicks, J.M. Butler, E. Connolly, L. Gusmão, B. Kokshoorn, N. Morling, R.A.H. van Oorschot, W. Parson, M. Prinz, P.M. Schneider, T. Sijen, D. Taylor, DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence - Guidelines highlighting the importance of propositions. Part II: Evaluation of biological traces considering activity level propositions, Forensic Science International: Genetics 44 (2020) 102186.

[2] S.M. Willis, L. McKenna, S. McDermott, G. O'Donell, A. Barrett, B. Rasmusson, A. Nordgaard, C.E.H. Berger, M.J. Sjerps, J.-J. Lucena-Molina, G. Zadora, C. Aitken, T. Lovelock, L. Lunt, C. Champod, A. Biedermann, T.N. Hicks, F. Taroni, ENFSI Guideline for Evaluative Reporting in Forensic Science, European Network of Forensic Science Institutes (available at http://enfsi.eu/sites/default/files/documents/external_publications/m1_guideline.pdf), 2015.

[3] D. Taylor, A. Biedermann, T. Hicks, C. Champod, A template for constructing Bayesian networks in forensic biology cases when considering activity level propositions, Forensic Science International: Genetics 33 (2018) 136-146.

[4] D. Taylor, The evaluation of exclusionary DNA results: a discussion of issues in R v. Drummond, Law, Probability and Risk 15(1) (2016) 175-197.

[5] B. Szkuta, K.N. Ballantyne, B. Kokshoorn, R.A.H. van Oorschot, Transfer and persistence of non-self DNA on hands over time: Using empirical data to evaluate DNA evidence given activity level propositions, Forensic Science International: Genetics 33 (2018) 84-97.

[6] K. Steensma, R. Ansell, L. Clarisse, E. Connolly, A.D. Kloosterman, L.G. McKenna, R.A.H. van Oorschot, B. Szkuta, B. Kokshoorn, An inter-laboratory comparison study on transfer, persistence and recovery of DNA from cable ties, Forensic Science International: Genetics 31 (2017) 95-104.

[7] B. Kokshoorn, B.J. Blankers, J. de Zoete, C.E.H. Berger, Activity level DNA evidence evaluation: On propositions addressing the actor or the activity, Forensic Science International 278 (2017) 115-124.

[8] I.W. Evett, On meaningful questions: A two-trace transfer problem, Journal of the Forensic Science Society 27 (1987) 375-381.

[9] C.M. Triggs, J. Buckleton, The two trace transfer problem revisited, Science and Justice 43(3) (2003) 127-134.

[10] S. Gittelson, A. Biedermann, S. Bozza, F. Taroni, Bayesian networks and the value of the evidence for the forensic two-trace transfer problem., Journal of Forensic Sciences 57(5) (2012) 1199-1216.

[11] P. Juchli, A. Biedermann, F. Taroni, Graphical probabilistic analysis of the combination of items of evidence, Law, Probability and Risk 11 (2012) 51-84.

[12] J.-A. Bright, D. Taylor, Z. Kerr, J. Buckleton, M. Kruijver, The efficacy of DNA mixture to mixture matching, Forensic Science International: Genetics 41 (2019) 64-71.

[13] K. Slooten, Identifying common donors in DNA mixtures, with applications to database searches, Forensic Science International: Genetics 26 (2017) 40-47.

[14] E. Dowlman, N. Martin, M. Foy, T. Lochner, T. Neocleous, The prevalence of mixed DNA profiles on fingernail swabs, Science and Justice 50 (2010) 64-71.

[15] T. Boyko, R.J. Mitchell, R.A.H. van Oorschot, DNA within cars: prevalence of DNA from driver, passenger and others on steering wheels, Australian Journal of Forensic Sciences 51(sup1) (2019) S91-S94.

[16] D. Taylor, D. Abarno, E. Rowe, L. Rask-Nielsen, Observations of DNA transfer within an operational Forensic Biology Laboratory, Forensic Science International: Genetics 23 (2016) 33-49.

[17] J.A. de Koeijer, M.J. Sjerps, P. Vergeer, C.E.H. Berger, Combining evidence in complex cases - a practical approach to interdisciplinary casework, Science & Justice 60(1) (2020) 20-29.
[18] R v Quist, Supreme Court of Australia (Court of Criminal Appeals), SASCFC 37, 2017.
[19] D. Taylor, J.-A. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Science International: Genetics 7(5) (2013) 516-528.
[20] D. Taylor, L. Samie, C. Champod, Using Bayesian networks to track DNA movement through complex transfer scenarios, Forensic Science International: Genetics 42 (2019) 69-80.
[21] S.K. Anderson, K.G. Olesen, F.V. Jennings, F. Jensen, HUGIN - a shell for building Bayesian belief universes for expert systems., Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (1989) 1080-1085.
[22] A. Biedermann, S. Bozza, F. Taroni, M. Furbach, B. Li, W.D. Mazzella, Analysis and evaluation of magnetism of black toners on documents printed by electrophotographic systems, Forensic Sci Int 267 (2016) 157-165.
[23] D. Taylor, A. Biedermann, L. Samie, K.-M. Pun, T. Hicks, C. Champod, Helping to distinguish primary from secondary transfer events for trace DNA, Forensic Science International: Genetics 28 (2017) 155-177.
[24] M. Goray, S. Fowler, B. Szkuta, R.A.H. van Oorschot, Shedder status - An analysis of self and non-self DNA in multiple handprints deposited by the same individuals over time, Forensic Science International: Genetics 23 (2016) 190-196.
[25] M. Goray, S. Fowler, B. Szkuta, R.A.H.v. Oorschot, Corrigendum to: Shedder status - An analysis of self and non-self DNA in multiple handprints by the same individuals over time; [Forensic Sci. Int. Genet. 23 (2016);190-196], Forensic Science International: Genetics 34 (2018) e26.
[26] B. Szkuta, K.N. Ballantyne, R.A.H. van Oorschot, Transfer and persistence of DNA on the hands and the influence of activities performed, Forensic Science International: Genetics 28 (2017) 10-20.
[27] E.M. Ton, J. Limborgh, L.H.J. Aarts, B. Kokshoorn, J. de Koeijer, M.C. Zuidberg, Plaats delict-onderzoek met vooruitziende blik: Anticiperen op alternatieve scenario's tijdens het sporenonderzoek op de plaats delict, Expertise en Recht 4 (2018) 144-149.

Supplementary: Probability assignments for BNs in Figure 1

Main BN (Part A of Fig 1)

Node 1 (Hp/Hd): The proposition node with possible states of 'Hp' for prosecution or 'Hd' for defence. The 'Hp' option is the prosecution scenario and 'Hd' considers the defence scenario as given in section 4.0. These two options are assigned with equal prior probabilities. Note that this does not mean the prior odds in this case are equal; equal prior probabilities are applied for the propositions so that the values obtained by the BN inform the likelihood ratio for the scientific evidence only.

Hp/Hd

Hp	0.5
Hd	0.5

Node 2 (D handled bottles): considers the activity of D placing the bottles in the ceiling space, which has occurred under Hp and not under Hd.

Hp/Hd

D handled bottles

	Hp	Hd
Yes	1	0
No	0	1

Node 3 (Police walked through crime scene): There is no dispute that this activity has occurred and therefore 1 is assigned for state 'yes' under both Hp and Hd.

Hp/Hd

Police walked through crime scene

	Hp	Hd
Yes	1	1
No	0	0

Node 4 (AO handled bottles): considers the activity of the AO placing the bottles in the ceiling space, which has occurred under Hd and not under Hp.

Hp/Hd

D handled bottles

	Hp	Hd
Yes	0	1
No	1	0

Node 5 (AO match D): This root node shows the probability that an unknown person will share the same DNA profile and here we use the profile probability for the defendant of 1 in 1 billion.

AO match D

Match	1.0E-9
No match	1 – 1.0E-9

Node 6 (Background DNA on D hands): Considers how often non-self DNA is found on the hands. For illustration of the BN performance we use values of 0.5 for 'yes' and 'no' but concede the presence of background DNA on hands has been shown to be more prevalent as demonstrated in (Szkuta et al. 2017).

Background DNA on D hands

Yes	0.5
No	0.5

Node 7 (Cells reinvigorated): The probability that DNA relocation can occur via shed cells in the pathway of investigators processing a crime scene. We consider the possibility of cell relocation during evidence collection as one event, and allocate a probability of 1 in 1000. If investigators did not walk through the pathway, then cell relocation cannot occur, therefore 0 is assigned for state 'yes' and 1 is assigned for state 'no'.

Cells reinvigorated

Police walked through crime scene	Yes	No
Yes	0.001	0
No	0.999	1

Node 8 (Background DNA on AO hands): The same treatment as non-self DNA on the hands of the defendant so states will have the same values as node 6.

Background DNA on AO hands

Yes	0.5
No	0.5

Node 9 (BG on AO hands matches D): The same treatment as node 5. Although already accounting for the presence of unknown DNA from the AO, this node considers unknown DNA present as background on this unknown person's hands.

BG on AO hands matches D

Match	1.0E-9
No match	1 – 1.0E-9

Node 10-19 (B1 and B2 same U?, B1 and B3 same U? …, B4 and B5 same U?): Considers the different sources of unknown DNA for bottle 1 and bottle 2. This accounts for whether the same unknown DNA that may be present from background DNA on the defendant's hands, unknown DNA from the AO, background DNA on the AO's hands, and background DNA on the bottles. The same reasoning then extends to the remaining nodes for all pairwise comparisons for each bottle.

Unknown_matcher

B1 and B2 same U?

	Yes	No
Yes	1	0
No	0	1

Node 20-24 (Bottle 1 DNA results …, Bottle 5 DNA results): These nodes combine the parental node values to consider the different transfer mechanisms and sources of DNA for each bottle (example of the conditional probability table is shown here for bottle 1). Here we use presence or absence of DNA for the various sources and consider possibilities of the defendant's DNA (D), defendant and unknown DNA (D+U), unknown DNA only (U) or no DNA at all (None).

Transfer mechanisms and sources of DNA for bottle 1

Bottle 1 DNA results

	D	D+U	U	None
D	1	0	0	0
D+U	0	1	0	0
U	0	0	1	0
None	0	0	0	1

Bottle class BN (Part B of Fig 1)

Node 25 (Unknown background DNA from AO hands): If the AO handled the bottles (Y), with background DNA present on their hands (Y) and DNA had transferred to the bottle from handling (Y), then we expect the probability to find unknown background DNA to be quite high. We have assigned this value as 0.95 for state 'yes' and 0.05 for state 'no'.

Alternatively, if the AO handled the bottles with background DNA on the hands, however their DNA did not transfer to the bottle from handling (N), then we expect the probability of unknown DNA to also transfer to be quite low. The values assigned in this scenario are 0.05 for state 'yes' and 0.95 for state 'no'.

Finally, if the AO did not handle the bottle (N), or background DNA was not present on the hands (N), then unknown background DNA cannot be present via the hands. All values here are assigned 0 for state 'yes' and 1 for state 'no'.

AO handled bottles	Y				N			
Background DNA on AO hands	Y		N		Y		N	
AO DNA transferred to bottle from handling	Y	N	Y	N	Y	N	Y	N
Y	0.95	0.05	0	0	0	0	0	0

Unknown background DNA from AO hands	N	0.05	0.95	1	1	1	1	1	1

Node 26 (AO DNA transferred to bottle from handling): Considers the probability for DNA to transfer via a direct contact to the bottle surface. For illustration we use a rate of transfer of 0.5 so from handling the five bottles directly, we would expect to obtain a DNA profile from the handler two or three times. Again, if the bottles were not handled, then DNA cannot transfer and we assign 0 for state 'yes' and 1 for state 'no'.

AO handled bottles		Yes	No
AO DNA transferred to bottle from handling	Yes	0.5	0
	No	0.5	1

Node 27 (Unknown background DNA from D hands): If the defendant handled the bottles (Y), with background DNA present on the defendant's hands (Y) and DNA had transferred to the bottle from handling (Y), then we expect the probability to find unknown background DNA to be quite high. We have assigned this value as 0.95 for state 'yes' and 0.05 for state 'no'.

Alternatively, if the defendant handled the bottles with background DNA on the hands, however the defendant's DNA did not transfer to the bottle from handling (N), then we expect the probability of unknown DNA to also transfer to be quite low. The values assigned in this scenario are 0.05 for state 'yes' and 0.95 for state 'no'.

Finally, if the defendant did not handle the bottle (N), or background DNA was not present on the hands (N), then unknown background DNA cannot be present via the hands. All values here are assigned 0 for state 'yes' and 1 for state 'no'.

D handled bottles		Y				N			
Background DNA on D hands		Y		N		Y		N	
D DNA transferred to bottle from handling		Y	N	Y	N	Y	N	Y	N
Unknown background DNA from D hands	Y	0.95	0.05	0	0	0	0	0	0
	N	0.05	0.95	1	1	1	1	1	1

Node 28 (D DNA transferred to bottle from handling): The same treatment as node 26, to inform the node on this DNA transfer mechanism we expect similar findings for the AO and the defendant. Indeed shedder status, time since hand washing and opportunity for DNA loading will all impact this value, and we do not have case specific information to inform us of the shedder status of the defendant. We are left guided by scientific literature alone.

D handled bottles		Yes	No
D DNA transferred to bottle from handling	Yes	0.5	0
	No	0.5	1

Node 29 (D DNA settle on bottle once in air): We consider DNA to settle on each of the five bottles as individual events, therefore we have the 'cells reinvigorated' node on the outer layer of the BN as an input node, and the 'D DNA settle on bottle once in air' node present in the class OOBN inner layer. We have assigned a probability of 1 in 100 for each bottle so state 'yes' is 0.01 and state 'no' is 0.99. Again, if cells were not present in the air environment then the defendant's DNA cannot settle on a bottle (state 'no' = 1).

Cells reinvigorated		Yes	No
D DNA settle on bottle once in air	Yes	0.01	0
	No	0.99	1

Node 30 (D DNA contaminated bottle): Considers the probability of laboratory contamination and we assign a probability of 1 in 1000 (state 'yes' = 0.001 and 'no' = 0.999)

D DNA contaminated bottle	Yes	0.01
	No	0.99

Node 31 (Unknown background DNA on bottle): Considers the prevalence of background DNA on bottles and we assign a value of 0.5 for state 'yes' and 0.5 for state 'no'.

Unknown background DNA on bottle	Yes	0.5
	No	0.5

Node 32 (BG on D hands matched AO): Considers the probability of background DNA on the defendant's hands to have same alleles as an unknown donor as we use the profile probability of 1 in 1 billion.

BG on D hands matched AO	Match	1.0E-9
	No match	1 – 1.0E-9

Node 33-36 (BG match probability): This series of nodes considers whether background DNA on the bottle is the same as the defendant, alternate offender, background DNA on the

defendant's hands, or background DNA on the AO's hands. Again we use the profile probability as with node 32.

BG match probability

Match	1.0E-9
No match	1 – 1.0E-9

Node 37-43 are accumulation nodes of the various sources of unknown DNA, or DNA from the Defendant, which may be present on the bottle.

Node 44 (Recovery): Considers the ability to recover DNA from the outer surface of the bottle lids using a swabbing technique followed by DNA extraction. We consider the ability to obtain a DNA profile with efficiency of 0.8 for state 'yes' and 0.2 for state 'no'.

Recovery

Yes	0.8
No	0.2

Single-bottle unknown matcher class BN (Part E of Fig 1)

Node 38 (BG1 = BG2): Considers whether background DNA observed on two bottles will share a common unknown donor. Again, we use the profile probability for whether the two sources of unknown DNA will share the same alleles.

BG1 = BG2

Match	1.0E-9
No match	1 – 1.0E-9

Node 46-49 are the accumulation of common sources of DNA that may be observed from background, and again we do not describe the probability assignments as data is not required to inform these nodes.