HAND DELIVERED
SGR: AEC REC'D BY: MRS
COPY RECEIVED
MAR 13 2002
TIME: 2:52
UNITED STATES ATTORNEY
Seattle, WA

Hon. Robert S. Lasnik

UNITED STATES DISTRICT COURT
WESTERN DISTRICT OF WASHINGTON
AT SEATTLE

UNITED STATES OF AMERICA,
Plaintiff,
v.
MICHAEL S. PRIME,
Defendant.

NO. CR 01-310L

AFFIDAVIT OF DEFENDANT'S EXPERT RE DAUBERT MOTION

COMES NOW, the defendant, Michael S. Prime, by and through his attorney of record, Lee Covell, and files herein the affidavit of Michael J. Saks regarding the Daubert hearing scheduled for Monday, March 18, 2002 at 10:00 a.m.

DATED this 12th day of March, 2002.

Respectfully submitted,

Lee Covell
Attorney for Defendant

119 FIRST AVENUE SOUTH, #500
SEATTLE, WASHINGTON 98104
(206) 682-6644

AFFIDAVIT OF MICHAEL SAKS

Michael J. Saks, being first duly sworn upon oath, deposes and states:

Professional Background

1. I am a Professor of Law and Professor of Psychology at the Arizona State University, where I am associated with the Center for the Study of Law, Science and Technology. My doctoral training emphasized research methodology and statistics. One of my principal areas of professional interest is scientific evidence, which I have taught in various forms and forums to law students, law professors, attorneys, and judges for the past 25 years. I am co-editor with David Faigman, David Kaye and Joseph Sanders of the treatise, Modern Scientific Evidence, 2nd edition (West, in press 2002), on which I am primarily responsible for the chapters on forensic science and research methodology. Of particular relevance to the case at bar, I also am co-author with Michael Risinger and Mark Denbeaux of several articles on the claims of expertise associated with handwriting identification, most notably, *Exorcism of Ignorance as a Proxy for Rational Knowledge: The Case of Handwriting Identification "Expertise,"* 137 U. PA. L. REV. 731 (1989). I have been an invited speaker at conferences organized by the National Institute of Justice and the Federal Bureau of Investigation, at which I outlined ways that forensic handwriting examination might, by undertaking appropriate empirical research, be able to place itself on a more sound and scientific footing. My curriculum vita is attached.

Overview

2. The conclusions of the proffered asserted handwriting expert witness or expert witnesses offered by the Government in this case, and of their field generally, stand on a weak or non-existent scientific (or technical) foundation. That foundation:

a. Embodies a detailed set of beliefs about the nature of handwriting (enumerated in Risinger, 1997, 2002). (Note: Articles and treatises are referred to briefly in the text of this affidavit and then are listed with fill citation at the end of the affidavit.) Some of these beliefs have been mentioned by Mr. Bottini in his "Analysis." These beliefs, although testable, have never been subjected to systematic empirical testing to determine which of them are valid and which are invalid. (Testing is a major factor to be considered under *Daubert*, at 593, and FRE 702(2)).

b. Consists of a remarkably thin body of published, peer reviewed research (where there is no research there can be no published research). (*Daubert,* at 593-594).

c. Suffers from high rates of error *(Daubert,* at 594), especially when dealing with hand printing.

d. Is lacking in standards (*Daubert,* at 594) because the technique is almost entirely subjective. Examiners look at a questioned and a known writing, apply no objective measures or standards, and when an examiner is intuitively comfortable with a conclusion that becomes the conclusion, which is to say that handwriting identification is essentially an exercise in *ipse dixit* (*Joiner,* at 137; FRE 702(3)).

e. Is "generally accepted" among those who earn their living performing handwriting examinations, but whether those beliefs and techniques are generally accepted in any wider relevant community – other forensic scientists, normal scientists who study handwriting using scientific methods, physiological psychologists, or statisticians – is either unknown or in the negative (*Daubert,* at 594). The discussion below will make clear why statisticians and normal scientists would not regard the claims of handwriting examiners to rest on a sound foundation. Indeed, two prominent forensic scientists reviewing a wide range of forensic sciences have offered the view that "[a]rguments that the identification of handwriting is a science are weak to the point of being indefensible" and that handwriting examination is "high" on vulnerability to a *Daubert* challenge (Thornton, 1997; Thornton & Peterson, 2002).)

3. Accordingly, it has been and continues to be my conclusion that, on present knowledge and methods, forensic handwriting examination fails to satisfy most, and perhaps all, of the validity criteria of FRE 702, as elaborated by the United States Supreme Court's decisions in *Daubert* (proof of a valid basis is a precondition for admission of an asserted expertise), *Joiner* ("Nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert"), and *Kumho Tire* (*Daubert's* evidentiary reliability requirement applies to all fields of proffered expert evidence, not only those denominated "science"). This is not to say that forensic handwriting examiners could not at some future time, after adequate research and development of their field, and depending upon what the results of that research revealed,

satisfy the validity requirements of FRE 702. But such work has only recently begun and remains in its infancy.

4. Recognizing the deficiencies outlined above, the National Institute of Justice has initiated a research funding program to facilitate the testing of the claims of asserted handwriting identification experts to determine which of the claims are true. (NIJ, 1999 ("Validation of the scientific basis for handwriting examination... is extremely important... priority must be placed on demonstrating the scientific basis for identifications claimed during handwriting comparisons.")

5. The limited relevant research that exists indicates that on various tasks forensic handwriting experts disagree with each other, have substantial error rates, and often are no better at reaching correct conclusions than laypersons. (These studies will be discussed below. Also see *Kumho Tire* (the focus of judicial gatekeeping is to be on the specific task-at-hand in the case rather than on a field globally and generally).)

6. Prior to *Daubert*, there was no reported judicial opinion in the United States that scrutinized the basis for handwriting examiners' claims of expertise. Wholesale admission was based on something other than a judgment about the validity of the field's claims. (For an extensive history of the pre-*Daubert* case law on admissibility of handwriting expert testimony, see Risinger et al., 1989.) Following *Kumho Tire*, forensic handwriting expert witnesses have regularly been limited or excluded entirely when challenged under *Daubert*.

7. The balance of this affidavit is devoted to explaining in more detail the scientific, empirical, theoretical, and statistical issues that have been summarized above.

Bases for the Claim of Handwriting Examiner Expertise

8. If the asserted handwriting expert or experts in this case were pressed to explain the basis of their conclusion about this (or any other handwriting identification task) – and assuming they based their explanation on the established literature of the field – they would, in substance, assert the following.

a. Forensic handwriting examiners assume that no two people write alike and that no one person writes exactly the same way twice. (The first of those statements posits infinite variability among writers; the second posits infinite variability within the writing of each individual writer.)

b. The assumed nature of between-writer infinite handwriting variation is based on the product rule of probability theory. ("This problem is capable of a mathematical solution if we can agree on the basis for the calculation.... We must first determine how often, or rather how seldom, each feature will be found separately and then, by a mathematical formula... we determine how often coincidence of all the features may be expected." Osborn (1929).) The belief that no one person writes exactly the same way twice assumes that within-writer variation is also infinite, but somehow is "less" infinite.

c. Handwriting examiners claim to have the skill to discern when two writings are so similar that they should be attributed to the same writer, and when they are sufficiently different that they should be attributed to the writing of two different persons, and that their ability to do this is superior to that of anyone not a member of their professional circle. (That is, not only do handwriting examiners all claim to be better than laypeople at making these observations and inferences, some groups of handwriting examiners claim to be superior to other groups of handwriting examiners. (*See* Moenssens, 1998, who identifies 55 different professional associations of handwriting examiners.)

d. Handwriting examiners offer as a primary indication of their expertise the assertion that all qualified handwriting experts always reach the same conclusion on the same writings. (E.g., FBI document section chief Ronald Furgerson asserted that all "certified" document examiners in the United States would reach the same conclusion as he would in any and every case. Fisher (1995), at 196.) (This is the claim of *reliability* of performance, that is, that handwriting experts' conclusions, right or wrong, are in complete agreement with each other.)

e. The asserted skill to make accurate comparisons and inferences is claimed to flow from the apprenticeship training of examiners and their later on-the-job experience. (Such apprenticeships provide the core training of handwriting examiners due to the absence of formal educational programs.) (This "training" is offered to support the claim of *validity* of performance, that is, that they produce correct results. The way this training typically works is that the apprentice works on cases along with his mentor and tries to reach the same conclusions as

his mentor does. After leaving the training phase, an examiner claims to benefit from the "experience" of working on cases of his own. The obvious problem is that during training the criterion of accuracy is not reality but the opinion of the mentor. And once an examiner moves beyond the training period, he has even less to test his accuracy, because the correct answer is rarely if ever known . Except on proficiency tests, to be discussed below.)

The Claim of Handwriting Examiner Expertise Remains Untested

9. Each and every one of the claims by handwriting examiners outlined above either remains untested after nearly a century of practice, or (of the few that have been tested) has been called into serious doubt by the research testing it. An exhaustive search for empirical research testing the claimed abilities of handwriting examiners found a grand total of one poorly conducted but published study by Inbau from 1939 and 5 unpublished proficiency studies (reviewed in Risinger et al., 1989). No studies overlooked by Risinger et al. have ever come to light. Several studies have been carried out since the Risinger et al. article, and which are summarized in Risinger, 1997, 2002).

10. Without exception, every published commentator agrees that the field of forensic handwriting examination has neglected to conduct research testing its claims, including commentators who explicitly set out to defend the field against the criticisms made by Risinger et al. (Galbraith et al., 1995: There is an "admittedly sparse history of carefully controlled empirical studies." "... there certainly has been a shortage of studies...." Kam et al., 1994: There is "an acute lack of empirical evidence on the proficiency of document examiners" and therefore "it is widely agreed that testing of professional document examiners and acquiring data on their abilities... are necessary." Moenssens: "Document examiners have not done the kind of empirical research that could have and should have been done.... On that the critics are absolutely correct." There is "indeed a dearth of published empirical information relating to the proficiency of document examiners....")

The Claim That No Two People Write Alike, and No One Person Writes Exactly the Same Way Twice

11. The real claim here is that inter-writer differences are infinite and intra-writer variation (though infinite) is within a small enough range, so that handwriting examiners can *reliably distinguish* all writers from all other writers. Document examiners have never had any basis for knowing whether this is correct or not, and can produce no data to support their contention. Their basis for this claim is nothing more than having been told it is true by their

mentors and from looking at a lot of writing themselves and seeing a lot of variability. In contrast to DNA examiners, for example, handwriting examiners have never collected systematic data on populations or subpopulations of writers and measured the variation in the writing of a large number of people. Unlike fingerprint examiners, for example, handwriting examiners do not have large databases of writings to which they refer in the course of making an examination in each case. They merely compare the questioned and the known writing given to them and make assumptions about the populations from which they came, or make inferences and guestimations about the population from what they can recall from other writing they've seen.

12. In fact, there has been one reported attempt to see how much variation there is, in signatures, and that study found the signatures of numerous different people named "John Harris" to be so alike that it was impossible for examiners to distinguish one writer from another. (Harris (1958) ("So many of these signatures lacked individuality and looked alike that they were not worth photographing.")) This extremely cautionary, if not devastating, study by a prominent handwriting examiner has simply been ignored by the field. In so doing, the field went from assuming the truth of the proposition without evidence to asserting the truth of the proposition in spite of evidence that the proposition is false.

The Assumption of Infinite Handwriting Variation Is Based Loosely on the Product Rule of Probability Theory

13. The principal basis for handwriting examiners' dogged belief in the uniqueness of the writing of different writers was offered by the father of the modern document examination field, Albert Osborn. Osborn adopted for handwriting the basic concept of the product rule of probability theory. (In doing so he was following the lead of Alphonse Bertillon, the father of forensic identification, who invented *anthropometry*, a once leading identification technique that is no longer in use.)

14. The concept is that if we know the individual probabilities of a set of independent attributes, we can calculate the probability of their joint occurrence by multiplying them together. As he explained in his book, QUESTIONED DOCUMENTS, 2nd ed. (1929), which is still the leading treatise in forensic document examination: "This problem is capable of a mathematical solution if we can agree on the basis for the calculation.... We must first determine how often, or rather how seldom, each feature will be found separately and then, by a mathematical formula... we determine how often coincidence of all the features may be expected."

15. One problem is that Osborn and his followers never progressed beyond using the theory as a loose metaphor for what they were doing. They take no measurements, they make no calculations, and they report no probability of coincidental matches. In other words, they do not apply the theory to the practice at all. (*Compare,* FRE 702(3).)

16. Another problem is that probability theory is incapable of supporting the claim of unique individuality. The mathematics and logic of probability theory can lead only to a probability value – not to unique, one-of-a-kind certainty. If the exercise suggested by Osborn were carried out, the end product of handwriting examination would be a specified probability of a coincidental match (that is, the probability that a person's writing selected at random from the population would have the same characteristics as the questioned writing, also known as the "random match probability") – rather than a bald assertion of identity. That probability theory is inherently incapable of leading to conclusions of unique individuality is obvious to mathematicians and statisticians, and has been recognized by some forensic scientists. (*See* Cummins & Midlo (1943): "It is unfortunate that this approach [the product rule of probability theory] carries the implication that a complete correspondence of two patterns might occur..." "... it is impossible to offer decisive proof that no two... bear identical patterns."; and Stoney (1991).) Also, see the everyday work of DNA typing experts, who calculate the probability of a coincidental match, because they know that is the most that can be done.)

17. Probability theory could, however, be employed to reduce uncertainty in trying to match the correct writer to a writing, or more correctly evaluate the likelihood that a signature was not made by the purported author. But handwriting examiners have never taken the trouble to do the research necessary to accomplish this. Instead of developing the basis for making objective probability estimates, they rely on what statisticians call "subjective probabilities" (personal guestimates of probability), or they simply act as if it is true that no two writings are indistinguishably alike (though Harris's research suggests this assumption is false.) Handwriting examiners have throughout this century recognized the possibility of moving beyond metaphor into actually measuring and calculating probabilities for their examinations, but they have repeatedly rejected undertaking the work needed to do that, and chose instead to remain subjective and intuitive (*see* Hilton (1982).)

18. The failure to adopt more objective, scientific, quantitative methods can be understood by contrasting what handwriting examiners do (they look and they opine) to the routine work of experts in DNA typing, who actually do have data on the probability of occurrence in the population of individual genetic systems, have tested them for independence, applied the product rule, and report the probability of a coincidental match (the random match probability, the probability of a false positive error). DNA experts do in actuality what handwriting experts do only metaphorically.

penmanship teacher

19. The failings of handwriting examination practice can also be understood by contrasting contemporary practice to the first and only time that handwriting was subjected to objective measurement and probability calculation – in the famous Howland Will Case of 1865, recounted recently in a magazine article by Menand (2001). That case dealt with a handwriting problem very much like the one in the case at bar. The 1865 case involved a claim to the substantial Howland fortune by the infamous Hetty Green (then Hetty Howland Robinson). Hetty was accused of forging the signature on the will upon which her claim relied. It was suspected that she had traced the signature from another of the purported testator's signatures. Renowned Harvard mathematician Benjamin Peirce and his (later) even more renowned son Charles Sanders Peirce were hired to examine the signature on the will. They devised a method which relied on defining the beginning point of downstrokes, of which there were 30 in the signature. Comparing 42 genuine Sylvia Ann Howland signatures, each to all others (which gave 25,830 comparisons), they determined that any given downstroke start position would coincide with any other approximately one time in five. They then compared the challenged will signature with the signature claimed to be the tracing model and found perfect coincidence of the start position of all 30 downstrokes. Using the baserate data they had gathered, they found the random match probability to be 1 divided by 5 to the 30th power, or one over 2666 followed by 10 zeros – an extremely small probability that the signature looked as it did owing to the normal variation in the testator's writing. In that same case there was no lack of handwriting experts who offered their subjective opinions on both sides of the question of whether the will was a forgery. But only the Peirces could offer real data and describe exactly how they arrived at it.

1784

20. In the present case, by contrast, the asserted handwriting expert or experts reached their conclusions concerning authorship merely by intuiting a subjective sense of the likelihood that the questioned writing was authored by the defendant. They cannot explain how their subjective probability judgment was arrived at, because it is entirely intuitive. (*See, McMahon v. Bunn-O-Matic*, 150 F. 3d 651, 658 (7th Cir. 1998) ("We have stated before and reiterate, that 'an expert who supplies nothing but a bottom line supplies nothing of value to the judicial process'.'"))

The Claim That Examiners Have Through Training and Experience Acquired a Special Ability to Discern Whether a Writing is Genuine or Not

21. Because there are few if any formal educational programs that teach the science or art or folklore of handwriting identification, the field relies on apprenticeships: an experienced examiner takes on an apprentice, and a new handwriting examiner is born. Such apprenticeship training of document examiners has been relatively informal, lacking in consistency, perhaps even idiosyncratic, and the final product of the training is poorly tested

before being certified as a handwriting expert. At most, what the examiner learns is to mimic the judgments of his mentor. Recent research from Australian document examiners, trying to figure out why error rates varied so greatly from one lab to another, found the most likely answer to be that high error rates cluster among examiners who studied under the same (high error rate) mentors and low error rates cluster among examiners who studied under different (low error rate) mentors. (Found & Rogers, unpublished).

22. Beyond the better or poorer intuitive skills of examiners, no training method or system can teach a handwriting examiner truths that have not been tested to discover whether they are indeed truths. Learning something from someone does not make the thing learned true. Unless a system is in place to test beliefs and over time to replace false hypotheses with correct ones, then the next generation of handwriting examiners may be in the possession of nothing more than the erroneous assumptions of the previous generation of handwriting examiners. Consider that Osborn's 1929 treatise, which changed very little from his 1910 edition, remains the core of handwriting examiner education, at least for the Orthodox Osbornian school, which is the faith of most government document examiners.

23. As to the claim of "experience," while it is true that many things can be learned from experience, it is equally true that many things cannot be learned from experience. It is important to inquire into what, exactly, the experience has contributed to the proffered expert witness' asserted expertise (see the discussion of experience-based expertise in the Advisory Committee's Note to revised Federal Rule of Evidence 702). In *Starzecpyzel*, Judge McKenna likened the asserted experience-based expertise of document examiners to that of harbor pilots, namely, that by doing something over and over one become good at it. But the analogy is inapt. Feedback is essential to the learning of many skills. In the course of his daily work, a harbor pilot learns instantly if he hits a sandbar or delivers a ship to the wrong wharf. A plumber usually can see immediately if he has failed to stop a leak. The same is true of many endeavors. From the feedback of their successes and failures people learn what works and what does not.

24. Not so with forensic handwriting examiners. Unlike harbor pilots and plumbers, forensic handwriting examiners receive no meaningful feedback concerning whether they have correctly or incorrectly linked a writing to its author, or whether a signature is genuine or not. It is precisely because only the writer knows whether he or she authored the questioned writing that the opinion of an "expert" is sought. Rarely if ever does the handwriting examiner receive feedback on whether he reached the correct conclusion or not. A person becomes better at playing darts by seeing how close each throw comes to the bull's-eye. But if the person were blindfolded, and had no idea how near or far to the bull's-eye each throw came, the person could not improve his skill no matter how much "experience" the person had.

Handwriting examiners are not in the position of harbor pilots. They are in the position of blindfolded dart players.

The Claim of Reliability of Performance

25. Handwriting examiners once asserted that they never disagreed with each other (*see* Todd, 1973; Fisher, 1995, at 196), the implication being that a lack of disagreement from other, equally qualified, examiners bespoke a high degree of accuracy in their conclusions. In reality, this is a claim of what researchers term "reliability" – consistency, sameness of response to the same stimulus.

26. In the wake of more than a decade of proficiency tests, in which many document examiners were given the same problems to which they often produced different answers, the claim of high reliability now should be heard less often. To take one example: In the 1985 Forensic Sciences Foundation document examiner proficiency test, examiners were asked to determine whether any of 12 checks all bearing a signature in the same name were signed by the same person. In reality, two had been signed by the same person, one was a freehand forgery, and a fourth was a tracing. Examiners gave a wide variety of answers in response, such that 41% were correct, 6% incorrectly attributed one of the forgeries to the real repeat signer, 31% could not reach a conclusion, and 22% gave answers that were substantially incorrect.

27. Even if document examiners did all agree with each other all the time, what might that prove? To a scientist there are two different aspects of measurement: *Reliability* refers to the extent to which a measuring device (including human handwriting examiners) produces consistent answers. *Validity* refers to the extent to which the measuring device produces accurate answers. (When the Supreme Court in *Daubert* wrote of "evidentiary reliability" the Court took pains to indicate that it was referring to something closer to what scientists label "validity.") Reliability (in the sense of consistency) is a necessary but certainly not a sufficient condition for judging the work of a field to be sound. It is one thing to ask whether a platoon of handwriting examiners all agree with each other on who wrote a document. It is something else to ask whether they were correct or not. The results of the 1984 FSF proficiency test dramatically illustrates that distinction. Examiners around the country were sent three letters containing bomb threats to be compared to each other and to exemplars from six suspects. In fact, two of the questioned letters were written by one person (whose writing was not submitted) and the third letter was written by one suspect simulating the writing of the other letters. Every one of the tested handwriting examiners failed to identify the author of the third letter. Thus, they were 100% reliable (all in agreement) but 0% valid (all wrong).

The Claim of Validity of Performance

28. At the end of the day, how often do handwriting examiners arrive at correct answers? Do they reach correct answers any more often than laypersons? There are two sources of answers to these questions, both severely limited in number and imperfect in design, but they are the only source of real knowledge that exists on the subject. One is the body of proficiency tests carried out by the Collaborative Testing Service, first under the supervision of the Forensic Sciences Foundation and later passed to the supervision of the American Society of Crime Laboratory Directors. The second is a series studies carried out by Moshe Kam and colleagues under contract to the FBI. All of these studies are discussed in detail in Risinger's chapter in MODERN SCIENTIFIC EVIDENCE (1st ed. 1997; 2nd ed. 2002). They will be summarized briefly below.

29. The proficiency studies suggest that handwriting examiners are more accurate in dealing with some kinds of writing and much less accurate with other kinds of writing. The simplest and most straightforward problems (two studies involving written letters in the natural hand of the writers compared to known exemplar writing of several suspects) produced correct results 89% and 52% of the time (erroneous answers 4% and 3% of the time). Tests involving the determination of genuineness of signatures ranged from 35% to 100% correct. The one test involving hand printing produced 13% correct and 45% incorrect answers. In a test involving a note said to be written by a teenaged vandal required comparing the crime scene note to samples from five teenaged suspects, none of whom was in fact the author of the crime scene note, examiners identified an innocent suspect as being the criminal 30% of the time. The most difficult task appears to be where examiners try to identify the author of a forgery, which led to 100% error.

30. Galbraith et al. (1995) reanalyzed the results of a series of the FSF proficiency studies in an effort to demonstrate handwriting examiner expertise by showing that they performed "significantly better than chance." The concept is puzzling. Who in any context would regard a performance that is merely better than sheer guessing to establish expertise? (For example, if a driver can remain on the correct side of the road statistically significantly more often than half the time, do we laud the driver as an expert at driving?) Astonishingly, the Galbraith et al. research found that in one-third of the tests, the examiners did no better than sheer guessing.

31. In these open proficiency tests, examiners may have been trying to avoid what they knew would be the most embarrassing of errors, namely false inculpation, by giving many "could not reach a conclusion" answers, and might offer this answer more often in tests

than they do in practice. This tendency toward inconclusive answers lowers the error rate, but also might account for the poor accuracy rate.

not a PDE recognized.

32. A serious related problem is that in real cases handwriting examiners often know what answer is expected or desired by the investigator or attorney who has engaged the document examiner, or suggest by other evidence in the case. (E.g., Dines, 1998 ("Before an attempt by the examiner to identify a handwriting, the investigator should consult and [obtain] as much circumstantial evidence as possible about the case.")) Though such information is extraneous to the expertise that is to be applied, the information exerts an inevitable pull on the conclusion reached by the examiner – especially where an expertise is overwhelmingly subjective and where the risk of a conclusion being shown to be assuredly wrong is virtually zero (unlike in proficiency tests, where the correct answer is known and will be revealed after the test is completed). In contrast, real scientists take pains to avoid exposed to context cues through the use of bias control procedures, such as blind or double blind testing. (The problem of context cues and cures for them in the forensic examination context are discussed in detail in Risinger et al., CAL. L. REV. (2002).

33. As unimpressive as the performance of handwriting examiners has been in proficiency tests, the possibility exists that they do better than laypersons do. Jurors or judges would be asked to make the comparisons by themselves if they had no expert to opine on authorship or genuineness.

34. Galbraith et al. were the first to make the comparison between experts and laypersons. They obtained the test materials from the 1987 FSF study and presented them to laypersons. The true positive accuracy rate of laypersons was exactly equal to that of handwriting experts (when comparing questioned writing to the known writing of a suspect who had in fact written the questioned note, both experts and amateurs reached the correct answer 52% of the time). On the other hand, in this particular problem, none of the experts tested identified an innocent suspect as the writer of the note, while 34% of the amateurs erroneously associated an innocent suspect to the crime scene note (the false positive error rate).

35. Kam and associates conducted four studies of this general question (Kam et al., 1994, 1997, 1998, and a fourth unpublished). The first three studies employed an examination task unlike anything that document examiners encounter in actual cases, and so is of uncertain relevance. These comparison problems were designed to test aggregate global skills, and not to allow isolation of, much less identification of, specific subtasks. Only the fourth, unpublished, study was designed to test something that might be a "task at hand" (*see Kumho Tire*) in an actual case, and that was to determine the genuineness of a signature.

36. In addition, problems with the design of each study led to the redesign of each successive study. For example, the first study offered the laypersons no incentives for accuracy and avoidance of inaccuracy, whereas the real stakes for handwriting examiners were unavoidably extremely high. The incentive scheme introduced into the second study differed from the inherent incentive regime for experts in such a way that at least some of the observed differences could have been produced by the incentive scheme rather than the differing abilities of the participants. The third study attempted to test whether different incentive schemes would produce different results, which led to confusing changes in performance levels.

37. These methodological problems call the findings of the studies into doubt. But if we temporarily disregard the methodological problems in the studies, what conclusions might they suggest?

38. The first study found experts to outperform nonexperts. Nonexperts, however, showed striking bi-modality, with the best nonexperts doing about as well as the experts. These distributional data, suggesting that a substantial subset of laypersons (with no prior experience and no motivation to do well) are as good as experts, were not provided in subsequent publications, so we can learn nothing more about this phenomenon. (I have asked Prof. Kam and the FBI to share the raw data with me, in accord with normal scientific custom, for further analysis. Although they initially promised to make the data available, they subsequently have refused to do so.)

39. The second study found the rate of true positives to be virtually identical for both experts (87.1%) and non-experts (87.5%) (a difference no greater than would be expected by chance), while experts committed fewer false positive errors (6.5% v. 38.3%). These results are similar to those of Galbraith et al.

40. Though the third study suggested that there were no differences as a result of differing incentive schemes, it also showed an overall 41% improvement in the ability of laypersons to avoid false positive errors. The latter finding has no apparent explanation.

41. The fourth study tested the ability of examiners and laypersons to determine whether signatures were genuine or not, much as in the case at bar. Experts and amateurs did not differ in their ability to detect forgeries (experts correct 96% of the time, laypersons correct 92% of the time, a difference no greater than would be expected by chance). Experts were better at discerning genuine signatures (correct 86% of the time) than amateurs (correct 70% of the time). Experts erroneously found forgeries to be genuine (0.5% of the time) less often than amateurs (6.5%). And experts erroneously judged genuine signatures to be forgeries (7% of the time) less often than amateurs did (26%). The meaning of

these findings is limited by the fact that the study suffers from some of the unsolved problems of Kam's earlier studies.

Application to the Case at Bar

42. For most of the past century, the claims of forensic handwriting examiners have been oversold and under-researched. On current theory and data, there is little basis for concluding that handwriting examiners can do what they claim to be able to do and much basis for believing that a large gap exists between what is claimed and what is actual. In particular, their theory of unique identification calls on probability theory to do what it simply cannot do, and their refusal to use actual data and compute actual probabilities means that their assertions almost certainly exaggerate the trustworthiness of their individualization decisions. What could and should be an objective and validated set of skills and methods remains an exercise in *ipse dixit*.

43. The body of empirical research on the performance of handwriting examiners is tiny by anyone's count. A field of purported science that has been practicing for a century would be expected to produce hundreds if not thousands of empirical studies conducted on their basic hypotheses, techniques and skills. Forensic handwriting experts can produce only a handful (which handful has been summarized above). Nevertheless, if one were compelled to apply the limited data to the case at bar, the following might be reasonable, though tentative, conclusions.

44. The FSF proficiency studies suggest that forensic handwriting experts make many errors, especially when dealing with hand printing (a test involving hand printing produced 13% correct and 45% incorrect answers).

45. If a questioned and a known writing do share a common origin, which is the government's theory in any case, where the claim is that the defendant is the author of the questioned writing, document examiners offer no help to the jury because they are no more accurate in this task than laypersons. (The study that comes nearest in its task to the one in the case at bar is that of Galbraith et al., which found that in evaluating the handwriting on a note, experts and amateurs both had a true positive accuracy rate of exactly 52%. If the task of the first three Kam et al. studies is considered have to involve a task comparable to the one in the case at bar, experts and amateurs both had a true positive accuracy rate of 87%.)

46. What looms over the limited research that exists is the massive absence of research. Thus, most of what one would like to know in order to determine whether an asserted

handwriting expert is likely to have reliable information to contribute to the task-at-hand in this case, beyond what a judge or jury can do without the examiner's help, remains unknown.

47. For the reasons given above, it is my conclusion that the proffered forensic document examiner testimony in this case fails to meet the validity criteria of FRE 702, as explicated in *Daubert* and *Kumho Tire*.

FURTHER YOUR AFFIANT SAYETH NAUGHT.

Michael J. Saks

SUBSCRIBED AND SWORN to before me this 7th day of March 2002.

Notary Public in and for

My Commission Expires: 11/7/02

OFFICIAL SEAL
KAREN S. BARNETT
NOTARY PUBLIC-ARIZONA
MARICOPA COUNTY
My Comm. Expires Nov. 7, 2002

References

Harold Cummins & Charles Midlo, Finger Prints, Palms and Soles: An Introduction to Dermatoglyphics 154 (1943).

J.E. Dines, Document Examiner Textbook (1998),

David Fisher, Hard Evidence (1995).

Oliver Galbraith, Craig S. Galbraith & Nanette G. Galbraith, The "Principle of the Drunkard's Search" as a Proxy for Scientific Analysis: The Misuse of Handwriting Test Data in a Law Journal Article, 1 Int'l J. of Forensic Document Examiners 7 (1995).

John Harris, How Much Do People Write Alike: A Study of Signatures, 48 J. Crim. L. & Criminology 647 (1958).

Ordway Hilton, Scientific Examination of Questioned Documents (revised ed., 1982)

Fred Inbau, Lay Witness Identification of Handwriting, 34 Ill. L. Rev. 433 (1939).

Moshe Kam, Joseph Wetstein & Robert Conn, Proficiency of Professional Document Examiners in Writer Identification, 39 J. Forensic Sci. 5 (1994).

Moshe Kam, Gabriel Fielding & Robert Conn, Writer Identification by Professional Document Examiners, 42 J. Forensic Sci. 778 (1997).

Moshe Kam, Gabriel Fielding & Robert Conn, The Effects of Monetary Incentives on Document Examination Professionals, 43 J. Forensic Sci. 1000 (1998).

Moshe Kam, Kishore Gummadidala, Gabriel Fielding & Robert Conn, Signature Authentication by Forensic Document Examiners (unpublished manuscript)

Louis Menand, She Had To Have It: The Heiress, The Fortune and the Forgery, The New Yorker, April 23 & 30, 2001, at 62-70.

Andre M. Moenssens, Handwriting Identification Evidence in the Post-Daubert World, 66 UMKC L. Rev. 251 (1998).