Victoria State Coroner's Inquest into Death of Jaidyn Leskie

Report Prepared by William C. Thompson

December 3, 2003

"How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?"
Sherlock Holmes in A. Conan Doyle's A Study in Scarlet

I am a professor in the Department of Criminology, Law & Society at the University of California, Irvine. I received a Ph.D. in psychology from Stanford University (specialty: scientific inference and probabilistic judgment) and a J.D. (Juris Doctorate) from the University of California, Berkeley. I have been studying legal applications of genetic technology since 1986. I have published more than 25 articles on forensic DNA evidence. My work addresses both scientific issues (particularly statistical characterization of DNA evidence and interpretation of DNA test results) and legal issues (e.g., admissibility standards, discovery rules, regulation of laboratory practice). I have also been involved as a lawyer in litigation involving DNA evidence and have served as a consultant on DNA issues to other lawyers throughout the United States. As a result of my academic and legal work I am familiar with the DNA testing procedures at issue in this matter. I have participated in several professional training courses on the operation of the instruments used for DNA testing in this matter and the interpretation of test results.

Ms. Rowena Orr and Mr. James Kennan, barristers who represent the State Coroner, asked me to prepare this report. At their request, I have reviewed the following materials: laboratory notes and reports related to DNA testing in both the Leskie toddler case (Case #6603/967) and the case involving the alleged rape of a woman referred to as "P" (Case #2831/978)(hereinafter referred to as the rape case); expert reports prepared by Professor Barry Boettcher, Professor Bruce Weir, Mr. John Scheffer, Mr. Robert Goetz, and Dr. Sally Ann Harbison; documents provided by the Victoria Forensic Science Centre, including laboratory protocols and statistical tables. I have also consulted with two scientists who have greater expertise than I concerning the details of laboratory procedure, mechanisms of DNA transfer, and the population genetics of STR loci: Professor Dan Krane of Wright State University, and Mr. Marc Taylor, a distinguished forensic scientist who operates an independent forensic DNA laboratory in California. In the passages that follow, I will endeavor to make clear which of my conclusions are based, in part, on information obtained from these other individuals and which rest entirely on my own analysis.

I have been asked to comment on a troubling event. Two DNA samples that were typed in connection with the Leskie toddler case have produced DNA profiles that match the DNA profiles found in an apparently unrelated case involving the alleged rape of a

woman identified as "P". The DNA profiles of samples that appeared to be blood on a bib and track pants in the Leskie case match the profile of a reference sample purportedly from P as well as the profile of the primary donor of DNA in two mixed samples from a condom that allegedly was employed in the rape of P. Assuming the DNA tests were accurate and were interpreted correctly, there are two possible explanations for the match:

(1) The DNA on the Bib and Track Pants came from an unknown person other than P who happens, by coincidence, to have the same DNA profile as the profile found in P's reference sample and the profiles found on the condom. If this explanation is true, then the DNA evidence connects this unknown person to the blood on the toddler's clothing and thereby suggests this unknown person may have been involved in the toddler case. However, this explanation requires the occurrence of an unlikely coincidence: two people in the state of Victoria both possessing a DNA profile that is sufficiently rare that some experts would declare it a "scientific certainty" that the profile is unique in all the world.

(2) The DNA on the Bib and Track Pants came from P. The problem with this explanation, I have been told, is that police have effectively ruled out any connection between P and the toddler's clothing at any time before the clothing was sent to the Victoria Forensic Science Centre for analysis. Consequently, if this explanation is true, then P's DNA must have gotten onto the toddler's clothing (or into two samples taken from the clothing) in the laboratory itself. This explanation requires either intentional misconduct by laboratory personnel (a possibility that I ignore as I am aware of no possible motivation for it) or an egregious laboratory error.

Overview

I believe the match is more likely due to accidental cross-contamination of samples in the laboratory than to the existence of an unknown person who happens to have same rare profile observed in the rape case. The key factor that weighs in favor of cross-contamination is that samples from the two cases were processed through the laboratory in close proximity in time. Although there is no direct evidence of cross-contamination, and it is unclear exactly how cross-contamination might have occurred, the opportunity for such an event did exist.

In reviewing the work of forensic DNA laboratories in the United States I have encountered a number of incidents in which samples were inadvertently mixed up or cross-contaminated. Cross-contamination, in this context, means that DNA is accidentally transferred from one sample into another. Although these events are relatively rare, they occur often enough that I personally have discovered several of them in reviewing perhaps 100 cases from laboratories that follow procedures similar to those of the Victoria Forensic Science Centre. Professional colleagues have informed me of many other errors of this type. Hence, accidental cross-contamination struck me, from the outset, as a plausible explanation for the match between the two cases.

My review of the laboratory notes revealed that DNA was extracted from the condom only one or two days before the laboratory began processing the bib and track pants. DNA from the condom was being processed and handled in the laboratory while the bib and track pants were being examined and prepared for DNA extraction. Although the laboratory notes reveal no specific evidence of cross-contamination, it is well known that DNA extracted from one sample can accidentally contaminate other items in a laboratory. Many of the cross-contamination episodes in U.S. laboratories have occurred during the DNA extraction process.

I consulted with Mr. Taylor and Dr. Krane about possible scenarios by which DNA might have been transferred from the condom to the bib and track pants. Their opinions coincide with the view expressed by Dr. Sally Ann Harbison, in her expert report, that DNA or cellular material from the condoms might have been deposited on a surface or implement that later came into contact with the bib and track pants, thereby effecting the transfer. The quantities of DNA involved are extremely small. The total amount of DNA detected on the bib and track pants is difficult to estimate precisely, but could well have been contained in as few as 400-500 cells, or even less—an amount too small to be apparent to the naked eye. Because I am not a laboratory scientist, I thought it wise to ask Dr. Krane and Mr. Taylor, who have extensive personal experience in such matters, whether they believed that inadvertent DNA transfer could have occurred under the circumstances of this case. Both agreed that it could have.

The alternative theory of a coincidental match seems less plausible. I am aware of no previous incidents in which two people have been discovered to share a DNA profile as rare as the matching profiles in this case. The only coincidental matches that have been reported involved DNA profiles that were thousands of times more common than the profiles at issue in this matter. For example, a widely reported false match in the United Kingdom involved DNA profiles estimated to occur in one person in 37 million.[1] By contrast, the matching profile in the two cases at issue here would be expected, by the most conservative estimate, to occur in less than one person in 171 billion (and might be considerably less common than that). In other words, one would expect to find more than 4000 people with profiles like that in the notorious U.K. case for each person who has the matching profile in this matter.

A coincidental match cannot, however, be ruled out entirely. When weighing this possibility, it is important to keep in mind that the match was discovered because the profiles from the two cases were entered into a database consisting of nearly 19,000 DNA profiles (of offenders and victims typed by the VFSC). As explained below, the probability that a coincidental match will occur between *any* two of 19,000 profiles in a database is far higher than the probability that a specific DNA profile will happen to match DNA from a random unrelated person. As I will explain below, the former probability may be more relevant than the latter in judging the likelihood that the match in this matter is due to coincidence. Nevertheless, after considering the matter in this

[1] This case, which arose in Manchester England in 1999, is discussed in Rudin and Inman, Introduction to Forensic DNA Analysis, 2nd Edition, CRC Press (2002).

light I still believe the cross-contamination theory is more plausible than the theory of a coincidental match.

I have been asked also to make recommendations on systems improvements that might contribute to more effective administration of justice in future cases. My comments in this regard are concerned primarily with the procedures used by the Victoria Forensic Science Centre to document its work and to make that documentation available to independent experts. As I will explain, there are some important gaps in the documentation that I received on this matter. I will propose some inexpensive procedural changes that will make the work of the laboratory more transparent and easier for independent experts to evaluate in future cases.

Frequency of the Matching DNA Profiles

The Victoria Forensic Science Centre (VFSC) used two genetic systems to type the DNA samples in these cases. One system, called Green I, allows examination of four loci called THO1, TPOX, CSF and Amelogenin. The other system, called ProfilerPlus, allows examination of nine additional loci (labeled for convenience D3, vWA, FGA, D8, D21, D18, D5, D13 and D7) as well as Amelogenin. Due to the low quantities of DNA in the samples, the laboratory was unable to obtain interpretable results at two of the ProfilerPlus loci for the bib, and at four of the ProfilerPlus loci for the track pants.

The profiles found on the two condom samples are consistent with a mixture of DNA from a primary contributor who is female and a secondary contributor who is male. The complainant in the rape case, "P," was tested using the ProfilerPlus system only. Her profile for the nine ProfilerPlus loci is identical to the profile of the primary donor to the two condom samples. (It is unclear why the laboratory failed to type her reference sample with the Green I system). The man accused of raping "P" was tested with both the Green I and ProfilerPlus systems. His profile matches that of the secondary contributor to the DNA on the condom.

Professor Weir estimated that there is one chance in 269 million that a randomly drawn Australian Caucasian would have the seven-locus ProfilerPlus profile that was found in the sample from the bib (70iii a) and in the reference sample from "P." Professor Krane put that probability at 1 in 1.3 billion. As I understand it, these two experts are working with the same data on allele frequencies. The difference in their estimates arises from the procedures they used for estimating genotype frequencies. Professor Krane employed a method recommended by the U.S. National Research Council (Method 4.1) that is widely used in the United States. Professor Weir used a somewhat more conservative method (similar to that employed by the VFSC) that allows a greater margin for possible errors arising from sampling variability and population structure.[2] The appropriate level of

[2] In March, 2000, the VFSC recorded genotype frequency estimates for this profile. Based on these estimates, the expected frequency of the seven-locus matching profile is approximately 1 in 227 million. In the calculations I report below, I will adopt the estimates of the VFSC because they are the most conservative.

"conservatism" for a computation of this type is a matter on which reasonable scientists may differ. I consider both estimates reasonable and informative.

It is important to note, however, that these estimates consider only the seven ProfilerPlus loci. When one compares the profile of the bib to the profile of the primary contributor to the condom samples, one finds not only the seven-locus match on the ProfilerPlus loci, but also a three-locus match on the Green I loci. Therefore I believe it is important to consider the probability of a random match across all ten loci. I computed the expected frequency of the combined 10-locus profile to be 1 in 171 billion (171,000,000,000).[3] Because I used extremely conservative methods, this estimate may understate the true rarity of the profile. Had I employed the procedure used by Professor Krane, my estimate would have been significantly lower.

Database Matches and the Birthday Problem

The foregoing discussion should make it clear that the ten-locus DNA profile found on both the bib from the Leskie case and the condom from the rape case is extremely rare. I doubt that any scientist would dispute that the probability of finding this profile in a randomly chosen individual is less than 1 in 171 billion. However, the so-called random match probability is not necessarily the same as the probability that *this* match was "adventitious" or coincidental. Consider that the match was found when the relevant profiles were entered into a database containing nearly 19,000 profiles.[4] The database allows each new profile entered to be compared with all other profiles in the database. Hence, the probability that a coincidental match will occur between *some* pair of profiles as new profiles are entered into the database is far higher than the probability that any single profile will match a single other profile.

The situation arising from the database search is analogous to what statisticians call "the birthday problem." What is the probability that a randomly chosen individual will have the same birthday that I have? Assuming (for purposes of simplicity) that birthdays are randomly distributed across the 365 days of the year, the chance that a randomly chosen individual will share my birthday is only 1 in 365. Consider next, however, the probability that *any two individuals* in a room will share the same birthday. If there are only two people in the room, the probability of a match will be 1 in 365, but as more people are added to the room the probability of a matching birthday will rise rapidly because each new person has the opportunity to match not just a single person but all of those already in the room. Probability calculations show that by the time the 23rd person has entered the room, the chance that *any two* will have the same birthday exceeds 1 in

[3] In making this estimate, I used the VFSC estimates for the frequency of the ProfilerPlus genotypes, the VFSC general Australian database for determining the frequency of the Green I genotypes, a population structure parameter (theta) of 0.03, and the same conservative assumptions about sampling error incorporated by the VFSC.

[4] I understand that some profiles were entered into the database and later removed pursuant to legal requirements. If so, the effective size of the database for purposes of the present discussion may actually be considerably larger than 19,000.

2.[5] Hence, the chances for an adventitious or coincidental match between *some* pair of individuals in a group may greatly exceed the probability of a random match between a single individual and another individual picked at random.

When evaluating the match between the bib and the condom, then, it may be worthwhile to consider the likelihood that *any two* individuals in a database of 19,000 will be found to share a ten-locus DNA profile. This probability is difficult to calculate, and will vary depending on the rarity of the profiles in the database. However, this probability clearly will be many orders of magnitude higher than the random match probability associated with any single profile. Although I can make only rough estimates at this point, the probability of a coincidental match between *any two* ten-locus profiles in a database of 19,000 could well be in the range of one in thousands rather than one in millions or billions. In other words, the odds that the VFSC, while compiling a database of 19,000 profiles, would at some point encounter a coincidental 10-locus match may be closer to the odds of being dealt an excellent five-card poker hand than the odds of winning a national lottery. Hence, while I believe the match in the present case is unlikely to have been coincidental, I believe that possibility cannot be ruled out.

Professor Krane's report discusses three approaches that have been suggested for analyzing "cold hit" DNA matches. The procedures he describes estimate the probability of a coincidental match between *a specific* DNA profile and any other profile in a database. The "birthday problem" estimates that I discuss here address a different issue: the probability that *any* profile in a database will be found to match *any* other. If the question one wishes to answer is "what is the probability that the profile of "P" will match another profile in the database?" then the approaches discussed by Professor Krane provide relevant answers. If, however, the question one wishes to address is "what is the probability that the VFSC would at some point find a coincidental match between two profiles when compiling a large database?" then the answer must be provided by the "birthday" analysis I have discussed.

Opportunities for Cross-Contamination

Samples from the two cases were processed through the laboratory in close temporal proximity. In many instances the same personnel worked on items from the two cases. Hence, it seems likely (although I cannot confirm this from the laboratory notes) that samples from the two cases were examined on the same tables and manipulated with the same instruments. The samples were quite literally touched by the same hands (albeit with disposable gloves).

Samples from the condom, made their way through the laboratory slightly in advance of the bib and track pants. On February 2, 1998, DNA was extracted from samples taken from the condom. On February 3, the condom DNA was checked for yield and then amplified with primers for the FES/vWA duplex and the Green I system. On February 4,

[5] See, Sayrafiezadeh, M. "The Birthday Problem Revisited." *Math. Mag.* **67**, 220-223, 1994.

the condom DNA was analyzed with a gel-based genetic analyzer (ABI 373). A second run, using slightly greater quantities of DNA, was performed on the same instrument the following day.

Meanwhile, the bib and track pants, and samples extracted from them, were in the laboratory. These items were initially examined on either February 2 or February 3, 1998. DNA was extracted from them on February 4. It is difficult to determine the exact timing of these events from the laboratory notes, but it appears that the same personnel processed the bib and track pants and that these items were either processed together (simultaneously) or consecutively, which makes it plausible that they could both have received contamination from the same source.[6] It is significant that DNA was extracted from the condom the very day (or perhaps the day before) the bib and track pants were first examined, and that DNA from the condom was being tested in the laboratory on the same day that DNA was extracted from the bib and track pants. Given the close temporal proximity, it is possible, as Sally Ann Harbison has suggested, that genomic DNA extracted from the condom could have contaminated a surface or instrument that later came into contact with the bib and track suit. Professor Krane and Mr. Taylor concur that the bib and track pants could have been contaminated in this manner. (I wish to make it clear that my conclusion on this point rests heavily on conversations with Professor Krane and Mr. Taylor, who are more knowledgeable than I about possible mechanisms of DNA transfer).

Mr. Scheffer points out that there were two contributors to the DNA on the condom and that DNA from only one contributor was detected on the bib and track pants. However, this finding does not rule out the possibility that the DNA on the bib and track pants came from the condom. The condom samples (particularly Item 1i "Inside condom") contain significantly more DNA from the primary female donor than the secondary male donor. When DNA tests are performed on samples with low levels of DNA, such as those found on the bib and track pants, it is common for the alleles of a weak secondary contributor to "drop out" and become undetectable. I have observed this phenomenon in a number of cases that were tested using the ProfilerPlus system. When the samples contain sufficient quantities of DNA, the profiles of both contributors can be detected. When the samples are more limited, only the primary donor can be detected.

Mr. Scheffer also points out, quite correctly, that there is no evidence of contamination in negative control samples. Of particular importance are the reagent blanks that were

[6] Based on the notes I received, I was unable to verify Mr. Scheffer's report (July 30, 2003, p. 5) that the bib and track pants were examined on different days (February 2 and 3, 1998, respectively). Having looked carefully at the notes regarding the examination of these items, I can find no indication of dates. The bib and track pants are described on consecutive pages, suggesting that the bib was examined first, but there is no indication that I can see that any appreciable length of time intervened between the examinations of these items. The "Sample Continuity and Results" sheets that Mr. Jones prepared for these items are both dated February 3, 1998. With respect to other items, the dates on these sheets correspond to the dates the items were examined. Because I cannot confirm that the items were examined on different days or at widely different times, I am not persuaded by Mr. Scheffer's claim that contamination of these two items would have required two distinct laboratory errors.

created during the extraction of DNA from the condom and from the bib and track pants. If there was widespread or endemic contamination in the laboratory, it should have been detected in these blank (no DNA) samples, but when these samples were tested they produced no results. According to the laboratory notes, the reagent blanks of other cases that were processed around the same time also showed no evidence of contamination. These findings argue against the contamination theory but are not definitive. I have observed instances in which there has been accidental transfer of DNA between samples and the reagent blanks were not affected. The type of secondary DNA transfer posited by Dr. Harbison and Professor Krane would not necessarily affect all samples in a particular group or batch, but might affect only samples that were exposed to the contaminant at a particular time or place. I have discussed this issue with Professor Krane and Mr. Taylor, who both support this conclusion. (This is another instance in which my opinion rests partly on my consultation with them).

Moreover, the laboratory's use of control samples was less than ideal. The laboratory failed to take substrate controls from the bib or track pants. Substrate controls (i.e., samples of unstained material from an area adjacent to a biological sample) are not always employed but would have been useful in this case given the uncertainty that existed about the nature of the biological material on the toddler's clothing. Substrate controls could have helped the laboratory assess whether the DNA extracted from the samples was actually from the putative biological material in the stain (which appeared to be blood) or was from another source, such as a contaminant, that affected the entire surface of these items.

Additionally, it is unclear from the laboratory notes exactly when the reagent blank samples were created. In the rape case, for example, the samples from the inside and outside of the condom were assigned DNA Numbers 9771 and 9772. The reagent blank was assigned DNA number 9854, which suggests it may have been created at a later time (perhaps much later) than the condom samples and hence may not have been exposed at the particular time and place that contamination occurred. The same uncertainty surrounds the reagent blank associated with the bib and track pants.

Yet another problem is that the laboratory apparently lost the reagent blank sample associated with the condoms (DNA #9854). This sample was available for an initial round of testing in February 1998, but could not be located when the laboratory conducted a subsequent round of testing in November 1999 with the ProfilerPlus system.[7] Hence, there was no opportunity to check for contamination of the sample with the ProfilerPlus system. Because the sample had been tested earlier with the Green I system, and reportedly produced no results, it seems likely that it contained no detectable amounts of extraneous DNA. But this conclusion is not definitive because ProfilerPlus may well be more sensitive than the procedures used earlier. (For ProfilerPlus the

[7] A reagent blank associated with the reference sample from the accused man in the rape case was also apparently misplaced. The fact that these important samples went missing raises concerns about the evidence control procedures of the VFSC.

laboratory used a new genetic analyzer that employed capillary electrophoresis; for Green I, the lab used an older gel-based system).

Finally, there is a gap in the laboratory notes I received with respect to the test results on reagent blanks. The reagent blank associated with the bib and track pants was reportedly typed with the ProfilerPlus on December 30, 2002, but I received no laboratory notes regarding the results of this test.

Issues of Interpretation and the Disclosure of Electronic Data

At the beginning of this report I say there are two possible explanations for the findings *assuming that the DNA tests were accurate and were interpreted correctly*. Before concluding the report I feel compelled to point out a significant limitation in my ability to evaluate whether the tests in these cases were in fact accurate and correctly interpreted.

When I review the casework of forensic DNA laboratories in the United States, I typically ask for and receive copies of the electronic data collected by laboratory instruments during DNA testing. These data show the actual results of the DNA tests. By examining these data, an independent expert can check whether the forensic laboratory interpreted its results correctly and can detect a host of potential problems that would not be apparent from examining the laboratory notes. Electronic data would be particularly helpful in evaluating the present cases. Beyond allowing a cross check of the analysts' interpretations, it would allow a detailed independent assessment of whether there is evidence of low-level contamination in any samples. It would also allow an independent check of whether, for example, there is trace evidence of the profile of the secondary male contributor (whose DNA was on the condom) in the bib or track pant sample, which would support the theory of cross-contamination.

I do not question the sincerity or honesty of the VFSC analysts. From extensive experience in the United States, however, I have learned that experts sometimes differ in their interpretation of DNA testing data. Disagreements can arise from honest differences over appropriate standards and from the fact that different analysts approach data with different expectations. It is well known that what one expects or desires to see can shape one's interpretation of ambiguous data.[8] In my academic work I have demonstrated that analysts' interpretation of DNA test results can be influenced, quite unintentionally, by such factors.[9] Accordingly, I believe it is very important that the

[8] See, Risinger, D.M., Saks, M.J., Thompson, W.C. & Rosenthal, R. The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion. **California Law Review**, 90(1) 1-56 (2002).

[9] Thompson, W.C. Subjective interpretation, laboratory error and the value of DNA evidence: Three case studies, **Genetica,** 96: 153-168 (1995); Thompson, W.C. Accepting Lower Standards: The National Research Council's Second Report on Forensic DNA Evidence. **Jurimetrics** 37(4) 405-424 (1997); Thompson, W.C. Examiner Bias in Forensic RFLP Analysis. **Scientific Testimony: An Online Journal**, **www.scientific.org** (1998).

original DNA testing data (and not just the conclusions recorded in the laboratory notes) be available for independent review.

Professor Krane and I have both requested copies of the electronic data from the two cases. As yet, we have not received it. According to Ms. Orr, the laboratory reported difficulty retrieving the electronic data from a tape archive system. In my opinion, forensic laboratories should take care to preserve electronic data in a manner that is easy to retrieve and disclose to independent experts. CD-ROMS are an excellent way to preserve and transfer such data. Because they are inexpensive, cost should be no barrier to full disclosure of electronic data. Full disclosure allows a thorough independent review of results and thereby helps assure that the underlying scientific methods are strong and appropriate, and that the laboratory's interpretations are fair and accurate. Until the electronic data are made available, I cannot say that I have performed a complete review of the underlying evidence in this matter.