The Discomfort of Thought
—a discussion with John Butler

Several years ago, we began to keep a list of topics that we thought would either be worthwhile topics for a Proceedings (POL in our vernacular, for Proceedings of Lunch, capitalization optional), or just fun to talk or write about. Recently we added *discussion with John Butler* to the list. Although one of us (NR) has had sporadic conversations with John over the years, we've never actually had the opportunity to share a meal. Fortuitously, all three of us attended the recent CAC meeting in Sacramento (We don't think we provided Mr. Houde with any photo ops, but there were reliable witnesses), and were able to huddle around the salad and other lunch offerings to at least begin this session. John has indicated that he routinely reads the *CACNews*, including this column. And he expressed some fascination with the process of how these Proceedings actually come about. What better way to find out than to participate in one? We agreed to present him with a list of questions to

What, we wonder, was the impetus for the SWGDAM 2010 Autosomal STR Interpretation Guidelines? What was wrong with the previous SWGDAM guidelines? Or what needed updating? John responds by saying that the Quality Assurance Standards (QAS) were, after a decade hiatus, revised in 2009. It was felt that the SWGDAM STR Interpretation Guidelines should also be updated to include more information and specifically to aid with mixture interpretation. The previous SWGDAM STR Interpretation Guidelines were released in 2000 and were very general. The 2010 guidelines expanded the text from 4 pages (1066 words) to 28 pages (9862 words) but followed the same general format. More information was needed on mixture interpretation and statistical approaches as the 2000 guidelines only had a few sentences on these topics without any real detail.

> "For the greatest enemy of truth is very often not the lie – deliberate, contrived and dishonest – but the myth – persistent, persuasive, and unrealistic. Too often we hold fast to the clichés of our forebears. We subject all facts to a prefabricated set of interpretations. We enjoy the comfort of opinion without the discomfort of thought."
>
> —*John F. Kennedy*

stimulate the discussion, and then to schedule a phone conference so that we could actually dialogue. John had a question of his own: don't we have to be eating lunch, he wanted to know? Isn't that a requirement to participate? We made him affirm (We don't think he swears) that he would be consuming some food while he typed the answers to our emailed questions, thus preserving the namesake integrity of the column.

We note that John wears many hats, and one of them is as unofficial spokesperson, if not ambassador, for SWGDAM (the Scientific Working Group on DNA Analysis Methods, caps required). So we begin with a few questions about this group and its body of work, which naturally segues into other relevant topics.

Having read and pored over these guidelines for almost 20 years now (the first TWGDAM guidelines were issued in the early 1990's), we wondered what went on behind the curtain; what sort of smoke-filled room and other prestidigitation is required to produce such a document?

At the NIJ Conference in June 2010, John introduced the SWGDAM Autosomal STR Interpretation Guidelines[1] to the forensic DNA analysts in attendance, wherein he described the process for creating SWGDAM guidelines:

---

[1] The slides are available on the NIST STRBase website at http://www.cstl.nist.gov/biotech/strbase/pub_pres/NIJ2010-SWG-DAM-Guidelines.pdf.

Photo © Norah L. Rudin

- Recognized need or request for guidance on a particular topic received
- A committee is formed and individuals selected to participate (the committee selects a chairperson that directs the efforts)
- Committee works to produce a document
- Committee product provided to full SWGDAM for comment
- Committee revises document based on comments received
- Full SWGDAM group evaluates and discusses the document
- SWGDAM approves based on a membership vote
- Guidance document released to the public most recently through the FBI website *(Forensic Science Communications)*

John amplified during our conversation that, because most work is done only during semiannual meetings with some periodic phone conferences or WebEx meetings, it can take several years to complete this process, depending on the scope of the work. The most difficult part was getting everyone on the committee on the same page so that all used the same nomenclature to mean the same thing when discussing, as one example, thresholds.

One word that grabs our attention is "Guideline;" we frequently see confusion between guidelines and rules or standards. We asked John if he sees the 2010 SWGDAM guidelines as a general guideline or a strict set of rules? Or something else entirely? John believes that guidelines contain principles that, when followed, will lead to better laboratory protocols and individual analyst practice. Guidelines are not standards and should not be used as an audit checklist. In his mind, following properly written guidelines is good science and should make good sense. That being said, guidelines are, almost by definition, a work in progress, and thus will likely require further revision with time and additional experience.

Norah pursues that idea by noting that science is dynamic, while a set of guidelines is static. The moment guidelines are written, they are out of date. As examples, she cites the "ceiling principle" (which of course is neither) from NRC I; the NP calculation for database searches from NRC II; and the 2p "rule", which is really a shortcut, and of course not a rule or even a guideline. And so she presses John: how should labs and analysts proceed in the years in between SWGDAM guideline updates? Is there a way to use and appropriately justify new procedures and protocols?

John points out that any active scientific field is always going to progress and improve, but sagely acknowledges the practical dilemma; the challenge is in dealing with that change while trying to conduct casework. This situation is similar to purchasing a computer. You still have to use the computer for practical reasons long after better and faster computers are released. A key issue is that a set of guidelines is not going to be able to cover every possible scenario. Therefore, thinking will be required based on an understanding of correct principles conveyed by appropriate guidelines.

Keith now asks what is the proper use, then, of these guidelines? John begins by stating that following correct (accurate) guidelines is good science, but Keith quickly interjects with a story. Conversing once with Jan Bashinski, he (Keith) used that phrase "good science," and Jan cut him off with, "Everyone thinks they are doing good science; that phrase carries no meaning without definition." So Keith asks John, what do you mean when you use the phrase "good science," and further, in what way is following guidelines good science? Aren't some guidelines just matters of policy, not necessarily matters

of science[2]? John, of course, has an answer. What he has in mind when using the phrase "good science" is that we should not convey meaning that is not directly supported by an analytical result, and that we convey meaning by providing data, rather than mere opinion. As an example, he stresses that we don't want to convey a false sense of the strength of a DNA result by failing to provide a statistic. Good science, in his view, uses data to communicate meaning to someone who is not a scientist, but who must nevertheless make a decision based on a scientific result that you, the analyst, have obtained. And the scientific means of doing that is via data, not opinion.

Norah, though, returns to the issue of thoughtfulness and understanding when it comes to the application of guidelines to casework. Why, she brazenly asks, do analysts follow guidelines without thought? Or, less accusingly, how do you think they follow them?

John responds with one of the most relevant quotes we have ever heard: While delivering the commencement address at Yale University in June 1962, President John F. Kennedy shared some valuable insights that John believes apply to DNA interpretation, particularly the interpretation of mixtures. President Kennedy said,

"For the greatest enemy of truth is very often not the lie – deliberate, contrived and dishonest – but the myth – persistent, persuasive, and unrealistic. Too often we hold fast to the clichés of our forebears. We subject all facts to a prefabricated set of interpretations. **We enjoy the comfort of opinion without the discomfort of thought.**"

Thinking, as discomforting as it may be, John insists, is required to perform mixture interpretation well. Analysts need to understand the principles that underlie mixture interpretation. In his training workshops, he typically focuses on principles, knowing (or at least hoping) that protocols and practice will improve if the basics are understood.

In contrast to his strong belief that interpretation requires thought, he regularly receives questions in which analysts solicit a "cookbook" response. What they often want is a simple recipe (protocol) that they should follow in order to ensure that they will, each and every time, bake the perfect cake (DNA interpretation and report conclusions).

So we put him on the spot a little; will he, we ask, provide an example where an analyst applied guidelines inappropriately, without thought or understanding? In his typical circumspect manner, John starts with a general example of a guideline that seems to be frequently misunderstood or misapplied:

SWGDAM Interpretation Guideline 4.6.2:

"It is not appropriate to calculate a composite statistic using multiple formulae for a multi-locus profile. For

---

[2] The establishment of an analytical threshold is an example of a guideline based more on policy than data. A 1998 article (Wallin et al., TWGDAM validation of the AmpFISTR blue PCR Amplification kit for forensic casework analysis. *J Forensic Sci* 1998;43(4):854–870.) suggests an analytical threshold of 150 RFU based on the desire of the authors to obtain a full DNA profile, and their data indicating that a specific amount of DNA coupled with a threshold of 150 RFU would achieve this goal. That paper clearly established, however, that DNA alleles could be detected far below that threshold. Subsequently, many laboratories adopted 150 RFUs as their threshold without regard to whether the goal of the analysis was a full profile or the detection of any and all real DNA alleles. A policy based on the arbitrary desire to report only full profiles should not be confused with or substituted for an empirically validated detection threshold designed to distinguish true signal from noise.

example, the **CPI** and **RMP** cannot be multiplied across loci in the statistical analysis of an individual DNA profile because they rely upon different fundamental assumptions about the number of contributors to the mixture."

He goes on to state that apparently many analysts simply follow a protocol written by their technical leader without understanding the validation studies upon which it (hopefully) relies. John further believes that analysts frequently fail to appreciate that guidelines based on single source or two-person mixtures don't apply to more complex mixtures with three or more contributors (more on this issue of complexity later). In particular, he notes that allele sharing can increase the height of STR peaks to a greater extent than might be predicted from the simpler samples, negating interpretation guidelines that depend on peak height ratios. One consequence of this might be a mis-estimation of the number of contributors. He emphasizes that treating loci within a profile differently, whether at the interpretation stage, or for calculating a statistic, is simply not supported by available data.

John then goes on to relate a question that he received, along with a portion of his response, as an example of failure to appreciate the scientific foundation for a guideline. The question asked:

In section 3.5.8 (Interpretation of Potential Stutter Peaks in a Mixed Sample), it is listed that 'If a peak is at or below this expectation, it is generally designated as a stutter peak. However, it should also be considered as a possible allelic peak, particularly if the peak height of the potential stutter peak(s) is consistent with (or greater than) the heights observed for any allelic peaks that are conclusively attributed (i.e., peaks in non-stutter positions) to the minor contributor(s).'

In your opinion, does this recommendation apply just to the statistical step of the analysis or does it mean that the stutter peak is to be considered a possible allele peak in the inclusion/exclusion phase of analysis? If we are doing unrestricted CPI, is it ok to estimate the number of contributors and delete a stutter peak if we are confident all contributors are already represented at a locus. Some are proposing that we delete the stutter peak at the inclusion/exclusion phase but then add it in when doing the CPI calculation. **Is it ok to have that disconnect between your interpretation and statistical method?** *Or if we have chosen to use CPI for our statistic should those rules apply to the interpretation of the sample (i.e. don't use assumptions of the number of contributors when determining alleles suitable for inclusion/exclusion.)*

In section 4.6.3, it is listed that 'When using CPE/CPI (with no assumptions of number of contributors) to calculate the probability that a randomly selected person would be excluded/included as a contributor to the mixture, loci with alleles below the stochastic threshold may not be used for statistical purposes to support an inclusion. In these instances, the potential for allelic dropout raises the possibility of contributors having genotypes not encompassed by the interpreted alleles'.

A portion of John's response included:

Your interpretation and statistical methods should have consistent assumptions and go together for each assumption being made (e.g., you may interpret a mixture under alternative sets of assumptions). Thus, in my opinion, you should be consistent with handling the stutter peaks. With use of unrestricted CPI and a peak at a stutter position below the stochastic threshold, the locus should be dropped from statistical consideration because it may be possible that allele dropout has occurred at that locus and a sister allele of the low level peak is missing. With the possibility of allele dropout at a locus, CPI is not an appropriate statistic at that locus.

John sees this exchange as an archetypal example of the failure to understand the very foundational knowledge upon which the guidelines were written. Keith adds that what the question illustrates is a lack of understanding of the limitations of the evidence, the test, or the interpretation. When analysts apply a statistical tool incorrectly, the disconnect can often be traced to a failure to understand and appreciate one or more of the aforementioned limitations. While CPI type calculations have been widely adopted due to their simplicity, many apply them to profiles for which they were never intended, thus pushing this simple tool far past where it should be used.

Keith continues on this theme by observing that understanding physical evidence analysis is much more about understanding limits rather than capabilities, both of the evidence and the test employed. Unless you have stressed your analytical system until it fails with known samples, you won't know when, with a real piece of physical evidence, you are standing on firm ground, when you are on thin ice, and when you are actually in the middle of a large body of water about to be permanently submerged. Norah provides an example; low template DNA methods and technology were developed without a concomitant development of interpretational schemata. She argues that we should not perform any type of analysis if we lack a theoretical framework for understanding the results that we get. John responds with two keen observations.

First, we should spend as much time developing our interpretation skills as we do our methodological skills. Technological progress (more sensitivity in detecting DNA, for example), can be a double–edged sword; without equivalent progress in interpretation skill, we are just as likely to cut ourselves as we are the target.

Second, John proposes the concept of a "complexity threshold," for which research theory and validation data establish the limits of our ability to interpret a result. He uses the stochastic threshold to illustrate his point; in most laboratories, this threshold is established by testing either single source or two-person mixtures. If we now attempt to apply that stochastic threshold to more complex data (three or more-person mixtures), we can easily be mislead because it does not take into account the additional layers of complexity. In other words, absent subsequent validation testing for three or more-person mixtures, we simply cannot support any conclusion drawn for results that are more complex than our validation data. We also discuss how this might be applied at the level of physical evidence. Should laboratories encourage or even accept gun and knife handle swabs, or other contact DNA samples, knowing that a high likelihood exists that they will struggle with interpreting the data?

As a member of the academic community, Keith wants to know whether more education might improve the situation? And if so, how? What kind of education would help? What do analysts need to know to use a set of guidelines thoughtfully and intelligently? He is quick to point out that it is not necessarily a matter of the degree conferred; many people know how to think, regardless of the piece of paper nailed to the wall in their office or study. And John's experience supports that notion.

A continuing frustration for John is that many analysts treat continuing education as a checklist; their agency sends them to a class, they sit there for 8 hours, with no feedback to determine whether or not they absorbed any of the material,

and everyone is content to check the accreditation requirement box for continuing education. He believes that some sort of assessment is required if such classes are truly meant to further the expertise of the analyst.

But Keith takes note of a large body of research on adult learners, especially work that emphasizes the need to produce life-long learners, defined as those who have learned how to teach themselves. Current work suggests that this is a matter of providing the appropriate challenge in the right atmosphere. The task of the profession is to engage forensic scientists as lifelong learners so that when given a new set of guidelines, they don't use them as a cookbook or as a checklist, but can successfully learn the concepts, theories, and experiments supporting the prose on the page in order to more competently analyze and interpret their evidence.

The training of a DNA analyst is the time when the attitude of learning and searching is best nurtured. Theoretical concepts such as experimental design would establish a foundation useful for reading literature on new technology and interpretation techniques, while further teaching and practice in the use of statistics should produce greater confidence when performing calculations on increasingly complex profiles. New analysts should also, as part of their training, familiarize themselves in detail with laboratory validation data. Why was a specific threshold selected—and based on what data? And what binder holds that data?

We collectively wonder what factors contribute to the seeming disinclination of forensic DNA analysts to continue deep learning. John provides an interesting laundry list of suggestions:
- Analysts are not engaged; they come to a class to be entertained rather than informed;
- Analysts want simple sound bites rather than in-depth explanations;
- Analysts believe that they are not allowed to fail, and yet they must be allowed to fail if they are to learn;
- Analysts are afraid of looking dumb in front of their peers; they would rather remain silent rather than ask a question that could clarify a concept for them;
- Some analysts are engaged and interested, but run into roadblocks back at the lab. "I can't implement this new technique because my Tech Lead won't let me;"
- A roadblock of fear also exists, including but not limited to:
  ◦ Fear of an inability to defend one's thinking,
  ◦ Fear of getting clobbered by a defense attorney who might question your approach,
  ◦ Fear of not being able to easily check all the boxes on the audit checklist, and possibly jeopardizing grant funding, CODIS participation, and accreditation.

John laments that the forensic DNA culture is not one in which members are encouraged to think deeply about what they are doing; they skate across the surface to get the job done, actually welcoming the ability to easily apply preconceived guidelines that don't require much thought. The guidelines, and the laboratory protocols written from them, are combed for the exact situation presented by an instant case, so that original thinking is not required. Keith suggests that the only pressure felt by caseworkers is to get the report out; any other pressures are squeezed out or ignored. The message heard is that the analyst can't afford to think about a problem or an issue for a week; results are needed today.

John reiterates that guidelines cannot be written to cover every situation. He looks at it as a pair of glasses; you want the best glasses to clearly see the data. The better the analysts understand the guidelines, the more efficient they will be in the laboratory. They won't waste time spinning their wheels, discussing some issue without ever understanding it, and therefore taking forever to get the case out. He continues that you need the optimal prescription to see the world properly and clearly. With the correct prescription, you suddenly realize what you weren't seeing – the fuzzy blurs become pinpoints of light. If analysts understand the guidelines better, they will be *more* efficient in the laboratory, and not waste time endlessly discussing what to do with results from a piece of evidence, or searching in vain for a solution to their situation in the protocol. When genetic, statistical, and forensic principles are understood more completely and fully, the process becomes more efficient, rather than less efficient.

We think we've taken up enough of his time, so we thank John for providing thoughtful responses, and ask whether we can prepare this into a Proceedings. His response:

"As long as I'm not eaten for lunch."

We promise.

<p style="text-align:center">◆➤✖◀◆</p>

Announcing in 2012...

Dark Side of Justice, a new novel by CAC member Raymond J. Davis, is due out early next year.

"Carl Bowman, a private forensic scientist in Seattle, Washington has been targeted by a secret cabal of cops investigating the Green River serial murder cases. They believe he has evidence implicating them for their failure to arrest the prime suspect, Gary Ridgway. Carl seeks safety in his ancestral homeland, Sweden, to wait out the threat and to take the opportunity to learn more about his heritage.
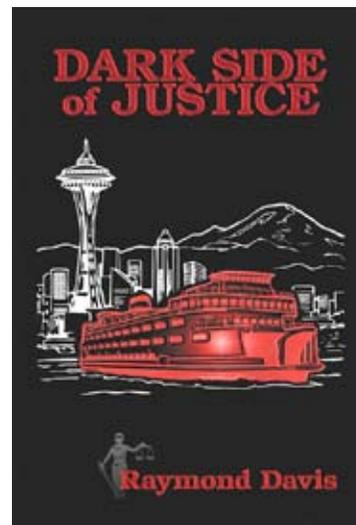
Carl's escape to safety lands him back into action when called upon to use his technical skills to solve an unexplained death at a European Conference only an hour's drive from his sanctuary. In the midst of the investigation, Carl receives startling news that will send him back to Seattle and a confrontation with law enforcement authorities.

Carl's odyssey to seek justice leads him instead to ultimately find the single most important thing in life."

RAYMOND J. DAVIS is a forensic scientist with over thirty years of experience in both private and government crime laboratories. He holds a degree in chemistry from CSU Sacramento. As the former editorial secretary of the CAC, he oversaw the publication of the quarterly journal, the *CACNews.* He also teaches law enforcement personnel in the techniques of effective courtroom communication.