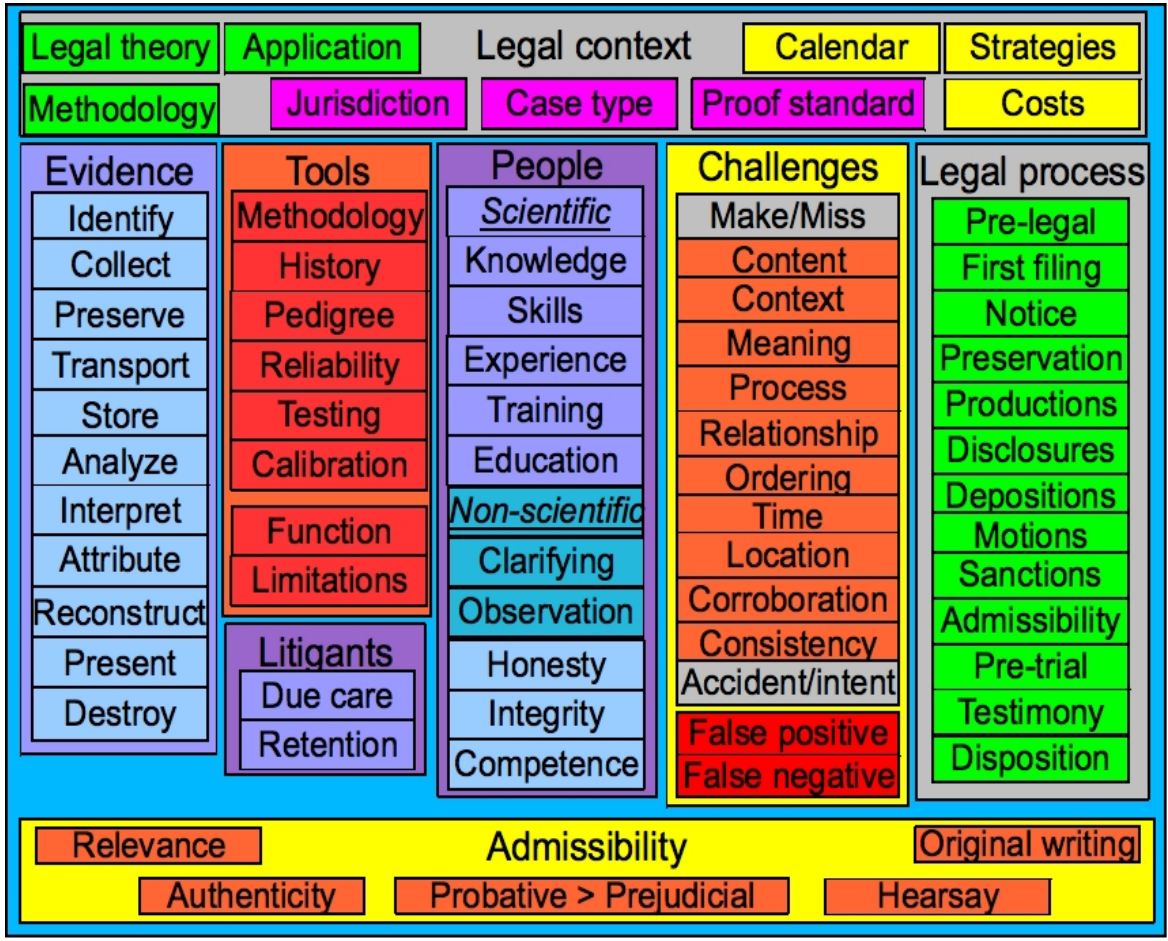
The physics of digital information and its application to digital forensics

Computer Forensics Show – Nov 2, 2010

Dr. Fred Cohen
President - California Sciences Institute
CEO – Fred Cohen & Associates

Outline

- Introduction
- A different physics
- Expert?
- Consensus?

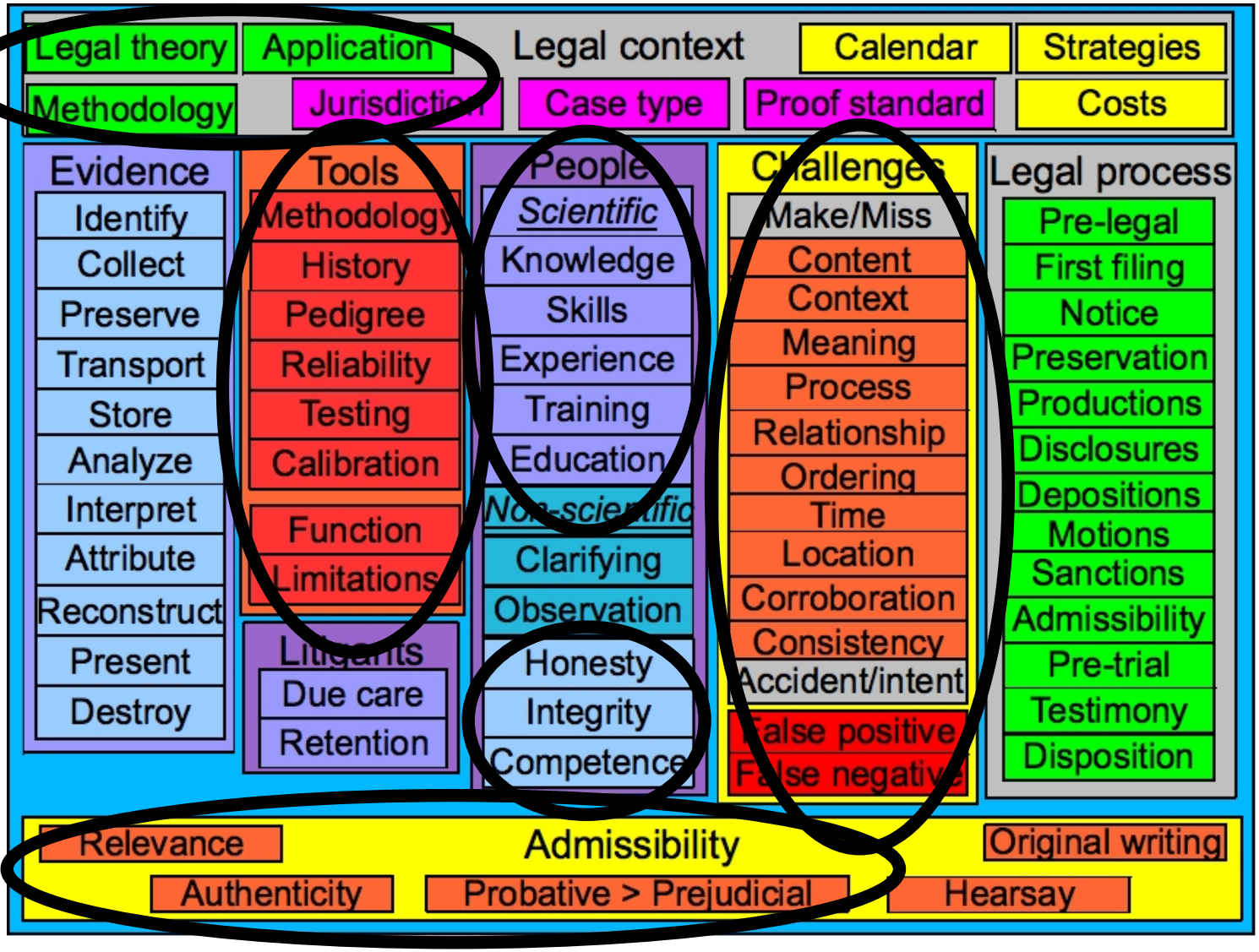


Your speaker

- CEO - Fred Cohen & Associates / President CalSci
  - Enterprise information protection architecture
  - Digital forensics for (usually high-valued) legal cases
  - 501(c)3 research and educational institution
  - M.S. Advanced Investigation / Ph.D. Digital Forensics
- B.S. EE (C-MU '77), M.S. Info Sci (Pitt '81), Ph.D. EE (USC '86)
- >30 years of information protection R&D, design, engineering, testing, implementation, operation, etc.
- >20 years since first digital forensics case
- POST certified instructor in digital forensics, Guest lecturer FLETC, PMTS Sandia National Labs, etc.
- >>100 peer reviewed publications, many conference talks, …

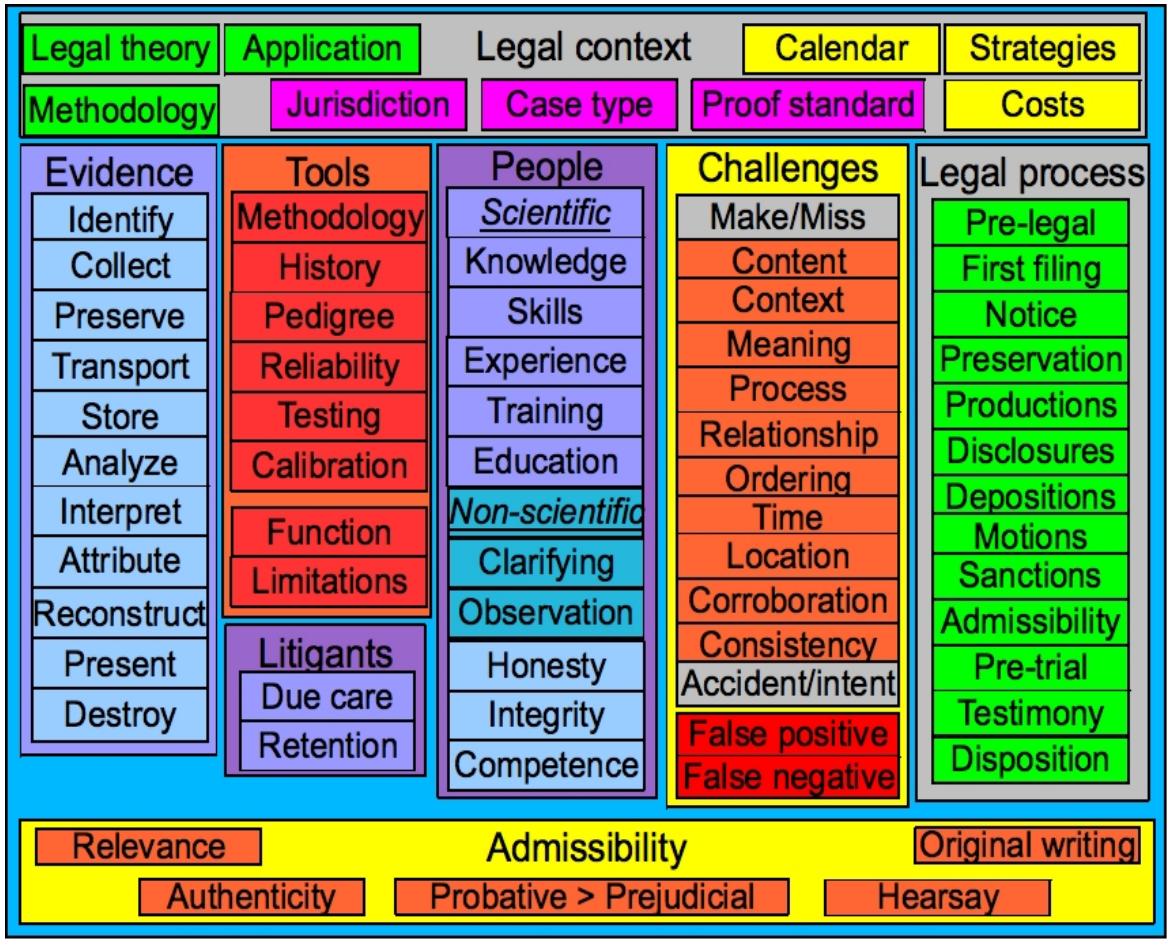
The issue at hand

- A scientific methodology properly applied?



Outline

- Introduction
- A different physics
- Expert?
- Consensus?



Differences in physics

- Some basic physics of digital evidence:
  - **Digital evidence** is entirely **sequences of bits**
  - The atomic unit is the "bit"
  - Nothing smaller (finite granularity)
    - No longer dealing with the digital evidence
    - Smaller than a bit it's physical evidence
  - Finite bit granularity → finite time granularity
    - Bits can only store traces (of time) at finite granularity (a finite bit sequence)
- **Normal space: infinite** granularity space/time
- **Digital space: finite** granularity space/time

Challenge!!!

- Finite granularity → time is a partial ordering
  - A before B (A<B), A after B (A>B), Can't tell (A≈B)
  - Traces as recorded are subject to Δt
    - What is the Δt for your traces / time stamps?
  - Is the claim a sequence of events?
    - Don't know Δt → don't know the sequence!
- Precision vs. accuracy
  - Trace time stamps are subject to delays, etc.
    - They look precise (2010-11-02 03:34:54.455)
    - But often aren't as accurate (off by 9 hours)
  - Mixed granularity misleading as to sequences
    - Some Windows time stamps at 1-day granularity

Differences in physics 2

- Observation without alteration:
  - **Normal space: Not possible** to observe a physical particle without altering it
  - **Digital space: Possible** to observe a bit without altering it - because the media storing bits is highly stable and engineered for this purpose.
- Duplication without removal:
  - **Normal space: No "exact" duplicates.** When we steal something, the original is gone.
  - **Digital space: Exact duplicates:** We can "steal" bits leaving the original intact and unaltered.

Challenge!!!

- Courts have held bit-for-bit copies acceptable as original writing for digital evidence - BUT:
  - A scientific basis is required to demonstrate that the duplication was properly done
- FRE 702: sound methodology properly applied
  - An underlying digital physics
  - Proper use of properly functioning tools
- We don't have a widely accepted and uniformly applied way to do this today
  - Each instance is a possible challenge
  - Each expert better understand it all

Differences in physics 3

- **DFE** is "**trace**" evidence
  - Finite State Machines (FSMs) execute
  - They produce outputs that may get stored
  - Stored outputs are "**traces**" of the event sequences in the FSMs
- **DFE** is **NOT "transfer"** evidence
  - **Normal evidence:** Two objects touch → each leaves part of itself with the other
  - **Digital evidence:** systems in "contact" with each other, do NOT leave parts
  - Systems **may** independently produce (different) traces as a result of "contact"

Challenge!!!

- Most digital forensics folks are unaware of the history of **natural world** forensics
  - **Natural world:** 1900 or so, **"transfer"**
  - **Transfer** is the scientific basis for trace evidence
- There is **no transfer in digital evidence**:
  - The **scientific basis** for evidence acceptance in the natural world **does not apply**
- But there are still **traces**
  - Products of the execution of FSMs
  - The basis for admission and use is different
  - Does your expert understand these principles?

Differences in physics 4

- FSMs have "perfect" forward predictability.
  - Given an FSM, initial state, and input sequence, all state and output sequences are precisely defined
- Thus **digital space "converges" with time**
  - **Normal space** admits to only one past but many possible futures.
  - **Normal space "diverges" with time!**
- **Many FSMs and input** sequences produce **identical output** sequences
  - Traces do not uniquely identify how they came to be!

Challenge!!!

- Suppose an asserted expert says:
  - Based on digital traces alone, a specific event sequence definitely happened
- But digital space converges with time:
  - Traces do not uniquely identify how they came to be!
- This is not a valid expert opinion
  - And it puts the expertise in question
- Be careful… precise wording is important
  - If you don't understand the physics, its easy to screw up

Information physics details

- Digital space converges with time
  - FSM: (I,O,S,m:{IxS}→{O,S'}) IF |I|>(|O|+|S|) THEN ∃(i,i')∈I:∃(o)∈O,∃(s)∈S, i→(o,s) and i'→(o,s)
  - Also note that $h(O) \leq h(I+S)$ (Shannon's h)
  - **Normal space diverges with time** *(2nd law of thermodynamics)*
  - **Digital space converges with time**
- You can't normally identify I^n from traces T
  - T: $|T|<|I^n|$, $\exists(i,i')\in I^n:\exists(t)\in T$, i→(t) and i'→(t)
  - In digital space, history is not uniquely determined by the present

Differences in physics 5

- FSMs are syntactic in nature
  - Semantics is driven entirely by context
  - The same sequence of bits can "mean" a lot of different things
  - Different sequences of bits can "mean" the same thing
- This means that "interpretation" is required for any meaningful use of digital evidence
  - There are a very large number of possible interpretations
  - But few of them are consistent, which is key

DFE scientific methodology

- The fundamental theorem of DFE examination:
  - **What is inconsistent is not true**
- DFE examination consists of testing hypotheses to try to refute them.
  - No matter how many tests are performed, except for special cases, ***you can't prove that any real world event is true***.
  - The ***best*** you can do, is show that your ***tests failed to refute*** the ***hypotheses*** at issue.
  - The ***most*** you can say (in proof) is that the ***results*** of the tests you did were ***consistent with*** some set of ***hypotheses***.

Refutation is king

- On the other hand...
  - One refutation disproves a hypothesis.
  - The ***least*** you can say based on refutation is that the ***hypothesis is not true***.
- Thus the methodology consists of:
  - Devise testable hypotheses (A ***consistent*** with B)
  - Test those hypotheses against the evidence
    - A scientific test should seek to refute a hypothesis and not to confirm it (seek ***inconsistency***)
  - Inductive and deductive logic are valuable tools for testing hypotheses
  - As is experimental technique

Differences in physics 6

- DFE is (normally) latent in nature
  - It can't be directly observed with human senses
  - The bits must be observed through tools
- How do we understand and trust the tools?
  - Most tools are computer programs (sequences of bits interpreted by FSMs)
  - How do we assess and present tool reliability?
- Most examiners today don't discuss this
  - But the Supreme court seems to think this is not up to snuff for other sorts of evidence

Challenge!!!

- DFE is latent → depends on tools
  - FRE702:"product of reliable principles & methods"
  - What are the principals and methods of the tools?
  - How reliable are the tools?
  - What are the limits of the tools?
- A scientific methodology to evaluate tools?
  - No methodology → regardless of what the tools tell us, we don't know how to interpret it
- What is the basis for trusting your tools?
  - In most cases, no basis is provided
  - Do you know the principals and methods?

Does your expert do this?

- How reliable?
  - What sort of errors are made by the tools?
  - To do this, we need an error model
    - See "Challenges to Digital Forensic Evidence"
- How do we calibrate and test tools?
  - Calibration → validation with known samples
    - What known samples are right for the matter?
    - What is the "right" answer and how do we tell?
  - Testing involves software verification
    - Mathematical proofs
    - Tests against error models

Even if the tool was "perfect"

- FRE 702: "the witness has applied the principles and methods reliably to the facts of the case"
  - Tools must be properly used w/in their limits
  - Results must be meaningfully interpreted
  - This implies relevant examiner knowledge, skills, experience, training, education
- A theory of measurement is needed:
  - What does the tool measure? How does it do it?
  - Do I need / can I use the same tool to test it?
  - Can I use a tool that doesn't reveal mechanisms producing its outputs?

Differences in physics 7

- **Normal space is limited by the speed of light**
  - Speed of light (c) ~186,000 mi/s ($3*10^8$m/s)
  - Matter can't be accelerated past c
  - Light and signals travel no faster than c
- **Digital space is also limited by c!!!**
  - Digital systems exist in the physical world
  - So these physical constraints apply to them
- **Digital space** and computational complexity
  - Computational complexity limits what operations can be performed with what computing capacity in what time frame: another "c" for digital space

Case 1 – The speed of light

- 2009 – civil matter – California
  - Respondent's "expert" claims that "When it takes **longer than a millisecond** to transfer the message from one server to another, it typically **means** that there is a **massive amount of information being sent**"
  - In this case, transfers were from **San Diego to San Francisco**
  - The distance from San Diego to San Francisco is in excess of **500 miles.**
  - The speed of light ~186,000 miles/second
  - 500 miles **= 2.68 milliseconds** – **OOPS!!!**
  - The "expert" never testified…

Case 2 – Other speed of light

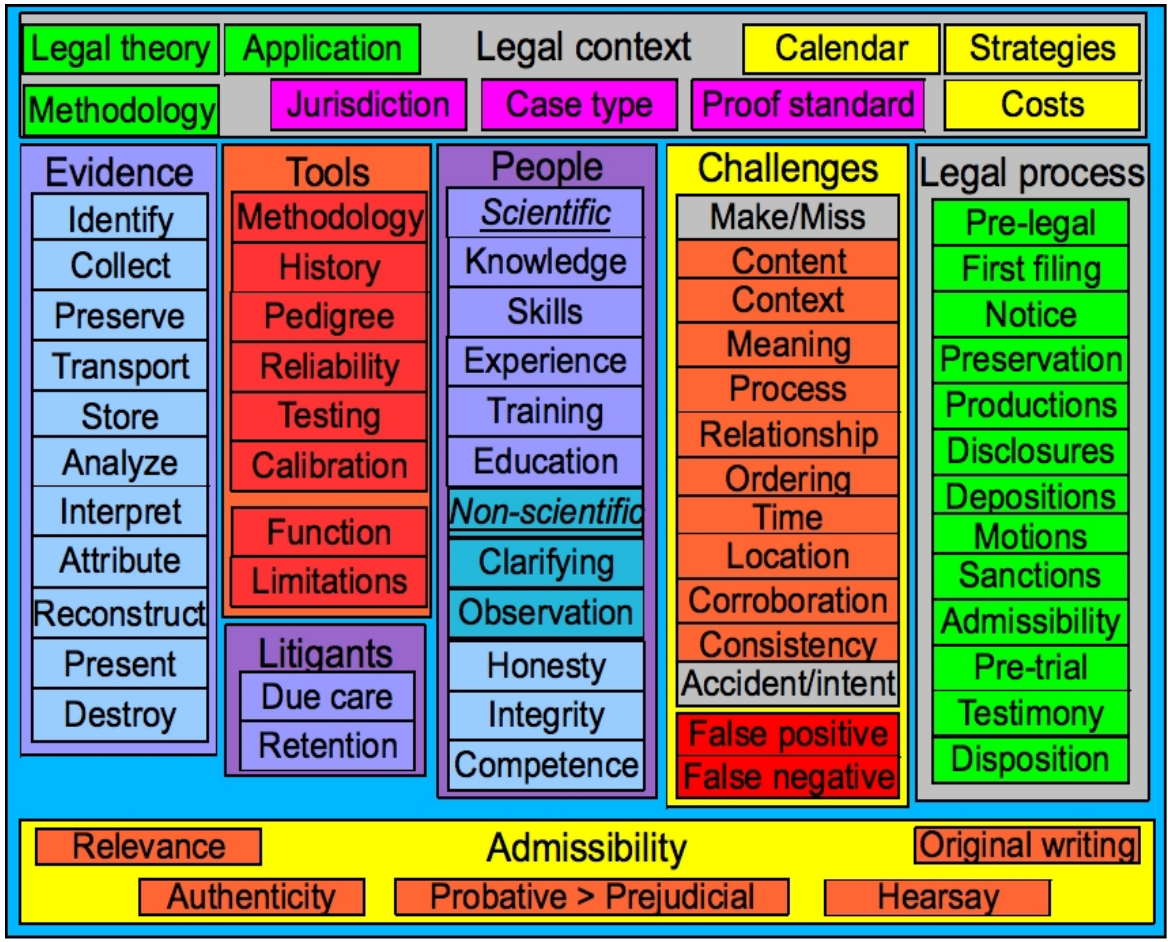
- 1989? – administrative hearing – Illinois
  - Plaintiff "expert" claims that audit trails were altered to remove all traces of a break-in.
  - In this case, the lack of traces was the claimed basis for the break-in.
  - Computational complexity of making the identified changes while retaining the consistency of other redundant audit trails was analyzed.
  - Complexity of making the asserted alterations was too high to allow it to be done with existing computer resources at that time.
  - The matter was dismissed.

Differences in physics ...

- There are many more examples of differences between the physics of digital information and the physics of the natural world.
- For details see:
  - F. Cohen, "Digital Forensic Evidence Examination - 2nd ed.", ASP Press, 2010

Outline

- Introduction
- A different physics
- Expert?
- Consensus?



How to tell a real expert?

- It takes one to know one...
  - But the courts seem to have a way...
- FRE, Daubert, and subsequent case law seem to have something to say on it...
  - Education, Knowledge, Experience, Training, and Skills
- Admissibility of their "opinions" (FRE 702)
  - (1) based on sufficient facts or data
  - (2) product of reliable principles and methods
  - (3) applied reliably to the facts in the case

Case 1 on expertise

- 9th circuit (testimony about proposition 8)
  - Perry, et. al. v. Schwarzenegger, et. al. US District Court Northern District of California, # C 09-2292 VRW
- Credibility determinations:
  - Plaintiffs called 9 expert witnesses, each with advanced degrees in relevant fields, published papers or studies on relevant topics in peer reviewed publications, each with years of experience in their fields, many acting as professors at universities.
  - All were accepted by the courts as experts

Case 1 (continued)

- Credibility determinations (continued):
- Proponents called 2 “expert” witnesses: (#1)
  - B.A. in social studies, M.A. in comparative social history. Worked as a community organizer, published 2 books, edited 4 books, co-authored others on related issues.
  - Publications challenged for lack of traditional peer review. Expertise challenged for non-relevance to relevant fields. (non-determinative)
  - Ruling: “None of … opinions is reliable.” … only “bald assurance” … “not supported by the evidence on which he relied”… “too great an analytical gap between the data … and opinion”

Case 1 (continued)

- Credibility determinations (continued):
- Proponents called 2 “expert” witnesses: (#2)
  - Ph.D. political science, Professor
  - Expertise challenged relative to issues in case.
  - Counsel provided most material considered, including >20 Web sites, did not know of many relevant issues, read few articles by other “experts” in the field, had not researched relevant issues, used results from non-scientific polls, failed to explain how data was consistent with conclusions, failed to review data used in his own report, was inconsistent between trial and deposition, etc. - Given “little weight”

Case 2 on expertise

- 6th circuit (testimony of a medical doctor)
  - 6th U.S. Circuit Court of Appeals in Tamraz v. Lincoln Electric Company (2010-09-08)
  - Ruing about "admissible opinion and inadmissible speculation under Rule 702"
  - Doctor testified that: "products triggered "***manganese-induced*** parkinsonism" in a welder who used them"
  - Verdict overturned – while the Doctor was qualified, the testimony was speculative because it was not adequately backed up by scientific research and results

Case 2 continued

- "That is a plausible hypothesis. It may even be right. But it is no more than a hypothesis, and it thus is not "knowledge," nor is it "based upon sufficient facts or data" or the "product of reliable principles and methods . . . applied . . . reliably to the facts of the case." FRE 702.
  - "speculative jumps involved in steps 4, 5 and 6 of this chain of causation..."
  - "step 4, ... literature hypothesizing a link between ... as "all theoretical."
  - "knew of no studies finding a link"

Case 3 on expertise

- An example in a lower court – the "Expert"
  - College degree in music
  - "Hacker" who broke into systems till caught and then "worked" for the FBI (w/out pay)
  - Systems administration experience ~5 years
- The expert never testified...
  - The "expert" report was ravaged by the other side's report indicating many serious flaws and downright errors.
  - After the expert on the other side testified, the case was settled.

How can you tell?

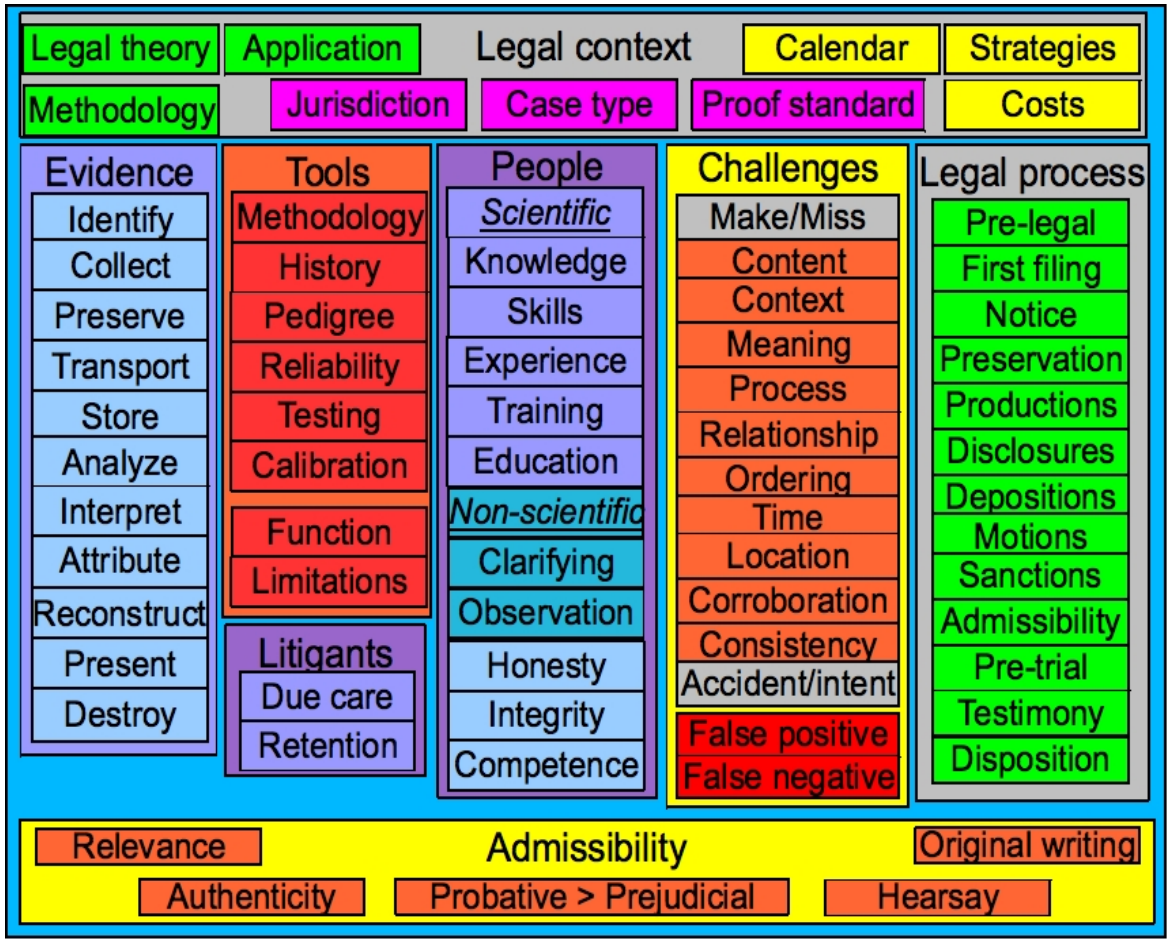
- Assess education, knowledge, experience, training, and skills like any other case
  - Ph.D. in a relevant field?
  - Published peer reviewed articles in the area?
  - Testifying in an area of their real expertise?
  - Years of experience in both research and practical work?
  - Professor in relevant university program?
  - Reputation from other lawyers / cases?
  - Experimental as well as theoretical work?
  - Honors, awards, keynotes, etc.?

How can you tell?

- Methods uses and how opinions formed:
  - Mostly facts presented that you can test without much expertise
    - Mostly: I did this – I saw that
    - Few “opinions” - limited – carefully worded
  - Mostly methods from peer reviewed sources
    - The detailed basis for opinions presented
    - No: “bridge too far” - Yes: “I don't know”
  - Look for certain “bad things”
    - e.g., No sound detailed bases in the report
  - Express tool/method reliability sensibly
    - History, pedigree, testing, calibration, limits

Outline

- Introduction
- A different physics
- Challenges
- Expert?
- Consensus?



The state of consensus

- The “scientific community” in digital forensics lacks consensus even around the very basic notions
- Compared to the consensus on human activity producing global climate change (86% or more) the basic notions of digital forensics are not at consensus levels:
  - Digital evidence is made of bit sequences.
  - You can observe bits without altering them.
  - You can duplicate bits without removing them.
  - Digital evidence is trace evidence

Challenge!!!

- There are only about 500 peer reviewed articles on digital forensics in the literature
  - Terminology is not widely agreed or uniformly applied – lots "made up"
  - Testability, validation, and scientific principles have not been widely addressed
  - The small corpus of published results limits the scientific basis for such statements
  - Claims w/out supporting experiments common.
- "The State of the Science of Digital Evidence Examination" - pending peer review for 2011 publication and presentation at IFIP

Thank You

http://calsci.org/ - calsci at calsci.org
http://all.net/ - fc at all.net