

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

UNITED STATES

v.

ERIC D. HORN

:

CRIMINAL ACTION NO. 00-946-PWG
: (Magistrate Judge Paul W. Grimm)

:

:

MEMORANDUM AND ORDER

At approximately 10:35 p.m. on June 28, 2000, Sergeant Eric D. Horn attempted to enter the Harford Road gate of the Army facility located at Aberdeen Proving Ground, Maryland. Officer Daniel L. Jarrell stopped Horn's vehicle for an identification check. As a result of his observations of Horn, Jarrell suspected that Horn was driving under the influence of alcohol, and he was detained and questioned. Three standard field sobriety tests ("SFSTs") were administered: the "walk and turn" test, the "one leg stand" test and the "horizontal gaze nystagmus" test.¹ As a result of his performance on these tests, Horn was charged with driving while intoxicated under Md. Code Ann., Transp. II § 21-902 (1999 Repl. Vol.),² as assimilated by

¹ Horn was given the opportunity to take a Breathalyzer test but refused, as he is entitled to do under Maryland law. Md. Code Ann., Cts & Jud. Proc. § 10-309 (1998 Repl. Vol. & 2001 Supp.).

² At the time of Horn's arrest, Md. Code Ann., Transp. II § 21-902 stated in pertinent part:

18 U.S.C. §§ 7, 13, the Assimilative Crimes Act, a Class A misdemeanor.

Horn has filed a motion *in limine* to exclude the evidence of his performance on the field sobriety tests, asserting that it is inadmissible under newly revised Fed. R. Evid. 702 and the *Daubert/Kumho Tire* decisions.³ The Government has filed an opposition, and Horn has filed a reply. In addition, a two day evidentiary hearing was held, pursuant to Fed. R. Evid. 104(a), on November 19 and 20, 2001, and additional testimonial and

(a) *Driving while intoxicated or intoxicated per se.* -
(1) A person may not drive or attempt to drive any vehicle while intoxicated.

(2) A person may not drive or attempt to drive any vehicle while the person is intoxicated per se.

(b) *Driving while under the influence of alcohol.* - A person may not drive or attempt to drive any vehicle while under the influence of alcohol.

Effective September 30, 2001, § 21-902 was amended; a person is now charged with either (a) driving under the influence of alcohol or under the influence of alcohol per se or (b) driving while impaired by alcohol. Md. Code Ann., Transp. II § 21-902 (2001 Supp.). Subsection(a), driving under the influence, is now the most serious charge. The change in lexicon is a result partly because of the change in the level of proof, in the form of blood alcohol content results obtained from breathalyzer tests, needed to convict under each subsection. For purposes of this opinion, this Court will continue to employ the driving while intoxicated and driving while under the influence language prevalent in most state court opinions.

³ *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993); *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999).

documentary evidence was received, which is discussed in detail below. At the conclusion of this hearing, the following ruling was made from the bench, the Court also announcing its intention subsequently to issue a written opinion on this case of first impression:⁴

(1) The results of properly conducted SFSTs may be considered to determine whether probable cause exists to charge a driver with driving while intoxicated ("DWI") or under the influence of alcohol ("DUI");⁵

⁴ Research has not revealed any other federal case on this subject applying newly revised Rule 702 and the *Daubert/Kumho Tire* tests. There have been a few prior federal cases to consider the admissibility of horizontal gaze nystagmus evidence but never with the factual record of this case or a challenge to this evidence such as rendered here. See, e.g., *United States v. Daras*, 1998 WL 726748 (4th Cir. 1998)(unpublished opinion) (court discussed in passing the SFSTs but did not analyze their admissibility as scientific or technical evidence because the evidence exclusive of the tests was sufficient to establish the defendant's guilt); *United States v. Ross*, CR No. 97-972M (D. Md. February 9, 2000)(unpublished memorandum order, in which Judge Connelly of this Court commented with his characteristic thoroughness and thoughtfulness on the state court decisions and narrowly held that SFST evidence is sufficient to establish probable cause to administer a breathalyzer test); *United States v. Everett*, 972 F. Supp. 1313 (D. Nev. 1997) (holding that "drug recognition examiner" testimony was governed by Rule 702 but not by *Daubert* on the basis that the testimony was not scientific in nature but utilizing the *Daubert* factors in analyzing the evidence).

⁵ Horn did not contest the Government's entitlement to rely on the results of properly conducted SFSTs for probable cause determinations related to DWI/DUI charges. To establish

(2) The results of the SFSTs, either individually or collectively, are not admissible for the purpose of proving the specific blood alcohol content ("BAC") of a driver charged with DWI/DUI;⁶

(3) There is a well-recognized, but by no means exclusive, causal connection between the ingestion of alcohol and the detectable presence of exaggerated horizontal gaze nystagmus in a person's eyes,⁷ which may be judicially noticed by the Court pursuant to Fed. R. Evid. 201, proved by expert testimony or otherwise;

probable cause to arrest a suspect all that is required is reasonably trustworthy information that would support a reasonable belief that the suspect committed an offense. *Beck v. Ohio*, 379 U.S. 89, 91 (1964). Probable cause determinations turn on practical, nontechnical determinations. *Id.* Thus, regardless of whether SFSTs are admissible as evidence, they may establish probable cause to arrest a motorist for DWI/DUI.

⁶ The Government acknowledged during the Rule 104(a) hearing that it was not seeking to admit the results of the SFSTs to prove Horn's specific BAC. Nonetheless, this opinion must discuss the admissibility of the SFSTs for this purpose to fully explain the ruling made regarding their use as circumstantial evidence of intoxication or impairment.

⁷ As will be discussed below, nystagmus always is present in the human eye but certain conditions, including alcohol ingestion, can cause an exaggeration of the nystagmus such that it is more readily observable. In this opinion, use of the phrase "nystagmus" or "horizontal gaze nystagmus" being "caused" by alcohol refers to the exaggeration of this natural condition and does not suggest, absent any alcohol, there would not be any nystagmus at all.

(4) A police officer trained and qualified to perform SFSTs may testify with respect to his or her observations of a subject's performance of these tests, if properly administered, to include the observation of nystagmus, and these observations are admissible as circumstantial evidence that the defendant was driving while intoxicated or under the influence. In so doing, however, the officer may not use value-added descriptive language to characterize the subject's performance of the SFSTs, such as saying that the subject "failed the test" or "exhibited" a certain number of "standardized clues" during the test;

(5) If the Government introduces evidence that a defendant exhibited nystagmus when the officer performed the horizontal gaze nystagmus test, the defendant may bring out either during cross examination of the prosecution witnesses or by asking the Court to take judicial notice of the fact that there are many causes of nystagmus other than alcohol ingestion; and

(6) If otherwise admissible under Fed. R. Evid. 701, a police officer may give lay opinion testimony that a defendant was driving while intoxicated or under the influence of alcohol. In doing so, however, the officer may not bolster the lay opinion testimony by reference to any scientific, technical or specialized information learned from law enforcement or traffic safety instruction, but must confine his or her testimony to

helpful firsthand observations of the defendant.

The issues addressed in this case likely will recur, given the large number of Class A and B misdemeanors prosecuted in this district under the Assimilative Crimes Act. Moreover, the admissibility of SFSTs implicates recent changes to the federal rules of evidence, as well as a large body of state cases on this topic, primarily decided under a different evidentiary standard than that governing the admissibility of the results of SFSTs in federal court.⁸ Accordingly, this opinion will discuss the basis for the above rulings in more detail below.

1. Applicable Rules of Evidence

Fed. R. of Evid. 104(a) requires the Court to make preliminary determinations regarding the admissibility of evidence, the qualifications of witnesses and the existence of privileges, and Rule 104(a) now permits the Court to make definitive pretrial evidentiary rulings *in limine*. During Rule 104(a) hearings the rules of evidence, except those dealing with privileges, are inapplicable, permitting the Court greater

⁸ See, e.g., *Kay v. United States*, 255 F.2d 476 (4th Cir. 1958) (The Assimilative Crimes Act "does not generally adopt state procedures . . . and federal, rather than state, rules of evidence are applicable under the Act."); *U.S. v. Sauls*, 981 F. Supp. 909, 915 (D. Md. 1997).

latitude to consider affidavits such as those filed by Horn and the Government. Fed. Rules of Evid. 104(a), 1101(d)(1).

Whether the results of SFSTs are admissible depends first on the purpose for which they are offered. Fed. Rule of Evid. 105. Second, the SFSTs must be relevant and not excessively prejudicial for the purposes offered. Fed. Rules of Evid. 401, 403. Third, if the SFSTs are introduced by the testimony of a sponsoring witness who is testifying as to scientific, technical or specialized matters, the admissibility of the SFSTs is dependent on whether the witness's testimony meets the requirements of newly revised Fed. Rule of Evid. 702 and the *Daubert/Kumho Tire* standards. Finally, Fed. Rule of Evid. 102 emphasizes that interpretations of the rules of evidence should be made with an eye towards promptly, fairly, efficiently and inexpensively adjudicating cases.

In this case, the results of SFSTs potentially could be offered for the following purposes: (1) to establish probable cause to arrest and charge a defendant with DWI/DUI, (2) as direct evidence of the specific BAC of a defendant who performed the SFSTs or (3) as circumstantial proof that a defendant was driving while intoxicated or under the influence of alcohol. Horn has acknowledged that the tests may be used to determine

probable cause, as the overwhelming majority of cases have held,⁹ and the Government acknowledges that they are not admissible to prove the defendant's specific BAC, a conclusion almost universally reached by state courts, including Maryland.¹⁰ Accordingly, the task at hand is to determine to what extent the results of SFSTs are admissible as circumstantial proof that a driver has consumed alcohol and was driving while intoxicated or under its influence. Because the results of the SFSTs invariably are introduced by the testimony of an arresting police officer, and, as will be seen, may involve application of scientific, technical or other specialized information, the requirements of Rule 702, as recently revised, are of paramount importance.

Rule 702 permits testimony in the form of an opinion or otherwise regarding scientific, technical or specialized matters from a qualified expert, provided the testimony is based on (a) sufficient facts or data, (b) is the result of methods or principles that are reliable and (c) is the result of reliable

⁹ See, e.g., *Ballard v. State*, 955 P.2d 931 (Alaska Ct. App. 1998); *State v. Superior Court*, 718 P.2d 171, 176-78 (Ariz. 1986); *State v. Ito*, 978 P.2d 191 (Haw. Ct. App. 1999); *State v. Baue*, 607 N.W.2d 191, 197 (Neb. 2000) and Appendix.

¹⁰ See cases cited *infra* at p. 44 and Appendix.

application of the methods or principles to the facts of the particular case. These three requirements, added in December 2000, are complimentary to, but not identical with, the four non-exclusive evaluative factors identified by the Supreme Court in the *Daubert/Kumho Tire* cases: (a) whether the opinions offered are testable; (b) whether the methods or principles used to reach the opinions have been subject to peer review evaluation; (c) whether a known error rate can be identified with respect to the methods or principles underlying the opinion, and, finally, (d) whether the opinion rests on methodology that is generally accepted within the relevant scientific or technical community.¹¹

As further will be seen, almost the entire universe of published case law regarding the admissibility of SFST evidence comes from the state courts, as would be expected, given the fact that there is no uniform federal traffic code, and DWI/DUI cases in federal court usually come about as a result of assimilating state drunk driving laws under 18 U.S.C. §§ 7 and 13. This is significant because the vast majority of the state cases that have analyzed this issue have done so under the *Frye*¹²

¹¹ *Daubert*, 509 U.S. at 593-94; *Kumho Tire*, 526 U.S. at 141.

¹² *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

standard for admitting scientific or technical evidence: whether the methods or principles have gained general acceptance within the relevant scientific or technical community.¹³ While this test has continued vitality as one of the four *Daubert/Kumho Tire* factors, a federal court must do more in determining the admissibility of scientific, technical or specialized evidence than focus on general acceptance.

The starting point for this analysis is the SFSTs themselves, followed by a discussion of the evidence produced by the parties in this case regarding their reliability and then a consideration of the state cases that have focused on this issue.

2. The SFSTs

The three SFSTs that are the subject of this case were developed on behalf of the National Highway Traffic Safety Administration ("NHTSA") beginning in the 1970's. They are discussed in detail by a series of NHTSA publications, including:

- * a student manual for DWI detection and standardized field sobriety testing;
- * a June 1977 final report prepared for NHTSA by Marcelline

¹³See state cases cited *infra* at pp. 44-45 and Appendix.

Burns, Ph.D.¹⁴ and Herbert Moskowitz, Ph.D. of the Southern California Research Institute ("SCRI") titled "Psychophysical Tests for DWI Arrests" (the "1977 Report");

* a March 1981 final report prepared for NHTSA by Dr. Burns and the SCRI titled "Development and Field Test of Psychophysical Tests for DWI Arrest" (the "1981 Final Report");

* a September 1983 NHTSA Technical Report, authored by Theodore E. Anderson, Robert M. Schweitz and Monroe B. Snyder, titled "Field Evaluation Of A Behavioral Test Battery For DWI" (the "1983 Field Evaluation");

* a November 1995 study of the SFSTs funded by NHTSA and conducted by Dr. Burns and the Pitkin County Sheriff's Office, Colorado, titled "A Colorado Validation Study of the Standardized Field Sobriety Test (SFST) Battery" (the "1995 Colorado Validation Study"); and

*an undated study, authored by Dr. Burns and a sergeant of the Pinellas County Sheriff's Office, Florida, titled "A Florida Validation Study of the Standardized Field Sobriety Test (S.F.S.T.) Battery (the "Florida Validation Study").

(Gov't. Opposition Memo. Exhs. 2-7).

¹⁴ Dr. Burns is perhaps the most ardent advocate of the SFSTs at issue in this case, having participated in the original NHTSA studies that developed them, and thereafter as an ubiquitous--and peripatetic--prosecution expert witness testifying in favor of their accuracy and reliability in a host of state cases, over a course of many years. See cases cited *infra* at pp. 46-47. Despite her enthusiasm for the tests that she helped to develop, few, if any, courts have agreed with her that the SFSTs, taken alone or collectively, are sufficiently reliable to be used as direct evidence of specific BAC, as a review of the state cases listed in the Appendix to this opinion readily demonstrates. Dr. Burns has achieved, however, nearly universal success in persuading state courts that the SFSTs developed by SCRI, if properly administered, are admissible as circumstantial evidence of alcohol ingestion.

These studies are very significant, as they have been cited repeatedly by the state courts in their opinions regarding the admissibility of SFSTs in connection with assessment of the reliability of the SFSTs and their general acceptance within the law enforcement and traffic safety communities. They also are important in this case because they have been the subject of critical analysis by Horn's experts, who provided detailed testimony regarding the limitations of these studies and the extent to which the SFSTs are reliable and valid tests for driver intoxication or alcohol impairment.¹⁵

¹⁵ This underscores an important point. When analyzing the many state decisions regarding the admissibility of SFST evidence, care must be taken to focus on the factual basis supporting the rulings made. In many instances, the primary evidence that the court had before it regarding the reliability of SFSTs was Dr. Burns' testimony and the above described NHTSA, Colorado and Florida studies, as well as testimony from law enforcement officers with a vested interest in the use of the SFSTs. In most, but not all, instances, the defendant in the state cases simply did not mount a challenge to the "science" underlying the SFSTs. This is not the case here, where Horn has provided a spirited and detailed attack on the tests' reliability. This highlights an inherent limitation in the process of judicial evaluation of the reliability and validity of any scientific or technical evidence: the court must, under Rule 104(a), act as the "gatekeeper" to decide whether the evidence is reliable and admissible. The court, however, is limited in its ability to do so by the quantitative and qualitative nature of the evidence produced by the parties, whatever research the court itself may do, and any help it may derive from courts that have addressed the issue before it. This process unavoidably takes place on a continuum, and a court faced with the present task of deciding the admissibility of scientific evidence must exercise care to consider whether new developments or evidence

The three SFSTs developed by the research sponsored by NHTSA are summarized in the NHTSA student manual. (Gov't. Opposition Memo., Ex.2). The manual describes the tests and evaluations conducted to develop the SFSTs, then provides detailed instruction on how to administer and score each of the three tests.

The most "scientific" or "technical" of the three is the Horizontal Gaze Nystagmus Test ("HGN Test"). Nystagmus is "the involuntary jerking of the eyes, occurring as the eyes gaze toward the side. Also, nystagmus is a natural, normal phenomenon. Alcohol and certain other drugs do not cause this phenomenon, they merely exaggerate it or magnify it." *Id.* at VIII-12. Horizontal gaze nystagmus "occurs as the eyes move to the side." *Id.* at VIII-13. The HGN SFST requires the investigating officer to look for three "clues": (1) the inability of the suspect to follow a slowly moving stimulus smoothly with his or her eyes, (2) the presence of "distinct" nystagmus when the suspect has moved his or her eyes as far to the left or right as possible (referred to as holding the eyes at "maximum deviation") and held them in this position for

require a reevaluation of the conclusions previously reached by courts that did not have the benefit of the more recent information. In short, neither science and technology may rest on past accomplishments--nor may the courts.

approximately four seconds and (3) the presence of nystagmus before the eyes have moved 45 degrees to the left or right (which, the manual states, usually means that the subject has a BAC above 0.10). *Id.* at VIII-14-15. The officer is trained to look for each of the above three "clues" for each of the suspect's eyes, meaning there are six possible "clues." If the officer observes four or more clues the manual asserts that "it is likely that the suspect's BAC is above 0.10 [and] [u]sing this criterion [one] will be able to classify correctly about 77% of [one's] suspects with respect to whether they are above 0.10." *Id.* at VIII-17. If the results of the HGN test are offered to establish that the suspect's BAC is above 0.10,¹⁶ it is readily apparent that much depends on the investigating officer properly performing the HGN test procedures and on his or her subjective evaluation of the presence of the "standardized clues." Indeed, the manual itself cautions with

¹⁶ At the time of Horn's arrest, Maryland law stated that, "if at the time of [taking the breathalyzer test], a person has an alcohol concentration of at least .07 but less than .10" such results would be "prima facie evidence that the defendant was driving with alcohol in the defendant's blood." Md. Code Ann., Cts. & Jud. Proc. § 10-307 (1998 Repl. Vol.). Effective September 30, 2001, a blood alcohol concentration between 0.07 and 0.08 will be prima facie evidence that the person was driving while impaired by alcohol. If the person's BAC is .08 or higher, the defendant shall be considered under the influence of alcohol per se. Md. Code Ann., Cts. & Jud. Proc. § 10-307 (d), (g) (2001 Supp.).

respect to each of the SFSTs:

[the tests are valid] only when . . . administered in the prescribed, standardized manner; and only when the standardized clues are used to assess the suspect's performance; and, only when the standardized criteria are employed to interpret that performance. If any one of the standardized field sobriety test elements is changed, the validity is compromised.

Id. at VIII-12 (emphasis in original).

The Walk and Turn ("WAT") test requires the suspect to place his feet in the heel-to-toe stance on a straight line. The subject then is instructed to place his right foot on the line ahead of the left foot, with the heel of the right foot against the toe of the left. The suspect also is told to keep his arms down at his side and to maintain this position until the officer instructs him to begin the test. *Id.* at VIII-18. Once told to start, the suspect is to take nine heel-to-toe steps down the line, then to turn around in a prescribed manner, and take nine heel-to-toe steps back up the line. *Id.* While walking, the suspect is to keep his hands at his side, watch his feet, and count his steps out loud. *Id.* at VIII-19. Also, the suspect is told not to stop the test until completed, once told to start. *Id.*

As with the HGN test, the Manual asserts that there are

standardized clues, eight in all,¹⁷ that “[r]esearch . . . has demonstrated . . . are the most likely to be observed in someone with a BAC above 0.10.” *Id.* at VIII-19. Further, it states “[i]f the suspect exhibits two or more distinct clues on this test or fails to complete it, classify the suspect’s BAC as above 0.10. Using this criterion, you will be able to correctly classify about 68% of your suspects.” *Id.* at VIII-21. Once again, it is the officer’s subjective evaluation of the suspect that results in the determination of whether a “clue” is present or not, and, if only two of the eight “standardized clues” are detected, NHTSA asserts that the suspect’s BAC is 0.10 or more.

The third SFST is the One Leg Stand (“OLS”) test. In this test the suspect is told to stand with her feet together, arms at her sides. She then is told not to start the test until told to do so. To perform the OLS test, the suspect must raise whichever leg she chooses, approximately six inches from the ground, toes pointed out. *Id.* at VIII-23. While holding this position, the suspect then must count out loud for thirty seconds, by saying “one-one thousand, two-one thousand,” etc.

¹⁷ The eight clues are the inability to keep balance while listening to instructions, starting the test before the instructions are finished, stopping to steady one’s self, failure to touch heel-to-toe, stepping off the line, using arms for balance, improper turning, and taking an incorrect number of steps. *Id.* at VIII-20.

Id. The NHTSA manual identifies four "standardized clues" for the OLS test¹⁸ and instructs law enforcement officers that "[i]f an individual shows two or more clues or fails to complete the [test] . . . there is a good chance the BAC is above 0.10. Using that criterion, [one] will correctly classify about 65% of the people [one] test[s] as to whether their BACs are above or below 0.10." *Id.* at VIII-24.

The NHTSA Manual advises that when the WAT and HGN tests are combined, using a decision matrix developed for NHTSA, an officer can "achieve 80% accuracy" in differentiating suspects with BACs in excess of 0.10. *Id.* at VIII-5. These conclusions are supported, it is claimed, by the results of research and testing done by Dr. Burns and her company that was reported in the 1981 Final Report, the 1983 Field Evaluation, the 1995 Colorado Validation Study and the Florida Validation Study.¹⁹ *Id.* at Exs. 4-8.

As next will be seen, Horn's experts have challenged the

¹⁸ The four clues are swaying while balancing, using arms for balance, hopping, and putting a foot down. *Id.* at VIII-24.

¹⁹ The Florida Validation Study is undated. During the Rule 104(a) hearing, there was testimony from Spurgeon Cole, Ph.D., one of Horn's witnesses, that a third validation test had been done in San Diego, but it was not offered as an exhibit. Dr. Cole did testify, however, as to its conclusions and the defects in its design.

reliability, validity and relevance of the SFSTs to prove driver intoxication and are sharply critical of the claims of accuracy advanced in the NHTSA publications and the so-called validation studies. They have framed these objections in terms of the factors discussed in the *Daubert/Kumho Tire* decisions, as amplified by this Court in *Samuel v. Ford Motor Co.*, 96 F. Supp. 2d 491 (D. Md. 2000).

3. Horn's Challenges to the Reliability/Validity of SFST Evidence

Rule 702 prohibits expert testimony if it is not the product of *reliable* methods or principles that *reliably* have been applied to the facts of the particular case. In the context of scientific or technical testing, such as may be the case with SFSTs, reliability means the ability of a test to be duplicated, producing the same or substantially same results when successively performed under the same conditions. *Daubert*, 509 U.S. at 595; *Samuel*, 96 F. Supp. 2d at 494. Thus, for the SFSTs, if reliable, it would be expected that different officers, viewing the same suspect performing the SFSTs, would reach the same conclusion regarding the level of the suspect's impairment or intoxication. Alternatively, the same officer re-testing the same suspect with the same BAC as when first tested would reach the same conclusion.

A related, though distinct concept, deals with the *validity* of a test. A test is valid if it has a logical nexus with the issue to be determined in a case. *Daubert*, 509 U.S. at 591; *Samuel*, 96 F. Supp. 2d at 494. In the context of SFSTs, they are valid if there is a logical nexus between what the tests measure and the true ability of a driver safely to operate a motor vehicle. Thus, for example, does the fact that a suspect missed two "cues" in the WAT test mean that the driver cannot safely drive a car, or does it simply mean that the driver has some inability to perform the test that is unrelated to his or her ability to drive? Horn has challenged both the reliability and validity of the SFSTs.

During the Rule 104(a) proceedings, Horn produced four experts, three of whom submitted affidavits, and two of whom also testified: Yale Caplan, Ph.D. (former chief toxicologist for the State of Maryland and former scientific director of the Maryland Alcohol Testing Program); Spurgeon Cole, Ph.D. (Professor of Psychology, Clemson University and author of a series of articles critical of the SFSTs); Harold P. Brull (a licensed psychologist and consultant specializing in industrial/organizational psychology, particularly the definition and measurement of human attributes in employment and related settings); and Joel Wiesen, Ph.D. (an industrial

psychologist with special expertise in experimental psychology, psychometrics and statistics. Dr. Wiesen worked for more than ten years for the Massachusetts Division of Personnel Administration, developing and validating civil service examinations and is an independent consultant in the field of development and validation of human performance tests).

In his testimony and published writings, Dr. Cole was highly critical of the reliability of the SFSTs if used to prove the precise level of a suspect's alcohol intoxication or impairment. His 1994 article "Field Sobriety Tests: Are They Designed for Failure?," published in the journal *Perceptual and Motor Skills*, analyzed the 1977 Report, the 1981 Final Report, and the 1983 Field Evaluation report published by NHTSA regarding the SFSTs. (Def's. Memo, Ex. C.).

Dr. Cole observed the following:

(1) 47% of the subjects tested in the 1977 NHTSA laboratory study who would have been arrested by the testing officers for driving while intoxicated (BAC of 0.10 or greater) actually had BACs below 0.10;

(2) in the 1981 Final Report, 32% of the participants in the lab study were incorrectly judged by the testing officers as having BACs of 0.10 or greater; and

(3) the accepted reliability coefficient for standardized clinical tests is .85 or higher, yet the reliability coefficients for the SFSTs, as reported in the NHTSA studies, ranged from .61 to .72 for the individual tests and .77 for individuals that were tested on two different occasions while dosed to the exact same BAC. More

alarminglly, inter-rater reliability rates (where different officers score each subject) ranged from .34 to .60, with an over-all rate of .57.

Id. at 100.

Dr. Cole theorized that the SFSTs, particularly the WAT and OLS tests, required subjects to perform unfamiliar, unpracticed motions and noted that a very few miscues result in a conclusion that the subject failed and had a BAC in excess of 0.10. *Id.* His hypothesis was that individuals could be classified as intoxicated/impaired as a result of unfamiliarity with the test, rather than actual BAC. *Id.* He tested this hypothesis by videotaping twenty-one completely sober individuals performing either "normal-abilities tests" (such as reciting their addresses or phone numbers or walking in a normal manner) or the WAT and OLS tests. *Id.* at 99-102. The results of the study were that 46% of the officers that viewed the videotape of the sober individuals performing the SFSTs rated the subjects as having had too much to drink, as compared to only 15% reaching this decision after seeing the videotape of the subjects performing the normal-abilities tests. *Id.* at 102. Dr. Cole concluded:

[The SFSTs] must be held to the same standards the scientific community would expect of any reliable and valid test of behavior. This study brings the validity of field sobriety tests into question. If law enforcement officials and the courts wish to continue to use field sobriety tests as evidence of driving impairment, then further study needs to be conducted addressing the direct relationship of

performance on these and other tests with driving. To date, research has concentrated on the relationship between test performance and BAC and officers' perception of impairment. This study indicates that these perceptions may be faulty.

Id. at 103.

During his testimony at the Rule 104(a) hearing, Dr. Cole repeated his criticism of the reliability of the 1977, 1981 and 1983 studies but also testified about the Colorado, Florida and San Diego studies performed by Dr. Burns, styled as "field validation studies." This testimony echoed Dr. Cole's written criticisms about the SFSTs' reliability as precise predictors of the level of alcohol intoxication and the SFST's validity as a measure of driver impairment in his 1994 article, co-authored with Ronald H. Nowaczyk, titled "Separating Myth from Fact: A Review of Research on the Field Sobriety Tests" and published in the *Champion* journal of the South Carolina Bar Association. Def's. Reply Memo, Exh. 1.

Dr. Cole's primary criticisms, as discussed in his 1994 article, include, first, that the 1981 Final Report published by NHTSA claims an 80% accuracy rate for users of the SFSTs. This is misleading because when the actual data is examined with respect to the success rate of using the SFSTs to differentiate between drivers with BACs above 0.10 and those without, the critical population, the officers had "a 50/50 chance of being

correct just on the basis of guessing." *Id.* at 539.

Second, the SFSTs have a combined test-retest reliability rates of .77, while the scientific community "expects reliability coefficients to be in the upper .80s or .90 for a test to be scientifically reliable." *Id.* at 540. When different officers tested the same subjects at the same BAC dose level on different days the reliability was only .59--a 41% error rate. Dr. Cole contrasted these substandard reliability coefficients with that of the BAC machine, which is .96 or 96% reliable. *Id.* at 540-41.

Third, Dr. Cole argued that in order for the SFSTs to be valid predictors of BAC they must "not only identify individuals above a BAC level of 0.10 as 'failing', but also identify individuals below .10 as 'passing'." *Id.* at 541. The data from the NHTSA 1977 Report, however, shows that the validity of the HGN, OLS and WAT SFSTs was ".67, .48, and .55, respectively, with a combined validity coefficient of .67." *Id.* This means that use of the SFSTs results in an unacceptably high erroneous arrest rate, if the tests are used by the officer to make arrest decisions based on BAC levels being in excess of .10.

Fourth, Dr. Cole was particularly critical of claims that the NHTSA SFSTs have been "validated" in a "field setting." In this regard, he stated that the 1977 and 1981 NHTSA studies were

done in a laboratory setting, and the difference in conditions in a controlled lab are dramatically dissimilar from field conditions that can be expected when officers employ SFSTs at all times of day and night in widely disparate weather and traffic conditions and where issues of officer safety may influence how the test is performed.²⁰ *Id.* at 542. Dr. Cole stated that the NHTSA 1983 Field Evaluation purported to be a field validation study, but it failed to meet the recommendations of the authors of the NHTSA 1981 Final Report that the SFSTs be validated in the field for eighteen months in locations across the country. *Id.* Dr. Cole also stated that Dr. Burns herself has testified that the SFSTs adequately have not been field tested.²¹ *Id.*

²⁰ This criticism is especially significant in light of the third evaluative factor in Rule 702. This factor requires that the expert's opinion testimony be based on the use of principles/methods themselves reliable but that also reliably have been applied to the facts of the particular case. Thus, even if the SFSTs are determined to be reliable measures of driver intoxication, an officer's testimony about their use in a particular case could not be allowed absent a showing that the officer properly had administered the tests.

²¹ During his testimony, Dr. Cole stated that the Colorado, Florida and San Diego "validation" studies performed by Dr. Burns with various sheriff's departments do not cure the defects contained in the original reports. The three studies involved officers that made stops of drivers that were driving unsafely, and the officers evaluated them using the SFSTs, but also had the benefit of preliminary breath analysis tests, in many instances, and the studies do not permit a

Finally, Dr. Cole disputed the claims of proponents of the SFSTs that the studies regarding them have been published in peer review journals. The 1977 and 1981 field studies were published in technical reports by NHTSA, but those reports excluded the "methods and results" sections because they were thought to be too lengthy. *Id.* at 543. Cole concluded "[i]t is difficult to see how the NHTSA could claim that the FST is accepted in the scientific community, when results of studies on the validation of the FST have never appeared in a scientific peer reviewed journal, which is a basic requirement for acceptance by the scientific community." *Id.* Cole concluded:

Because of its widespread use, the FST battery has been **assumed** to be a reliable and valid predictor of driving impairment. NHTSA has done little to dispel that assumption. Law enforcement cannot be blamed for its use of the FST battery. Training documents refer to NHTSA reports and provide what appears to be supporting evidence for the validity of the FST battery. In addition, there is little doubt that individuals who have high BAC levels will have difficulty in performing the FST battery. However, what the law enforcement community and the courts fail to realize is that the FST battery may mislead the officer on the road to incorrectly judge individuals who are not impaired. The FST battery to be valid must discriminate accurately between the impaired and non-impaired driver.

critical reviewer to determine whether the officer's arrest decision was based on the SFSTs alone, or on the totality of the information available to the officer, including the results of the breath test. Thus, the studies were not controlled, and there were multiple variables that affected the ultimate decision. He concluded, therefore, that these "validation" studies were scientifically unacceptable.

NHTSA's own research on that issue . . . has not been subjected to peer review by the scientific community. In addition, a careful reading of the reports themselves provides support for the inadequacy of the FST battery. The reports include low reliability estimates for the tests, false arrest rates between 32 and 46.5 percent, and a field test of the FST that was flawed because the officers in many cases had breathalyzer results at the time of the arrest. NHTSA clearly ignored the printed recommendations of its own researchers in conducting that field study.

Id. at 546. (Emphasis in original).

Horn also introduced the affidavit of Joel P. Wiesen, Ph.D. Dr. Wiesen is an industrial psychologist with special expertise in experimental psychology, psychometrics and statistics. His experience includes more than ten years working with the Commonwealth of Massachusetts developing civil service examinations and an equal number of years as an independent consultant in the area of test development and validation. In addition, he is a published author of a mechanical aptitude test used nationwide. Although he is most familiar with written tests, he does have experience in the development of human performance tests. Def's. Reply Memo, Exh.6 at 1.

Dr. Wiesen reviewed the NHTSA 1977 Report, the 1981 Final Report, the 1983 Field Evaluation, the 1995 Colorado Validation Study, the undated Florida Validation Study, and the NHTSA student manual for the SFSTs. He was highly critical of these

studies, as the following summary illustrates:²²

1977 Report	1981 Report (Lab & Field Phases)	1983 Report	1995 Colo. Study	Fla. Study
1. In the lab the HGN test was administered using a chin rest which facilitated making HGN observations. This was not done in the field.	1. Serious flaws include 20% false positive evaluations of intoxic.; very high error rates in reliability if using SFSTs to predict BAC.	1. Report seriously flawed, does not meet professional standards of testing community.	1. Report describes results of impaired driving arrests from seven Colorado law enforcement organizations. Report too incomplete to draw any conclusions about the validity of the test.	1. Report too incomplete to permit meaningful evaluation.
2. A single set of data was used to determine criterion score and to evaluate accuracy of test, which artificially inflates estimate of accuracy.	2. HGN test affected by time of day, no adjustment in scoring.	2. Failure to monitor data collection by officers. Cannot tell if decisions based on SFSTs or prelim. breath test (PBT).	2. Methodology results and data sections of report are missing.	2. Methodology not described, and data regarding methodology not provided in report.
3. Tests are not age & gender neutral, and age/gender differences can affect ability to perform SFSTs.	3. Test/retest reliability rates very low.	3. Arrest decisions made on PBT results as well as SFSTs. Not possible to tell reliability of SFSTs.	3. Data generated by "volunteer" officers-- suggesting possible bias.	3. Data incompletely described.

²² The information reported in the chart is found in Def's. Reply Memo, Ex.6 at 1-13.

<p>4. In lab tests officers were monitored to insure correct performance of tests, not done in field.</p>	<p>4. Report states testing officers did not necessarily base decisions on results of SFSTs, making validity suspect.</p>	<p>4. Authors fail to report the data from N.C. Test site-over 25% of data for whole test.</p>	<p>4. No monitoring of data collection to verify reporting methodology. Officers merely reported results.</p>	
<p>5. Test results differ in statistically significant respects depending on time of day that HGN test was performed, yet test scoring did not account for difference in time of day test was administered.</p>	<p>5. Authors admit field test data not appropriate for statistical significance testing, and could be biased.</p>	<p>5. No statistical tests conducted on data.</p>	<p>5. Results unclear, particularly because two different arrest standards used (one for intoxication, another for impaired)</p>	
<p>6. The study was not peer reviewed, and would not have been accepted if offered.</p>	<p>6. High error rates. 28.6% of subjects with "legal" BAC arrested, and 50% of subjects w/ BAC > 0.10 not arrested.</p>	<p>6. SFSTs not administered in standard fashion.</p>		
	<p>7. Officers selected for study not representative of police officers across the board.</p>	<p>7. Authors acknowledge "extreme caution" needed in analyzing data collected in study. Accuracy of data suspect.</p>		
	<p>8. Authors reported that in field some officers forgot or ignored standardized procedure to administer SFSTs.</p>			

Dr. Wiesen concluded his evaluation of the SFST reports with the following observation:

the studies give only a general indication of the level of potential validity of the tests as described in the NHTSA manual Rather than the five studies supporting each other, they evaluate somewhat different combinations of test content and test scoring. The differences are large enough to change the validity and accuracy of the tests. The older studies are probably less germane, due to the changes in test content and scoring over time. The reports for the newer studies are grossly inadequate. Given this, and in light of the specific critiques above (which are not exhaustive), I can only conclude that the field sobriety tests do not meet reasonable professional and scientific standards.

Id. at 12-13.

Harold P. Brull testified on behalf of Horn and supplied an affidavit as well. Mr. Brull is a licensed psychologist with many years experience consulting in connection with the design and implementation of procedures to measure human attributes, especially in employment settings. He has designed and evaluated tests and procedures measuring human characteristics for over twenty years. Def's. Reply Memo, Exh. 5 at 2.

Mr. Brull reviewed the NHTSA 1977 Report, the 1981 Final Report, the 1983 Field Evaluation, the 1995 Colorado Validation Study, the Florida Validation Study, and the NHTSA officer training manual. Among his general observations of these materials was the opinion that there was a complete absence of

evidence "which would allow one to predict a known error rate in the field," where there is no ability to control the performance of the SFSTs like there is in a laboratory setting. Def's. Reply Memo, Exh. 4 at 6. He was especially critical of the assertions in the Florida and Colorado studies regarding the reliability of the SFSTs, primarily because of their use of lower BAC thresholds (0.05 and above instead of 0.10), the fact that the population of drivers evaluated were those stopped because of unsafe driving and the complete absence of any data in the reports to enable meaningful evaluation. *Id.* at 6-7. He further expressed the opinion that none of the reports was published in peer review literature. While Brull was not critical of the methodology used in the 1977 and 1981 laboratory studies, he stated that the results from these studies were inconclusive, and the subsequent field tests "simply do not contain sufficient detail or rigor to support any hypothesis that field sobriety studies, as conducted by police officers in the field, are valid and reliable." *Id.* at 7.

Brull's evaluation of the data contained in the 1977 and 1981 reports was consistent with that of Dr. Cole and Dr. Wiesen. Regarding the 1981 Final Report, he observed that "the degree of predictive error in the field appeared to be substantially larger than in the laboratory," and that "[w]hile training clearly

brought about improvement, it does not compare favorably to the laboratory condition and is [sic] a margin of error substantially higher than one would find acceptable for predicting with any degree of certainty." *Id.* at 11.

Brull was most critical of the Colorado and Florida "validation" studies. He noted that they "are merely summary reports, without foundation, of findings," and suffered from a "serious methodological flaw," in that the tests were done on actual motorists stopped by officers because their driving was unsafe, leading the officers automatically to suspect that they were intoxicated. *Id.* Use of this population likely will produce results that Brull characterized as "highly inflated." *Id.* He further noted that these field studies predicted 90% accuracy in identifying drivers with BAC's above 0.05, a level only one half that used in the earlier tests and below the level of legal intoxication. While the validation studies provided no data to assess the accuracy of the SFSTs in identifying drivers with BACs of 0.10 or higher, Brull suspected that the accuracy rate would be far lower than 90%. *Id.* at 12.

Brull's final conclusions were summarized as follows:

(1) the laboratory studies that form the foundation of the SFSTs (the 1977 and 1981 studies) were well designed;

(2) the accuracy of the SFSTs, even under laboratory conditions, is less than desired and below the level

expected for tests of human performance;

(3) the field studies were not well documented, produced unknown error rates, but which, if known, likely would have been unacceptable in real world situations;²³

(4) the error rate of SFSTs as actually performed by officers in the field is unknown;

(5) the only peer review article analyzing the SFST's was written by Dr. Cole and is highly critical of the accuracy of the SFSTs.

Id. at 14.

Finally, Horn offered the affidavit of Yale H. Caplan, Ph.D., Defs.' Motion, Ex. E. Dr. Caplan has more than thirty

²³ The concern about the reliability of SFSTs performed by officers in the field under actual stop and detain conditions is not fanciful, given the fact that the NHTSA officer training manual itself cautions that the reliability of the SFSTs depends on strict compliance with the standardized procedures. Gov't. Opposition Memo, Exh. 2 at VIII-12. Further, there is clear evidence that given the conditions under which SFSTs actually are performed in real life situations, officers often do not follow the prescribed methodology. See Def's. Reply Memo, Exh.8 at 116 ("End-position nystagmus as an indicator of ethanol intoxication," *Science and Justice Journal* 2001)(author studied videotapes of actual traffic stops where HGN test was administered. Over 98% of the roadside HGN tests were improperly conducted); 1981 Final Report at 18-19 (stating that officers did not necessarily follow the standardized decision criteria used with the SFSTs). The fact that officers may not perform the SFSTs properly in the field has special significance when evaluated under Rule 702, as the third factor in that rule requires the court to find that the opinion testimony is based on reliable methods or principles that reliably were applied to the facts of the particular case. Thus, if reliable methods exist, but are not used in a particular instance, the results of the misapplication of the methodology are not admissible.

years experience in the field of forensic toxicology and alcohol and drug testing. He served for many years as the chief toxicologist for the Maryland Medical Examiner's office and now is a consultant in the field of toxicology. *Id.* Dr. Caplan stated that a determination that a person is impaired by alcohol consumption may be made in one of two fashions: by direct evidence of impairment derived from the chemical analysis of a breath or blood specimen; or indirectly by assessing performance indicators of the subject through field sobriety tests. *Id.* With respect to the latter, Dr. Caplan stated:

Although physiological assessments (e.g. standardized field sobriety tests) when coupled with the odor of alcohol on breath and alcohol's relatively high epidemiological prevalence in drivers may suggest alcohol as the causative agent, the use of drugs or the concomitant use of alcohol and drugs or other medical conditions must be considered as causes for the impairment. In fact, field sobriety tests alone were never designed for or demonstrated to be unequivocally capable of indicating alcohol impairment.

Id. He expressed the following opinions: (1) that field sobriety tests can be used to define impairment but that a specific blood/breath alcohol test is needed to confirm that the cause of the impairment is alcohol ingestion; (2) that an alcohol test of a suspect's breath or blood can alone be used to establish impairment, but field sobriety tests alone cannot establish alcohol impairment "with absolute certainty." *Id.*

4. The Government's Evidence

In response to the evidence submitted by Horn, the Government introduced the affidavit of Officer Jarrell, the arresting officer, describing the stop, detention and arrest of Horn and the SFSTs administered to him. The Government also introduced the 1977, 1981, and 1983 NHTSA reports, the California and Florida "validation studies," the NHTSA student manual regarding the SFSTs, and an article titled "Horizontal Gaze Nystagmus: The Science & the Law," published by the American Prosecutors Research Institute's National Traffic Law Center ("NTLC").²⁴ Gov't's. Opposition Memo, Exhs. 1-7.

Additionally, the Government introduced the affidavit of Lieutenant Colonel Jeff C. Rabin, O.D., Ph.D., a licensed optometrist on active duty in the Army, assigned as the Director of Refractive Research at the Walter Reed Army Institute for Research, Walter Reed Army Medical Center.²⁵ *Id.* Exh. 8. Colonel

²⁴ The NTLC was "created in cooperation with . . . (NHTSA) and works closely with NHTSA and the National Association of Prosecutor Coordinators to develop training programs." The NTLC is a program of the American Prosecutors Research Institute, the principal function of which "is to enhance prosecution in America." Gov't. Opposition Memo, Exh. 1 at 2. The foreword to this publication was written by Dr. Marcelline Burns.

²⁵ The Government also had intended to introduce the affidavit of Sergeant Thomas Woodward of the Maryland State Police but ultimately was unable to do so.

Rabin, who also testified at the Rule 104(a) hearing, has testified as an expert witness on the effects of alcohol and drugs on eye movements, given presentations to Army doctors and optometrists on this subject and reviewed the NHTSA publications regarding the HGN and other SFSTs. *Id.* Exhs. 8, 9. His affidavit and trial testimony confirmed the fact that alcohol ingestion can enhance the presence of nystagmus in the human eye at BAC levels as low as .04. He expressed the opinion that "there is a very good correlation between the results of the . . . [HGN] test and breath analysis for intoxication." *Id.* He also stated that the three "clues" that officers are taught to look for in connection with the HGN SFST "are indicative of alcohol consumption with possible intoxication." *Id.* Colonel Rabin expressed his belief that police officers could be trained adequately to administer the HGN test and interpret its results.

Colonel Rabin's testimony was consistent with his affidavit. He did acknowledge, however, that he acquired his knowledge of, and formed his opinions about, the SFSTs in connection with performing duties as an expert witness for Army prosecutors in two courts martial, not as a result of any independent research that he had done as an optometrist. It further was acknowledged that Colonel Rubin was not asked to analyze in any detail the reliability and validity of the NHTSA SFST studies, and he had no

opinion on this subject. Further, the references to the HGN SFST that he read in peer review literature published by the American Journal of Optometry was based primarily on the NHTSA studies, rather than any independent research by that organization. He also acknowledged, in response to questions from the Court, that there are many causes of exaggerated nystagmus in the human eye that are unrelated to the ingestion of alcohol.

DISCUSSION

A. The State Case Law

State courts have wrestled with the admissibility of SFST results in drunk driving cases since 1986, when the Supreme Court of Arizona decided *State v. Superior Court*, 718 P.2d 171 (Ariz. 1986). In that decision, based on the testimony before the trial court by Dr. Burns and three police officers, and using the *Frye*²⁶ test, the court held that the results of a HGN test were sufficiently reliable to be used to establish probable cause to arrest a motorist for DWI/DUI, and that it had achieved general acceptance among behavioral psychologists, highway safety experts, neurologists and law enforcement personnel. *Id.* at 180. The court therefore held that HGN evidence was admissible

²⁶ *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

to prove driver intoxication/impairment.²⁷ *Id.* at 181.

Since the 1986 Arizona decision, a majority of the states have ruled on the admissibility of HGN and SFST evidence. A reading of these cases reveals that there are a core of decisions that have attempted to undertake a thorough review of the facts relating to admissibility of SFST evidence. Other state courts have relied more on the rulings of courts that previously had addressed the issue than on their own independent evaluation. It would unnecessarily lengthen this opinion to discuss all the state cases in detail. Thus, the Appendix attached to this opinion includes a chart that identifies the majority of state cases and briefly summarizes their holdings.²⁸ I will, however,

²⁷ The court cautioned that it was not ruling that HGN test results were admissible to prove that a driver had a BAC in excess of 0.10 "in the absence of a laboratory chemical analysis." *Id.* at 181. In *State v. City Court of the City of Mesa*, 799 P.2d 855 (Ariz. 1990), the Arizona Supreme Court clarified that in cases where no independently admissible chemical test of a driver's BAC had been performed, HGN evidence was admissible only as circumstantial evidence that the driver had consumed alcohol and not to prove a specific BAC. *Id.* at 860.

²⁸ The Appendix is intended to aid future courts called upon to research the issues presented in this case. The Court gratefully acknowledges the assistance of Ms. Jennifer Warfield, Mr. Kevin Cross, Ms. Jennifer Thomas, and Mr. Rodney Butler, interns who worked tirelessly on the Appendix. If the future of the legal profession may be predicted by these law students' work, it is a bright one. It also should be noted that, in addition to appointed counsel, Horn was also represented by Mr. Ryan Potter, a law student in the

discuss certain of the state cases in this opinion, as they are essential to understanding the rulings reached herein.

Maryland's appellate cases discussing the admissibility of HGN and other SFST evidence fall into the category of state court cases that have undertaken a comprehensive evaluation of the admissibility of this evidence. The principal case, *Schultz v. State*, 664 A.2d 60 (Md. App. 1995), has been cited repeatedly by other state courts in support of their own rulings on the admissibility of SFST evidence.

The defendant in *Schultz* was convicted of DUI. At the trial in the circuit court, the state's only evidence that the driver was driving under the influence of alcohol came from the arresting officer. Accordingly, the Court of Special Appeals was deprived of any evidence of record regarding the reliability of the HGN test. Its decision in *Schultz* was based on the court's own evaluation of other cases and the published literature regarding the HGN test from which the court took judicial notice of its reliability and general acceptance. *Id.* at 69-74. In

University of Maryland's much respected clinical law program. Admitted to practice under Local Rule 702, and under the skillful supervision of Professor Jerry Deise, these clinical law students offer significant assistance to their clients while concomitantly gaining invaluable trial experience. Ms. Claudia Diamond, my law clerk, also was instrumental in helping to revise and edit this opinion for which I am also very thankful.

doing so, the court observed that under Rule 5-702²⁹ of the Maryland Rules of Evidence, it was required to apply the *Frye* test, adopted in Maryland in *Reed v. State*, 391 A.2d 364 (Md. 1978).³⁰ In doing so, the court used a three prong test to determine whether HGN evidence satisfied the *Frye/Reed* test: (1) whether the scientific theory underlying the HGN test was *reliable*; (2) whether the *methods* used in connection with the HGN

²⁹ The Maryland rules of evidence were adopted in 1994 after the *Daubert* decision had been rendered by the United States Supreme Court. In the commentary to Rule 5-702, which is the state equivalent to Fed. R. Evid. 702, the drafters, however, noted that it was not their intent to adopt the *Daubert* test, then widely viewed as applicable only to issues regarding the admissibility of scientific evidence. Instead, the Maryland rule was intended to maintain the *Frye* test, which had been adopted by the state in the case of *Reed v. State*, 391 A.2d 364 (Md. 1978). To this day, Maryland has declined to adopt the *Daubert* test. *Burrall v. State*, 724 A.2d 65, 80 (Md. 1999) ("We have not abandoned *Frye* or *Reed*."); *Clark v. State*, 781 A.2d 913, 935 & n.13 (Md. Ct. Spec. App. 2001); *State v. Gross*, 760 A.2d 725, 757 (Md. App. 2000); *Schultz*, 664 A.2d at 64 n.3. Thus, in federal court, under the most recent version of Rule 702 and the *Daubert/Kumho Tire* decisions, the proponent of any expert testimony, whether scientific, technical or the product of some specialized knowledge, must undertake an analysis of reliability of the methods/principles underlying the opinion, as well as the reliability of the application of the methodology used by the expert to the particular facts of the case. Under Maryland evidence law, the *Frye/Reed* test applies only to introduction of scientific evidence, and Rule 5-702 alone covers all other types of expert opinion testimony.

³⁰ Maryland cases routinely refer to the *Frye* test as the "*Frye/Reed*" test. This opinion will as well.

test had been accepted by scientists familiar with the test and its use; and (3) whether the police officer in the case at bar properly had been trained to administer the test and administered it properly.³¹ *Id.* at 64. The *Schultz* court based its findings regarding the HGN test on the Arizona Court's decision in *State v. Superior Court*, the decisions of other state courts, as well as its reading of various studies and articles. *Id.* at 72-73. Its consideration regarding the reliability of the HGN test, however, is most significant with respect to the ruling made in this decision. Because it lacked the robust evidentiary record available to this court regarding the reliability of the HGN, OLS, WAT tests, the Court of Special Appeals was required to look at case law and published materials to determine whether the HGN test was reliable and generally accepted. The primary bases

³¹ As noted at pp. 7-8, in December 2000 the Federal Rules of Evidence were amended. Among the rules that were changed was Rule 702, the expert opinion rule. The amendment added three additional foundational requirements before expert testimony in any subject, whether scientific, technical or other specialized knowledge, is admissible: the opinion must be based on *sufficient facts or data*; it must be the product of *methods* and *principles* shown to be *reliable*, and the proponent must show that the methods/principles *reliably had been applied* to the facts of the case at hand. These factors are required by the rule itself and are independent from the factors identified by the Supreme Court in the *Daubert/Kumho Tire* decisions. The Maryland Rules of Evidence did not adopt the 2000 changes to the federal rules, and the Maryland expert opinion rule, Rule 5-702, does not contain the three additional foundational requirements as does Rule 702.

for its conclusion that it was, and that it therefore could take judicial notice of this fact, were a decision by the Texas Supreme Court in *Emerson v. State*, 880 S.W.2d 759 (Tex. Crim. App. 1994), a 1986 article authored by Edward B. Tenney and published in the New Hampshire Bar Journal,³² and the NHTSA 1983 Field Evaluation. *Id.* at 73 and n. 12.

In *Emerson*, the Texas court based its conclusions regarding the reliability of the HGN test on the NHTSA studies. *Emerson*, 880 S.W.2d at 766-67. The Tenney article cited only the NHTSA studies regarding the scientific basis for the HGN test and reached the conclusion that “[i]f the State of New Hampshire is still a true *Frye* jurisdiction, then the likelihood that results from horizontal gaze nystagmus testing will be admitted into evidence in this state is extremely thin,”³³ making it a questionable source to cite for the reliability of HGN testing. Finally, the conclusions of the NHTSA 1983 Field Evaluation have been aggressively challenged by Horn’s experts in this case. In short, the foundation of the Court of Special Appeals’ decision that the HGN test was sufficiently reliable and generally

³² Edward B. Tenney, *The Horizontal Gaze Nystagmus Test and the Admissibility of Scientific Evidence*, 27 New Hampshire Bar Journal 179 (1986) (hereinafter “Tenney article”).

³³ Tenney article at 187.

accepted rests on taking judicial notice of studies and articles that, at the time of their publication, had not been subject to the type of critical evaluation presented in this case.

The doctrine of judicial notice is predicated upon the assumption that the source materials from which the court takes judicial notice are reliable.³⁴ Where, as here, that reliability

³⁴ Indeed, in this regard, the Maryland and Federal Rules of Evidence are substantially identical. Rule 5-201 and Fed. R. Evid. 201 permit the taking of judicial notice of adjudicative facts if: (a) the facts are generally known within the territorial jurisdiction of the court or (b) capable of accurate and ready determination by resort to sources whose accuracy cannot reasonably be questioned. Obviously, the scientific basis underlying HGN tests is not a matter generally known within the state; so, if judicial notice is to be taken, it must be by reference to sources whose accuracy cannot reasonably be questioned. While the sources relied on in the *Schultz* case may not have been subject to reasonable question at the time that court considered them, given the lack of any evidentiary facts in the record regarding the reliability of the HGN test, and the fact that judicial notice was taken on appeal--not at the trial level where the parties might have had an opportunity to develop a factual basis to challenge the propriety of judicial notice-- the same cannot be said given the record in this case. Further, Rule 201(e) and 5-201(e) permit a party to be heard on the propriety of taking judicial notice, which did not occur in the *Schultz* case because judicial notice was taken on appeal. As one commentator has noted "where judicial notice of an adjudicative fact is taken by an appellate court on its own motion, an issue arises as to whether the provisions of Rule 201(e) concerning an opportunity to be heard are to be applied. At the moment, the question is unresolved." Graham, *Handbook of Federal Evidence* § 201.07 (5th ed. 2001). In any event, Rule 201(g) provides that in criminal cases, the court must instruct the jury that "it may, but is not required to, accept as conclusive any fact judicially noted." Implicitly, the rule would permit a

has been challenged, the court cannot disregard the challenge, simply because a legion of earlier court decisions reached conclusions based on reference to the same then-unchallenged authority. For the reasons that will be explained below, on the record before me, I cannot agree that the HGN, WAT and OLS tests, singly or in combination, have been shown to be as reliable as asserted by Dr. Burns, the NHTSA publications, and the publications of the communities of law enforcement officers and state prosecutors. While I ultimately agree, in large part, with the conclusions reached by the vast majority of state courts that the results of the HGN tests are admissible as circumstantial evidence of alcohol consumption, I must do so by recognizing their limited reliability and with substantial doubts about the degree of their general acceptance within an unbiased scientific or technical community.

This is not to say that I am critical of the decisions in *Schultz* or the other state courts. To the contrary, they are, for the most part, well-reasoned and written, based on the information then available to the deciding courts and the

defendant in a criminal case to offer evidence to rebut any adjudicative fact noticed by the Court. Thus, if a Court took judicial notice of the reliability and general acceptance of the HGN test, the defendant initially could object to it doing so under Rule 201(e). Then, if unsuccessful in preventing the court from taking judicial notice, the defendant could introduce evidence contesting the fact judicially noted.

inherent limitations of the process by which courts receive proof--either from evidence introduced by the parties themselves or by the taking of judicial notice from decisions of other courts or published materials. The Court of Special Appeals itself noted the danger inherent in such a process:

We note with some caution the dissent in *Emerson, supra*, which initially noted that, by taking judicial notice of the reliability of HGN testing and technique, the appellate court had relieved the State of its burden of establishing the reliability of the test at trial. We acknowledge that we, in taking judicial notice of the reliability of the test . . . are likewise relieving the State of that burden. We shall, nevertheless, take judicial notice that HGN testing, a scientific test, is sufficiently reliable and generally accepted in the relevant scientific community. . . . To do otherwise at this stage in the development of the science would leave to individual courts within the twenty-three jurisdictions of this State (and the various courts and judges within each jurisdiction) to determine, on a case-by-case basis, the scientific reliability of the test. In each of the various jurisdictions, the determination of the reliability and acceptability of such evidence would depend upon the competence, energy, and schedules (and even budgets) of the various prosecutors throughout the State in obtaining, and producing the attendance of experts at the thousands of trials involving alcohol related offenses in which HGN testing is sought to be admitted. Disparate results and decisions might result in many instances, not from the actual scientific reliability of the tests themselves, but from the differing abilities and resources of prosecutors and the availability of witnesses from the scientific community.

Schultz, 664 A.2d at 74.

The practical truth of the above reasoning cannot be denied. None today can doubt the serious public safety concerns related to driving by intoxicated or impaired motorists or the magnitude

of this problem.³⁵ Neither can it be disputed that, given the volume of DWI/DUI cases, the press of other criminal cases, and the limited resources and time of prosecutors to prepare them for trial, it is highly desirable to have available a simple, inexpensive, and reliable test that can be administered by police officers on the road, which would facilitate a prompt and inexpensive trial. Indeed, Rule 102 would militate in favor of interpreting the rules of evidence in such a fashion as to accomplish this end, if fairly possible. What cannot be lost in the process, however, is the requirement that the trial be a fair one and that the sum of the evidence introduced against the defendant must be sufficiently probative to prove guilt beyond a reasonable doubt.³⁶ Expedient as it may be for courts to take

³⁵ In FY 2000/2001, 35,962 DWI/DUI cases were filed in Maryland. *Administrative Office of the Maryland Courts Judicial Information System*, Maryland District Court Traffic System Citation Statistics, Report No. A70TM214, Run Date July 15, 2001.

³⁶ In addition, if local prosecutors may lack sufficient resources to prove the reliability and general acceptance of the SFSTs, which it is their burden to do in the first instance, it can be expected, *a fortiori*, that individual defendants charged with DWI and DUI will have even fewer resources to challenge the science and technology underlying these tests. If, once accepted by the application of the judicial notice rule, SFSTs are ever after immune from reconsideration, even in the face of new evidence challenging their reliability, then the burden will have been shifted from the state or government to establish the admissibility of the SFSTs to the defendant to disprove their admissibility. This

judicial notice of scientific or technical matters to resolve the crush of DWI/DUI cases, this cannot be done in the face of legitimate challenges to the reliability and accuracy of the tests sought to be judicially noticed. As will be seen, there is a place in the prosecutor's arsenal for SFST evidence, but it must not be cloaked in an aura of false reliability, lest the fact finder, like the protagonist in the Thomas Dolby song, be "blinded by science" or "hit by technology."³⁷

From a review of the state court decisions regarding the admissibility of HGN evidence in particular, and SFST evidence in general, a number of observations may be made. First, most of the states that have ruled that HGN evidence is admissible have not allowed it to be used to prove specific BAC but instead only as circumstantial proof of intoxication or impairment. See, e.g., *Ballard v. State*, 955 P.2d 931 (Alaska Ct. App. 1998); *State v. City Court of the City of Mesa*, 799 P.2 855 (Ariz. 1990); *State*

is a high price to pay in the interest of conserving limited prosecutorial resources.

³⁷ "She blinded me with science!
And hit me with technology."

Thomas Dolby, "She Blinded Me With Science," <http://www.prebble.com/sheblinded.htm>. See also *State v. Ferrer*, 23 P.3d 744, 765 n.6 (Haw. Ct. App. 2001)(quoting *State v. O'Key*, 899 P.2d 663, 672 n.6) (jurors may be "overly impressed with the aura of reliability surrounding scientific evidence").

v. Ruthardt, 680 A.2d 349 (Del. Super. Ct. 1996); *State v. Garrett*, 811 P.2d 488 (Idaho 1991); *State v. Buening*, 592 N.E.2d 1222 (Ill. App. Ct. 1992); *State v. Taylor*, 694 A.2d 907 (Md. 1997); *Wilson v. State*, 723 A.2d 494 (Md. App. 1999); *State v. Baue*, 607 N.W.2d 191 (Neb. 2000); *City of Fargo v. McLaughlin*, 512 N.W.2d 700 (N.D. 1994); *State v. Bresson*, 554 N.E.2d 1330 (Ohio 1990); *State v. O'Key*, 889 P.2d 663 (Or. 1995); *State v. Sullivan*, 426 S.E.2d 766 (S.C. 1993); *State v. Emerson*, 880 S.W.2d 759 (Tex. Crim. App. 1994).

Second, most of the states that have ruled that HGN evidence is admissible have employed the *Frye* standard requiring general acceptance of the test within the relevant scientific or technical community. *See, e.g., Malone v. City of Silverhill*, 575 So. 2d 101 (Ala. Crim. App. 1989); *State v. Superior Court*, 718 P.2d 171 (Ariz. 1986); *People v. Leahy*, 882 P.2d 321 (Cal. 1994); *Williams v. State*, 710 So.2d 24 (Fla. Dist. Ct. App. 1998); *Hawkins v. State*, 476 S.E.2d 803 (Ga. Ct. App. 1996); *Garrett*, 811 P.2d 488 (Idaho 1991); *State v. Buening*, 592 N.E.2d 1222 (Ill. Ct. App. 1992); *State v. Witte*, 836 P.2d 1110 (Kan. 1992); *State v. Armstrong*, 561 So.2d 883 (La. Ct. App. 1990); *Schultz*, 664 A.2d 60 (Md. App. 1995); *People v. Berger*, 551 N.W.2d 421 (Mich. Ct. App. 1991); *State v. Klawitter*, 518 N.W.2d

(Minn. 1994); *State v. Baue*, 607 N.W.2d 191 (Neb. 2000); *State v. Cissne*, 865 P.2d 564 (Wash. Ct. App. 1994). Some courts, however, have used other evidentiary standards. See, e.g., *Connecticut v. Russo*, 777 A. 2d 965 (Conn. App. Ct. 2001) (remanding case to trial court to evaluate admissibility of HGN evidence under *Daubert* standard adopted by the Connecticut Supreme Court in 1997); *State v. Ito*, 978 P.2d 191 (Haw. Ct. App. 1999); *Hulse v. State*, 961 P. 2d 75 (Mont. 1998);³⁸ *New Hampshire v. Duffy*, 778 A.2d 415 (N.H. 2001) (using state evidence Rule 702 that requires showing of reliability before HGN evidence can be admitted; remanding to trial court to hold a hearing on the test's reliability); *State v. Torres*,³⁹ 976 P.2d

³⁸ The *Hulse* court held that neither the *Frye* nor *Daubert* tests were applicable to admissibility of HGN evidence because those tests were restricted to admissibility of "novel" scientific evidence and HGN test was not "novel" science. 961 P.2d at 91. Instead, the court applied Montana Evidence Rule 702, which was identical to the then current version of Fed. R. Evid. 702. The court did not rule on the admissibility of HGN evidence in a DWI/DUI criminal trial, as the appeal arose from a trial court decision denying Hulse's petition to reinstate driving privileges after they were suspended because Hulse refused to take a breathalyzer, and the only legal issues presented were the existence of probable cause to arrest for DWI/DUI, and the driver's refusal to take a breath test. *Id.* at 91-92.

³⁹ In *Torres*, the court made several significant rulings. First, it held that police officers are not qualified to testify about the scientific bases underlying the HGN test and are not competent to establish that the test is reliable. 976 P.2d at 32. It further held that it "is improper to look for

20 (N.M. 1999) (reversing trial court's ruling that HGN evidence was admissible, remanding for hearing using *Daubert* test).⁴⁰

Third, of the state cases where the courts undertook the task of evaluating the admissibility of HGN evidence, the NHTSA studies and, in many instances, the testimony of Dr. Burns,

scientific acceptance only from reported case law," and it declined to take judicial notice of the reliability of the HGN test because "[w]e are not persuaded that HGN testing is 'a subject of common and general knowledge,' or a matter 'well established and authoritatively settled.'" *Id.* at 33.

Finally, the court held that, although a qualified expert was needed to testify about the reliability of the HGN test and its results, a properly trained police officer could testify about the administration of the test "after an appropriate foundation regarding such [scientific] knowledge has been laid by another, scientific expert." *Id.* at 34. The care taken by the *Torres* court illustrates the difference in application of the *Daubert* test from the *Frye* test. *Daubert* requires analysis of the methodology used, its reliability and validity. *Frye*, on the other hand, may tempt a court faced with determining the admissibility simply to see what other courts have done in the past, as well as review publications supplied by the parties, or found by the court's own efforts, without engaging in the sometimes difficult analysis of the reliability of the science or technology underlying those sources.

⁴⁰ *Ito* used Hawaii Evidence Rule 702, which, in addition to the requirements of the then current version of Fed. R. Evid. 702, added the provision that the court "may consider the trustworthiness and validity of the scientific technique or mode of analysis employed by the proffered expert." 978 P.2d at 200. The court held that judicial notice of the reliability of HGN evidence was not proper under Hawaii Evidence Rule 201 but that judicial notice of its reliability was proper under Hawaii common law which permits a trial court to take judicial notice of facts judicially noticed in case law from other jurisdictions. *Id.* at 208-09. In doing so, the court relied heavily on the Maryland *Schultz* opinion.

figured prominently in their conclusions that the HGN tests were admissible as evidence of intoxication or impairment. See, e.g., *Ballard v. State*, 995 P.2d 931 (Alaska Ct. App. 1998)(court relied on trial testimony of Dr. Burns, NHTSA training video and testimony of state trooper. Defendant called a psychology professor and neuro-ophthalmologist); *State v. Superior Court*, 718 P.2d 171 (Ariz. 1986) (court considered trial court testimony of Dr. Burns, two police officers, NHTSA studies, and published articles on HGN test); *People v. Joehnk*, 42 Cal. Rptr. 2d 6 (Ca. Ct. App. 1995)(court considered trial testimony of Dr. Burns, NHTSA studies, testimony of a "criminalist" and a toxicologist. Defendant called an emergency room doctor to testify); *State v. Ruthhardt*, 680 A.2d 349 (Del. Super. Ct. 1996) (court considered trial testimony of Dr. Burns, NHTSA studies, testimony of police officer, behavioral optometrist and neuro-ophthalmologist, defense introduced testimony of Dr. Cole, one of the defense witnesses in the pending case); *Williams v. State*, 710 So.2d 24 (Fla. Ct. App. 1998) (Dr. Burns, a neurologist and three state doctors called as witnesses by the state); *Hawkins v. State*, 476 S.E.2d 803 (Ga. Ct. App. 1996) (court relied on NHTSA studies, other state court rulings and articles); *State v. Hill*, 865 S.W.2d 702 (Mo. Ct. App. 1993) (Dr. Burns only witness called at

trial on HGN test); *State v. O'Key*, 889 P.2d 663 (Or. 1995)(court considered testimony of Dr. Burns, an optometrist, police officer and NHTSA studies).

Finally, those courts that did not undertake an independent evaluation of the admissibility of HGN evidence tended simply to cite to the decisions of other state courts. *See, e.g., Malone v. City of Silverhill*, 575 So. 2d 101 (Ala. Crim. App. 1989); *Hawkins v. State*, 476 S.E.2d 803 (Ga. Ct. App. 1996); *State v. Garrett*, 811 P.2d 488 (Idaho 1991); *State v. Buening*, 592 N.E.2d 1222 (Ill. App. Ct. 1992); *State v. Murphy*, 451 N.W.2d 154 (Iowa 1990); *State v. Breitung*, 623 So. 2d 23 (La. Ct. App. 1993); *State v. Bresson*, 554 N.E.2d 1330 (Ohio 1990); *State v. Cissne*, 865 P.2d 564 (Wash. Ct. App. 1994); *State v. Zivcic*, 598 N.W.2d 565 (Wis. Ct. App. 1999).

B. Difference between Daubert/Kumho Tire/New Rule 702 and Frye.

The difference in approach between the *Daubert/Kumho Tire*/New Rule 702 and the *Frye* tests reveals an unmistakable irony. The *Frye* approach to admissibility of scientific evidence was criticized widely as being too "rigid" because it would deny admissibility to evidence that was the result of new scientific discovery that, while factually sound and methodologically

reliable, had not yet gained general acceptance. Christopher Mueller & Laird Kirkpatrick, *Evidence* § 7.8 (4th ed. 1995); 29 Charles Alan Wright & Victor James Gold, *Federal Practice and Procedure* § 6266 (1997). Under the *Daubert* test, however, general acceptance was but one of the evaluative factors and, provided the evidence at issue was subject to being tested, did not suffer from an unacceptably high error rate and favorably had been peer reviewed, the evidence would be admitted because it was reliable. Under *Daubert*, therefore, it was expected that it would be easier to admit evidence that was the product of new science or technology.

In practice, however, it often seems as though the opposite has occurred--application of *Daubert/Kumho Tire* analysis results in the exclusion of evidence that might otherwise have been admitted under *Frye*. Although this may have been an unexpected outcome, it can be explained by the difference in methodology undertaken by the trial courts when measuring proffered evidence under *Daubert/Kumho Tire*, as opposed to *Frye*. Under *Daubert*, the parties and the trial court are forced to reckon with the factors that really do determine whether the evidence is reliable, relevant and "fits" the case at issue. Focusing on the tests used to develop the evidence, the error rates involved, what the learned publications in the field have said when evaluating it

critically, and then, finally, whether it has come be generally accepted, is a difficult task. But, if undertaken as intended, it does expose evidentiary weaknesses that otherwise would be overlooked if, following the dictates of *Frye*, all that is needed to admit the evidence is the testimony of one or more experts in the field that the evidence at issue derives from methods or procedures that have become generally accepted. Wright & Gold, 29 Federal Practice and Procedures § 6266 ("Daubert's focus upon multiple criteria for scientific validity compels the lower courts to abandon long existing per se rules of admissibility or inadmissibility grounded upon the *Frye* standard.").

Daubert's challenge is unmistakable. While courts may be skilled at research and analysis, the task of deciding the admissibility of new or difficult scientific or technical evidence involves subject matters that are highly specialized, and there is a risk that the court, forced to resolve an issue without the luxury of unlimited time to reflect on it, will get it wrong. This is especially true because judges do not determine the reliability of scientific or technical issues in the abstract but rather in the context of deciding a specific

dispute.⁴¹

The principle shortcoming of *Frye* was that it excused the court from even having to try to understand the evidence at issue. 4 Jack B. Weinstein & Margaret A. Berger, *Weinstein's Federal Evidence*, § 702.05[1] (2d ed. 1997) (Under *Frye* "[t]he court itself did not have to comprehend the science involved . . . [it] only had to assure itself that among the people involved in the field, the technique was acceptable as reliable."). Further, given the impact of the *stare decisis* doctrine, once a court, relying on *Frye*, had ruled that a doctrine or principle had attained general

⁴¹ Justice Stephen Breyer, all too aware of this problem, wrote in the introduction to the *Reference Manual on Scientific Evidence* 4 (2d ed. 2000):

[M]ost judges lack the scientific training that might facilitate the evaluation of scientific claims or the evaluation of expert witnesses who make such claims. Judges are typically generalists, dealing with cases that can vary widely in subject matter. Our primary objective is usually process-related: seeing that a decision is reached in a timely way. And the decision of a law court typically . . . focuses on a particular event and specific individualized evidence.

See also Mueller & Kirkpatrick, *Evidence* § 7.8 (4th ed. 1995) ("The main difficulty [with the Daubert case] is that courts are ill equipped to make independent judgments on the validity of science. Most judges are not scientists, and they do not have the time to spend at trial or beforehand to make fully considered decisions on validity.").

acceptance, it was all too easy for subsequent courts simply to follow suit. Before long, a body of case law could develop stating that a methodology had achieved general acceptance without there ever having been a contested, detailed examination of the underpinnings of that methodology. The admissibility of SFST evidence illustrates this hazard, as a review of the state cases reveals that, despite more than sixteen years of case law relating to this evidence, the number of instances where there have been factually well-developed and detailed challenges to the reliability and validity of the tests is extremely small.

Following the *Kumho Tire* decision and the December 2000 changes to Rule 702, a detailed analysis of the factual sufficiency and reliability of the methodology underlying expert testimony is required for all scientific, technical or specialized evidence, not just "novel scientific" evidence. This has required, at times, a reexamination of the admissibility of evidence that long has been admitted under the *Frye* test, which may result in exclusion of evidence that for years routinely has been admitted. See, e.g., *United States v. Llera Plaza*, 2002 WL 32697 (E.D. Pa. January 9, 2002) (excluding aspects of evidence of latent fingerprint identification evidence on the basis of *Daubert/Kumho Tire* and Rule 702 analysis). As lawyers and courts become fully aware of the relatively recent additional

requirements of *Kumho Tire* and revised Rule 702, this process of reexamination can be expected to continue. It may mean, in a very real sense, that "everything old is new again" with respect to some scientific and technical evidentiary matters long considered settled. Alarmists may see this as undesirable, envisioning courtrooms populated by mad scientists in white lab coats and overzealous judges in black robes, busily undoing established precedent. The more probable outcome is that judges, lawyers and expert witnesses will have to learn to be comfortable refocusing their thinking about the building blocks of what truly makes evidence that is beyond the knowledge and experience of lay persons useful to them in resolving disputes. The beneficiaries of this new approach will be the jurors that have to decide increasingly complex cases. *Daubert*, *Kumho Tire*, and now Rule 702 have given us our marching orders, and it is up to the participants in the litigation process to get in step.

C. Applying Daubert/Kumho Tire and Rule 702 in this Case

Many of the state cases debate whether SFST evidence is "scientific" or "novel science," and therefore subject to *Frye* analysis in the first instance.⁴² Under the Federal Rules of

⁴² See, e.g., *Schultz v. State*, 664 A.2d 60 (Md. App. 1995) (discussing whether HGN and other SFSTs are "scientific

Evidence, this debate is irrelevant, as newly revised Rule 702 and the *Daubert/Kumho Tire* cases require the same analysis for any evidence that is to be offered under Rule 702. Thus, if the SFSTs in this case are being offered as direct evidence of intoxication or impairment, they then become cloaked in a scientific or technical aura, and the factors articulated in *Daubert/Kumho Tire* and Rule 702 must be evaluated by the district court under Rule 104(a) before such evidence may be admitted.⁴³

With regards to the HGN test, from the testimony before me, the materials submitted for my review by counsel, my review of all of the state cases decided to date, and many of the articles cited in those cases, it cannot be disputed that there is a sufficient factual basis to support the causal connection between observable exaggerated horizontal gaze nystagmus in a suspect's eye and the ingestion of alcohol by that person. This connection is so well established that it is appropriate to be judicially

evidence"); *Hulse v. State*, 961 P.2d 75(Mont. 1998).

⁴³ If offered only as circumstantial evidence of intoxication/impairment, the HGN test still clearly invokes scientific and technical underpinnings. The WAT and OLS SFSTs, however, involve only observations of the suspect's performance, and therefore, it may be argued that they are not couched in science and technology if used for that purpose.

noted under Rule 201.⁴⁴ That being said, however, it must quickly be added that there also are many other causes of nystagmus that are unrelated to alcohol consumption. The *Schultz* court identified thirty-eight possible causes of nystagmus,⁴⁵ and, in

⁴⁴ The existence of a causal connection between alcohol ingestion and observable horizontal gaze nystagmus is the type of discrete adjudicative fact that properly may be judicially noticed under Rule 201 because it is a fact that can be accurately and readily determined by resort to sources whose accuracy cannot reasonably be questioned. This use of judicial notice is far more narrow than attempting to take judicial notice, as did the Court of Special Appeals in *Schultz*, that the SFSTs have attained general acceptance within the relevant scientific or technical community. Alternatively, the government may prove the causal relationship between alcohol consumption and exaggerated nystagmus by expert testimony, but in this regard I agree with the New Mexico Supreme Court's decision in *State v. Torres*, which held that a police officer is unlikely to have the qualifications needed to testify under Rule 702 as to the scientific principles underlying the HGN test or as to whether there is a causal link between alcohol use and exaggerated nystagmus. 976 P.2d at 32, 34. Accordingly, asking the court to take judicial notice of this causal connection likely will be the most frequent method used by the government to prove this essential fact. An alternative would be to use learned treatises, under Rule 803(18), if a proper foundation first is established. The police officer will, of course, be qualified to testify as to the training received in how to administer the HGN test, and to demonstrate his or her qualifications properly to administer it. Because Officer Jarrell did not testify at the Rule 104(a) hearing, there is no factual basis before me at this time to permit me to make findings regarding the final factor under Rule 702, i.e., whether Jarrell properly administered and interpreted the SFSTs given to Horn.

⁴⁵ The court recognized the following causes or possible causes of nystagmus: problems with the inner ear labyrinth; irrigating the ears with warm or cold water; influenza; streptococcus infection; vertigo; measles; syphilis;

his testimony, Colonel Rabin agreed that most of the *Schultz* factors did, or possibly could, cause nystagmus in humans. Thus, the detectable presence of exaggerated HGN in a driver clearly is circumstantial, not direct, evidence of alcohol consumption.

As for the sufficiency of the facts and data underlying the assertions in the NHTSA articles that SFSTs are reliable in predicting specific BAC, the testimony of Horn's experts, as well as the literature that is critical of these studies, establishes that presently there is insufficient data to support these claims of accuracy. The early NHTSA laboratory tests were too limited

arteriosclerosis; Korchaff's syndrome; brain hemorrhage; epilepsy; hypertension; motion sickness; sunstroke; eye strain; eye muscle fatigue; glaucoma; changes in atmospheric pressure; consumption of excessive amounts of caffeine; excessive exposure to nicotine; aspirin; circadian rhythms; acute head trauma; chronic head trauma; some prescription drugs; tranquilizers, pain medication, and anti-convulsant medicine; barbiturates; disorders of the vestibular apparatus and brain stem; cerebellum dysfunction; heredity; diet; toxins; exposure to solvents; extreme chilling; eye muscle imbalance; lesions; continuous movement of the visual field past the eyes; and antihistamine use. 664 A.2d at 77. The fact that there are many other causes of nystagmus in the human eye also is the type of adjudicative fact that may be judicially noticed under Rule 201. Thus, the defendant in a DWI/DUI case may ask the court to judicially notice this fact, once the government has proved the causal connection between alcohol ingestion and exaggerated nystagmus. Alternatively, the defendant may seek to prove the non-alcohol related causes of nystagmus by other means, such as the testimony of an expert witness, cross examination of any such witness called by the government or through a properly admitted learned treatise. (Fed. Rule of Evid. Rule 803(18)).

to support the claims of accuracy, and the subsequent field and validation testing insufficient to establish the reliability and validity of the tests if used to establish specific BAC. Indeed, the great weight of the state authority, including that in Maryland, agrees that BAC levels may not be proved by SFST test results alone, and I adopt that holding here.

The conclusion I have reached regarding the reliability of the methods and principles underlying the SFSTs takes into account the evidence introduced by Horn about the methods used to develop these tests, and the error rates associated therewith--the first two *Daubert/Kumho Tire* factors. This alone precludes their admissibility to prove specific BAC, and it therefore is not necessary to discuss in detail whether the many articles written about these tests constitute peer review analysis or something else, and whether they generally have been accepted in a relevant, unbiased scientific or technical community, the third and fourth *Daubert/Kumho Tire* factors. I do note, however, the testimony of Horn's experts that the NHTSA publications regarding the SFSTs do not constitute peer review publications, a conclusion that seems correct. As Dr. Cole testified, peer review as contemplated by *Daubert* and *Kumho Tire* must involve critical analysis that can expose any weaknesses in the methodology or principles underlying the conclusions being

reviewed.

Further, as testified to by Horn's experts, the process of selection of articles for publication in a peer review journal involves an evaluation by one or more experts in the field, to insure that the article meets the rigors of that field. Under this standard, most of the publications regarding the SFST tests, including the publications in bar journals, likely do not meet this criteria.

Similarly, despite the conclusion of many state courts that the SFSTs have received general acceptance among criminologists, law enforcement personnel, highway safety experts and prosecutors, I remain skeptical whether this is sufficient for purposes of *Daubert* and *Kumho Tire*. Acceptance by a relevant scientific or technical community implies that that community has the expertise critically to evaluate the methods and principles that underlie the test or opinion in question. However skilled law enforcement officials, highway safety specialists, prosecutors and criminologists may be in their fields, the record before me provides scant comfort that these communities have the expertise needed to evaluate the methods and procedures underlying human performance tests such as the SFSTs. Some might say the same about judges, without fear of too much disagreement, but judges are the ones obligated to do so by Rule 104(a) when

the admissibility of evidence is challenged. As to the conclusion of the state courts, more often than not expressed in passing and without analysis, that the SFSTs generally are accepted among psychologists like Dr. Burns, the evidence presented to me by the three psychologists called by Horn leads me, respectfully, to beg to differ. Thus, based on the foregoing, I conclude that the SFST evidence in this case does not, at this time, meet the requirements of *Daubert/Kumho Tire* and Rule 702 as to be admissible as direct evidence of intoxication or impairment.

A more difficult question, however, is whether the SFSTs may be used as circumstantial evidence of alcohol consumption and, if so, just how. The state courts overwhelmingly have concluded that the results of SFSTs are admissible as circumstantial evidence of alcohol consumption but have offered little guidance about what exactly the testifying officer may tell the fact finder about the SFSTs, their administration, and the performance of the suspect when doing them. The possibilities range from simply describing the tests--without explaining the scientific or technical bases underlying them or their claimed accuracy rates and describing only what the officer observed when they were performed, absent any opinions regarding whether the suspect "passed" or "failed" or assessment of the degree of intoxication

or impairment--to a full explanation of the tests, their claimed accuracy, the number of "standardized clues" the suspect missed, and an opinion that the suspect "failed" the test--in short everything up to testimony about the specific BAC of the driver.

On the record before me there are not sufficient facts or data about the OLS and WAT SFSTs to support the conclusion that, if a suspect exhibits two out of eight possible clues on the WAT test or two out of four clues on the OLS, he has "failed" the tests. To the contrary, Horn introduced Dr. Cole's study that showed an alarmingly high error rate when police officers were asked to evaluate completely sober subjects performing the WAT and OLS.⁴⁶ Def's. Motion Exh. C. To permit a police officer to testify about each of the SFSTs in detail, their claimed accuracy rates, the number of standardized clues applicable to each, the number of clues exhibited by the suspect, and then offer an opinion about whether he or she passed or failed, stopping just short of expressing an opinion as to specific BAC, invites the risk of allowing through the back door of circumstantial proof evidence that is not reliable enough to enter through the front door of direct proof of intoxication or impairment. Such

⁴⁶ See *supra* at pp. 17-18. Cole reported that 46% of the officers that observed videotaped subjects with BAC levels of .0% performing the WAT and OLS tests reported that the subjects had had too much to drink to be driving.

testimony clearly is technical, if not scientific, and may not be admitted unless shown to be reliable under the standards imposed by Rule 702 and *Daubert/Kumho Tire*, which has not been done in this case.

There is no factual basis before me to support the NHTSA claims of accuracy for the WAT and OLS tests or to support the conclusions about the total number of standardized clues that should be looked for or that missing a stated number means the subject failed the test. There is very little before me that suggests that the WAT and OLS tests are anything more than standardized procedures police officers use to enable them to observe a suspect's coordination, balance, concentration, speech, ability to follow instructions, mood and general physical condition--all of which are visual cues that laypersons, using ordinary experience, associate with reaching opinions about whether someone has been drinking.

Indeed, in *Crampton v. State*, 525 A.2d 1087 (Md. App. 1987) the Maryland Court of Special Appeals described field sobriety tests--other than the HGN test--administered by police to motorists as follows:

field sobriety tests are essentially personal observations of a police officer which determine a suspect's balance and ability to speak with recollection. There is nothing 'new' or perhaps even 'scientific' about the exercises that an officer requests a suspect to perform. Those sobriety tests

have been approved by the National Highway Traffic Safety Administration and are simply guidelines for police officers to utilize in order to observe more precisely a suspect's coordination. It requires no particular scientific skill or training for a police officer, or any other competent person, to ascertain whether someone performing simple tasks is to a degree affected by alcohol. The field sobriety tests are designed to reveal objective information about a driver's coordination. . . . The *Frye-Reed* test does not apply to those field sobriety tests because the latter are essentially empirical observations, involving no controversial, new or 'scientific' technique. Their use is guided by practical experience, not theory.

Id. at 1093-94. The same conclusion has been reached by many other state courts that have considered this issue. For example, in *State v. Ferrer*, 23 P.3d 744 (Haw. Ct. App. 2001), the court stated:

It is generally recognized, however, that the foundational requirements for admission of psychomotor FST evidence differ from the foundational requirements for admission of HGN evidence. Psychomotor FSTs test balance and divided attention, or the ability to perform multiple tasks simultaneously. While balancing is not necessarily a factor in driving, the lack of balance is an indicator that there may be other problems. Poor divided attention skills relate directly to a driver's exercise of judgment and ability to respond to the numerous stimuli presented during driving. The tests involving coordination (including the walk-and-turn and the one-leg-stand) are probative of the ability to drive, as they examine control over the subject's own movements. Because evidence procured by administration of psychomotor FSTs is within the common experience of the ordinary citizen, the majority of courts that have addressed the issue generally consider psychomotor FSTs to be nonscientific evidence.

Id. at 760-62 (citations omitted).⁴⁷ As the Florida District Court of Appeals said in *State v. Meador*, 674 So. 2d 826 (Fla. App. 1996):

While the psychomotor FSTs are admissible, we agree with defendants that any attempt to attach significance to defendants' performance on these exercises is beyond that attributable to any of the other observations of a defendant's conduct at the time of the arrest could be misleading to the jury and thus tip the scales so that the danger of unfair prejudice would outweigh its probative value. The likelihood of unfair prejudice does not outweigh the probative value as long as the witness simply describe their observations. Reference to the exercises by using terms such as 'test,' 'fail' or 'points,' however, creates a potential for enhancing the significance of the observations in relationship to the ultimate determination of impairment, as such terms give these layperson observations an aura of scientific validity. Therefore, such terms should be avoided to minimize the danger that the jury will attach greater significance to the results of the field sobriety exercises than to other lay observations of impairment.

Id. at 832.

I agree with this reasoning. If offered as circumstantial evidence of alcohol intoxication or impairment, the probative value of the SFSTs derives from their basic nature as observations of human behavior, which is not scientific, technical or specialized knowledge. To interject into this

⁴⁷ The court cites to decisions from Alabama, Arizona, California, Georgia, Illinois, Maryland, Massachusetts, New York, Pennsylvania, Florida and Oregon that have reached the same conclusion about the nature of psychomotor FSTs like the WAT and OLS tests. *Id.* at 760-62.

essentially descriptive process technical terminology regarding the number of "standardized clues" that should be looked for or opinions of the officer that the subject "failed" the "test," especially when such testimony cannot be shown to have resulted from reliable methodology, unfairly cloaks it with unearned credibility. Any probative value these terms may have is substantially outweighed by the danger of unfair prejudice resulting from words that imply reliability. I therefore hold that when testifying about the SFSTs a police officer must be limited to describing the procedure administered and the observations of how the defendant performed it, without resort to terms such as "test,"⁴⁸ "standardized clues," "pass" or "fail," unless the government first has established a foundation that satisfies Rule 702 and the *Daubert/Kumho Tire* factors regarding the reliability and validity of the scientific or technical

⁴⁸ It would be preferable to refer to the standardized field sobriety tests as "procedures," rather than tests, as the use of the word test implies that there is an accepted method of determining whether the person performing it passed or failed, and this has not been shown in this case. I recognize, however, that the HGN, WAT and OLS procedures have been referred to as field sobriety "tests" for so many years, that it is likely that it will be impossible to stop using this terminology altogether. Occasional reference to the HGN, WAT and OLS procedures as "tests" should not alone be grounds for a mistrial in a jury case. However, repeated use of the word "test" to describe these procedures, particularly when testifying as to how the defendant actually performed them, would be improper.

underpinnings of the NHTSA assertions that there are a stated number of clues that support an opinion that the suspect has "failed" the test.

This is not to say that a police officer may not express an opinion as a lay witness that the defendant was intoxicated or impaired, if otherwise admissible under Rule 701. As recently amended, Rule 701 permits lay opinion testimony if: (a) rationally based upon the perception of the witness, (b) helpful to the fact finder and (c) if the opinion does *not* involve scientific, technical or specialized information.⁴⁹ There is near universal agreement that lay opinion testimony about whether someone was intoxicated is admissible if it meets the above criteria. *See, e.g., Singletary v. Secretary of Health*, 623 F.2d 217, 219 (2d Cir. 1980) ("The testimony of lay witnesses has always been admissible with regard to drunkenness."); *United States v. Mastberg*, 503 F.2d 465 (9th Cir. 1974); *Malone v. City of Silverhill*, 575 So. 2d 101 (Ala. Crim. App. 1990); *State v. Lummus*, 950 P.2d 1190 (Ariz. App. 1997); *Wrigley v. State*, 546 S.E.2d 794, 798 (Ga. App. 2001) ("A police officer may give opinion testimony as to the state of sobriety of a DUI suspect and whether appellant was under the influence."); *State v.*

⁴⁹ Maryland's equivalent evidence rule, 5-701, does not contain the third requirement imposed by the federal rule.

Ferrer, 23 P.3d 744 (Hawaii Ct. App. 2001); *Com. v. Bowen*, 754 N.E.2d 1083 (Ma. App. 2001); *State v. Hall*, 353 N.W.2d 37, 43 (S.D. 1984); *Beats v. State*, 2000 WL 921684 (Tex. Crim. App. 2000) ("A lay witness, including a police officer, may express an opinion about a person's intoxication."). See also John W. Strong, *McCormick on Evidence*

§ 11 (5th ed. 1999) ("The so-called 'collective fact' or 'short-hand rendition rule' [permits] opinions on such subjects as. . . a person's intoxication."); Graham, *Handbook of Federal Evidence* § 701.1 (5th ed. 2001)(lay witness permitted to offer opinion testimony that a person was intoxicated); Mueller and Kirkpatrick, *Evidence* § 7.4 (4th ed. 1995) ("One common example [of the collective facts doctrine] is lay testimony that someone was intoxicated, and here the witness is not confined to descriptions of glazed eyes, problems in speech or motor coordination, changes in behavior or mood or affect, but may say directly (assuming adequate observation and common experience) that the person seemed drunk or under the influence").

In DWI/DUI cases, however, the third requirement of Rule 701, that the lay opinion is "not based on scientific, technical, or other specialized knowledge," will take on great importance. A police officer certainly may testify about his or her observations of a defendant's appearance, coordination, mood,

ability to follow instructions, balance, the presence of the smell of an alcoholic beverage, as well as the presence of exaggerated HGN, and the observations of the defendant's performance of the SFSTs-- consistent with the limitations discussed above. The officer should not, however, be permitted to interject technical or specialized comments to embellish the opinion based on any special training or experience he or she has in investigating DWI/DUI cases. Just where the line should be drawn must be left to the discretion of the trial judge, but the officer's testimony under Rule 701 must not be allowed to creep from that of a layperson to that of an expert--and the line of demarcation is crossed if the opinion ceases to be based on observation and becomes one founded on scientific, specialized or technological knowledge.

CONCLUSION

To summarize, the Court holds that the following rulings apply to the case at bar:

(1) The results of properly administered WAT, OLS and HGN SFSTs may be admitted into evidence in a DWI/DUI case only as circumstantial evidence of intoxication or impairment but not as direct evidence of specific BAC. Recognizing that Officer Jarrell, the arresting police officer in this case, may be the sponsor for this evidence, he must first establish his

qualifications to administer the test. Unless qualified as an expert witness under Rule 702 to express scientific or technical opinions regarding the reliability of the methods and principles underlying the SFSTs, Officer Jarrell's foundational testimony will be limited to the instruction and training received and experience he has in administering the tests and may not include opinions about the tests' accuracy rates. If Officer Jarrell testifies about the results of the HGN test, he may testify as to his qualifications to detect exaggerated HGN, and his observations of exaggerated HGN in the Horn, but may not, absent being qualified under Rule 702 to do so, testify as to the causal nexus between alcohol consumption and exaggerated HGN. When testifying about Horn's performance of the SFSTs, Officer Jarrell may describe the SFSTs he required Horn to perform and describe Horn's performance, but Officer Jarrell may not use language such as "test," "standardized clues" or express the opinion that Horn "passed" or "failed," because the government has not shown, under Rule 702 and the *Daubert/Kumho Tire* decisions, that these conclusions are based on sufficient facts or data and are derived from reliable methods or principles.

(2) The government may prove the causal connection between exaggerated HGN in Horn's eyes and alcohol consumption by one of the following means: asking the court to take judicial notice of

it under Rule 201; the testimony of an expert qualified under Rule 702; or through learned treatises, introduced in accordance with Rule 803(18). In response to proof of the causal connection between alcohol consumption and exaggerated HGN, Horn may prove that there are other causes of HGN than alcohol by one of the following methods: asking the court to take judicial notice of this fact under Rule 201; cross-examining any expert called by the government; by calling a defense expert witness, qualified under Rule 702, or through leaned treatises, introduced in accordance with Rule 803(18).

(3) Assuming the government can establish the elements of Rule 701, Officer Jarrell may give lay opinion testimony that Horn was intoxicated or impaired by alcohol. Such testimony must be based on Officer Jarrell's observations of Horn and may not include scientific, technical or specialized information.

Date _____

Paul W. Grimm
United States Magistrate Judge