2008

From a Plane Crash to the Conviction of an Innocent Person: A call on lawmakers to establish that forensic evidence is inadmissible unless forensic equipment is developed as a safety-critical system

Dr. Boaz Sangero
Dr. Mordechai Halpert

From a Plane Crash to the Conviction of an Innocent Person: Why Forensic Science Evidence Should Be Inadmissible Unless it has been Developed as a Safety-critical System

Mordechai Halpert and Boaz Sangero[*]

"When playing Russian roulette the fact that the first shot got off safely is little comfort for the next"

– Richard Feynman[1]

A B S T R A C T

According to existing law, a criminal conviction may be based on a single piece of scientific (forensic) evidence. Thus, for example, a DNA match could, on its own, lead to a conviction and a prolonged term of imprisonment, or even a death sentence. A testing error might result in the conviction of an innocent person. Therefore, the state ought to have a duty to ensure that such evidence is as reliable as possible. This article protests an inconceivable situation: that the development of forensic equipment, which is designed to produce evidence that can be relied on in a criminal trial, is not monitored and regulated by the state – and such equipment is not developed as carefully as it should be. This is in sharp contrast to the development of safety-critical systems, such as medical devices or airplanes, which is monitored by the authorities and must withstand strict standards of quality assurance. The basic principle in developing safety-critical systems is that it is impossible to guarantee the safety of the device by only testing the final product. Instead, safety must be built in to the entire life cycle of the device. For this purpose, the government has established authorities, such as the US Food and Drug Administration (FDA), whose role is to ensure that safety-critical medical devices are developed carefully, according to accepted principles of safety engineering. These principles are set forth in regulations. However, this is not the case with forensic equipment. Regulations dealing with the quality assurance of forensic devices do not relate to their development, but only to the process whereby evidence is produced using them. Thus, for example, there is no requirement for the manufacturer of a breathalyzer – used to detect

[*] Dr. Mordechai Halpert is a physicist involved, among other things, in the research and development of voice biometric systems. Dr. Boaz Sangero is the Head of the Department of Criminal Law and Criminology at the Academic Center of Law and Business, in Israel. Our thanks go to Dr. Rinat Kitai, Attorney Moshe Pardess, and Attorney Neil Zwail for their helpful comments on previous drafts of this article.

[1] Richard P. Feynman, *Richard Feynman's Minority Report to the Space Shuttle* Challenger *Inquiry*, in RICHARD P. FEYNMAN, THE PLEASURE OF FINDING THINGS OUT: THE BEST SHORT WORKS OF RICHARD P. FEYNMAN, 155 (Jeffrey Robbins, ed., 1999), also available in the NASA Space Shuttle Launch Archive at the Kennedy Space Center website, as *Appendix F - Personal Observations on the Reliability of the Shuttle*: http://science.ksc.nasa.gov/shuttle/missions/51-l/docs/rogers-commission/Appendix-F.txt.

drunken driving – to develop the device with proper caution, according to standards of quality assurance. This device is only evaluated based on its final performance, as a "black box," contrary to accepted principles of safety engineering, which could lead to errors. We show why devices used to produce forensic evidence for a criminal trial should be treated like safety-critical systems. We call on lawmakers to regulate the monitoring of the development of forensic equipment just as this has been done for other safety-critical systems. We also call on lawmakers to enact legislation establishing the inadmissibility of forensic evidence in a criminal trial unless the device producing it has been developed carefully, as a safety-critical system.

I. Introduction

In a criminal trial, it is acceptable to rely on scientific (forensic) evidence. Forensic evidence is often produced through devices comprising both hardware and software, such as a breathalyzer – which measures the level of alcohol in an expired breath – and the equipment used to create a DNA profile.[2] According to existing law, a person may be convicted on the basis of a single piece of evidence.[3] Once a person has been convicted of a serious offense, such as murder or rape, an "ultimate penalty"[4] may be imposed on him: life imprisonment or even a death sentence. Many studies conducted in the past few decades have revealed a significant phenomenon whereby innocent persons are convicted – a phenomenon that can no longer be doubted and ignored.[5] The inaccuracy or unreliability of forensic equipment could lead to the death or prolonged incarceration of an innocent person.

Other scientists and legal scholars have pointed out that some areas within the forensic sciences are nearly barren of a scientific foundation, that error rates are in fact higher than usually acknowledged by forensic scientists, and that laboratory regulation and quality control are weak compared to medical, research, other

---

[2] William C. Thompson et al., *Evaluating Forensic DNA Evidence: Essential Elements of a Competent Defense Review: Part 1*, 27 THE CHAMPION 16, 18 (2003); Jennifer N. Mellon, *Manufacturing Convictions: Why Defendants Are Entitled to the Data Underlying Forensic DNA Kits*, 51 DUKE L.J. 1097, 1098 (2001).

[3] Boaz Sangero & Mordechai Halpert, *Why a Conviction Should Not Be Based on a Single Piece of Evidence: A Proposal for Reform*, 48 JURIMETRICS J. 43, 62-63 (2007).

[4] *See* LEON SHELEF, ULTIMATE PENALTIES (1987).

[5] *See*: Boaz Sangero, *Miranda Is Not Enough: A New Justification for Demanding "Strong Corroboration" to a Confession*, 28 CARDOZO L. REV. 2791, 2792-94 (2007); Hugo A. Bedau & Michael L. Radelet, *Miscarriages of Justice in Potentially Capital Cases*, 40 STAN. L. REV. 21 (1987); ARYEH RATTNER, CONVICTING THE INNOCENT: WHEN JUSTICE GOES WRONG (Ph.D. dissertation, Ohio State University, 1983); Aryeh Rattner, *Convicted but Innocent: Wrongful Conviction and the Criminal Justice System*, 12 LAW & HUM. BEHAV. 283 (1988); Richard A. Leo & Richard J. Ofshe, *The Consequences of False Confession: Deprivations of Liberty and Miscarriages of Justice in the Age of Psychological Interrogation*, 88 J. CRIM. L. & CRIMINOLOGY 429 (1998) (and see the references to numerous additional sources in footnotes 1-4 of their article); BARRY SCHECK ET AL., ACTUAL INNOCENCE: FIVE DAYS TO EXECUTION AND OTHER DISPATCHES FROM THE WRONGFULLY CONVICTED (2000); Website of the Innocence Project *at* www.innocenceproject.org (last visited July 17, 2008); Keith A. Findley, *Learning from Our Mistakes: A Criminal Justice Commission to Study Wrongful Convictions*, 38 CAL. W. L. REV. 333 (2002); Elizabeth V. Lafollette, State v. Hunt *and Exculpatory DNA Evidence: When Is a New Trial Warranted?*, 74 N.C. L. REV. 1295 (1996); David De Foore, *Postconviction DNA Testing: A Cry For Justice From The Wrongly Convicted*, 33 TEX. TECH. L. REV. 491 (2002); Karen Christian, *"And the DNA Shall Set You Free": Issues Surrounding Postconviction DNA Evidence and the Pursuit of Innocence*, 62 OHIO ST. L.J. 1195 (2001); DONALD S. CONNERY, ED., CONVICTING THE INNOCENT (1996); THE ROYAL COMMISSION ON CRIMINAL JUSTICE: REPORT PRESENTED TO PARLIAMENT (July 1993), Chairman: Viscount Runciman of Doxford CBE FBA; Lissa Griffin, *The Correction of Wrongful Convictions: A Comparative Perspective*, 16 AM. U. INT'L L. REV. 1241 (2001).

government, and private industry labs.[6] The present paper adds something new to all of that: it draws attention to the development of software and hardware used in crime laboratories, and of other forensic equipment used by the police, and questions why those are not held to the same development standards as other safety-critical devices in other realms where the safety of the public is concerned.

As we shall see below, safety-critical systems, which pose a danger to human life or body integrity or could lead to significant economic loss, are found in many fields and industries, such as medicine, aeronautics, and transportation. In such fields, statutory rules have been enacted that are designed to minimize the danger of mishaps that could lead to catastrophes and significant damage; and regulatory bodies have been established whose role it is to monitor the procedures used to develop such systems and to guarantee the safety of the end product. These bodies are not satisfied with a test of the final product as a "black box," which is only concerned with the device's output and not the way it works or its method of development.

Instead, regulatory authorities demand that the device be developed safely at all stages. In particular, the manufacturer is subject to strict requirements related to the software of the device. One of the principles of software validation is that testing the program after it has been coded is not enough to verify software integrity and coherence; it is also necessary to ensure that the design, development and coding of the software, as well as its testing, were conducted in a manner that has prevented the introduction of defects.

As we shall see below, the same caution is not applied to devices used to produce forensic evidence, despite the potential of such evidence to determine the fate of an individual. This is the case in particular with regard to evidence obtained through the use of computer software, which the courts are still willing to admit with an unfounded, dangerous optimism. This lack of caution is reflected in the evaluation of scientific evidence, while the device used to produce it is treated solely as a "black box." The result is that findings obtained from a device such as a breathalyzer – which would not be approved based on standards applied to medical devices – are

[6] Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCIENCE 892, 895 (2005); Paul C. Giannelli, *Regulating Crime Laboratories: The Impact of DNA Evidence*, 15 J.L. & POL'Y 59, 72 (2007); Jonathan J. Koehler et al., *The Random Match Probability in DNA Evidence: Irrelevant and Prejudicial?*, 35 JURIMETRICS J. 201, 206–11 (1995); William C. Thompson, *A Sociological Perspective on the Science of Forensic DNA Testing*, 30 U.C. DAVIS L. REV. 1113 (1997).

considered admissible evidence for the purpose of convicting a person in a court of law.

In Part II of this article, we will first explain, in detail, what safety-critical systems are, the importance of the standards applied to their development, and what system safety engineering is. In Part III, we will illustrate the current attitude in a court of law towards scientific evidence, using the breathalyzer, which is used to detect drunken driving, as a test case. In Part IV, we will try to convince the reader that it is essential to develop forensic equipment as a safety-critical system. Finally, in the Epilogue, we call on lawmakers to subject forensic evidence to the same standards applying to safety-critical systems, in order to reduce the danger of convicting the innocent.

II. Safety-critical systems

A. Catastrophes caused by a failure to follow special rules of safety

One famous case, from the field of medicine, occurred in 1937.[7] Sulfanilamide, a drug for treating streptococcal infections, was shown to have dramatic curative effects and had been used safely for some time in both tablet and powder form. However, the manufacturer discovered that there was also a demand for the drug in liquid form. For this purpose, sulfanilamide was dissolved into diethylene glycol. Since this was changed little from its original state, which was safe, the manufacturer did not bother to test the toxicity of the new formulation of the drug sold in liquid form. The result was fatal – 100 people died from taking the drug. Apparently, liquid diethylene glycol above a certain dosage is a deadly poison.[8] These deaths led to the passage of the 1938 Food, Drug, and Cosmetic Act, which increased the authority of the Food and Drug Administration (FDA) to regulate drugs.[9]

[7] Carol Ballentine, *Taste of Raspberries, Taste of Death - The 1937 Elixir Sulfanilamide Incident*, FDA CONSUMER MAGAZINE (June 1981), *at* http://www.fda.gov/oc/history/elixir.html.
[8] The lethal dose is estimated at 1–1.5 ml/kg of body weight, or approximately 100 ml for an adult. *See* Martina Krenova & Daniela Pelclova, *Complete Recovery After Repeated Suicidal Ethylene Glycol Ingestion*, 107(1) PRAGUE MED. REP. 130, 131 (2006), *available at* http://pmr.cuni.cz/Data/files/PragueMedicalReport/PMR_01-2006%20Krenova.pdf.
[9] Ballentine, *supra* note 7.

A further example is the crash of the maiden flight of the Ariane 5 rocket, on June 4, 1996.[10] Forty seconds following the initiation of the flight sequence, the rocket veered off its flight path at an altitude of about 3,700 meters, broke up, and exploded.[11] This failure was caused by the complete loss of guidance and attitude information thirty-seven seconds into the start of the main engine ignition sequence. The loss of information was the result of specification and design errors in the software of the inertial reference system.[12] The inquiry board investigating the disaster found that the extensive reviews and tests that were carried out during the Ariane 5 development program failed to include sufficient analysis and testing of the inertial reference system or the complete flight control system, which could have identified the potential failure.[13] This is what the inquiry board had to say about the inadequate manner in which the developers had dealt with the software (italics added):

> An underlying theme in the development of Ariane 5 is the bias towards the mitigation of random failure. The supplier of the SRI was only following the specification given to it, which stipulated that in the event of any detected exception the processor was to be stopped. The exception was detected, but inappropriately handled because *the view had been taken that software should be considered correct until it is shown to be at fault.* The Board has reason to believe that this view is also accepted in other areas of Ariane 5 software design. *The Board is in favour of the opposite view, that software should be assumed to be faulty until applying the currently accepted best practice methods can demonstrate that it is correct.*[14]

---

[10] *ARIANE 5: Flight 501 Failure*, Report by the Inquiry Board (Paris, July 19, 1996), *at* http://www.ima.umn.edu/~arnold/disasters/ariane5rep.html.
[11] *Id.* at Foreword.
[12] *Id.* at Section 3.2. The software exception was caused during the execution of a data conversion from 64-bit floating point to 16-bit signed integer value. The floating-point number that was converted had a value greater than what could be represented by a 16-bit signed integer. This lead to an Operand Error that occurred in a part of the software that serves no purpose after the launcher lifts off. (*Id.* at Section 2.1).
[13] *Id.* at Section 3.2.
[14] *Id.* at Section 2.2.

Both of these catastrophes – the first leading to a loss of life and the second to a substantial financial loss – as well as others,[15] could have been avoided had proper techniques for developing safety-critical systems been applied.

B. The definition of a safety-critical system

A safety-critical system may be defined as "one in which a malfunction could result in death, injury or illness, major economic loss, mission failure, environmental damage, or property damage."[16] Title 14 (Aeronautics and Space) of the Code of Federal Regulations offers the following definition: "*Safety-critical* means essential to safe performance or operation. A safety-critical system, subsystem, condition, event, operation, process or item is one whose proper recognition, control, performance or tolerance is essential to system operation such that it does not jeopardize public safety."[17] Such systems are found in airplanes, space shuttles, drugs, medical devices, railway control systems, automobiles, etc.[18] Other examples of safety-critical systems are the aeronautics of airplanes and the anti-lock braking system currently found on automobiles.[19] Failure in such systems could lead to catastrophic damage. Accordingly, their development and manufacture demand the use of engineering methods designed to reduce such danger.

C. A short history of system safety and its lessons

[15] For other examples, see Nancy Leveson & Clark S. Turner, *An Investigation of the Therac-25 Accidents*, 25(7) IEEE COMPUTER 18 (1993) (discussing six software-related accidents between June 1985 and January 1987, involving the Therac-25 computerized radiation therapy machine, which caused massive overdoses of radiation with resultant deaths and serious injuries), http://courses.cs.vt.edu/~cs3604/lib/Therac_25/Therac_1.html.

[16] Frances E. Zollers et al., *No More Soft Landings for Software: Liability For Defects in an Industry That has Come of Age*, 21 SANTA CLARA COMPUTER & HIGH TECH. L.J. 745, 751 (2005).

[17] 14 C.F.R. § 401.5, *available at* http://law.justia.com/us/cfr/title14/14-4.0.2.7.2.0.24.3.html.

[18] NEIL STOREY, SAFETY-CRITICAL COMPUTER SYSTEMS 1-2 (1996); US DEP'T OF DEFENSE, STANDARD PRACTICE FOR SYSTEM SAFETY (2000), http://www.faa.gov/library/manuals/aviation/risk_management/ss_handbook/media/app_h_1200.PDF (hereinafter: "DOD STANDARD PRACTICE"); US FEDERAL AVIATION ADMINISTRATION, SYSTEM SAFETY HANDBOOK (2005), http://www.faa.gov/library/manuals/aviation/risk_management/ss_handbook (hereinafter: "FAA HANDBOOK"; US AIR FORCE, SYSTEM SAFETY HANDBOOK (2000), http://www.system-safety.org/Documents/AF_System-Safety-HNDBK.pdf (hereinafter: "USAF HANDBOOK").

[19] STOREY, *supra* note 18, at 1.

Although engineers have long been concerned with the safety of their products, safety engineering began as a separate discipline only after the Second World War.[20] Up until the development of modern safety engineering, the accepted approach in aircraft production was "fly-fix-fly."[21] This meant that a plane was flown until the occurrence of an accident. The cause of the accident was then determined through an investigation,[22] after which lessons were drawn and action was taken to prevent or minimize the reoccurrence of accidents with the same cause.[23] These lessons were then introduced as a developmental standard and even as a regulation, until the next accident, when the process would be repeated in order to ascertain the new cause of failure.[24]

The "fly-fix-fly" method helps to prevent future accidents by dealing with flaws that have been revealed as the result of a past accident.[25] It does not provide a good solution for preventing new accidents that occur because of other defects.[26] The problem is even more serious when technology is rapidly changing or when learning from experience proves too costly. In the 1940s, there were ten aircraft mishaps per every hundred thousand flight-hours.[27] When the cost of the planes rose, the damages from plane crashes became too expensive. Preventing accidents before they occur became the goal of safety engineering,[28] i.e., to make a product "first-time safe."[29]

System safety began in the 1940s as a grass-roots movement. Gaining further momentum in the 1950s, it became established in the 1960s.[30]

An early leader in the development of system safety was the United States Air Force.[31] Boeing developed the first formal system safety program plan (SSPP) in 1960, for the Minuteman program. The Bureau of Naval Weapons issued the first military specification for safety design requirements in 1961. In 1964, at the

---

[20] Nancy Leveson et al., *Effectively Addressing NASA's Organizational and Safety Culture: Insights from Systems Safety and Engineering Systems*, p. 2 (paper presented at the Engineering Systems Symposium, MIT: March 29-31, 2004), *at* http://esd.mit.edu/symposium/pdfs/papers/leveson-c.pdf.
[21] *Id.* See also USAF HANDBOOK, *supra* note 18, at 3.
[22] Leveson et al., *supra* note 20, at 2.
[23] *Id.*
[24] *Id.*
[25] *Id.*
[26] *Id.*
[27] USAF HANDBOOK, *supra* note 18, at 2.
[28] Leveson et al., *supra* note 20, at 2.
[29] ADEDEJI BODUNDE BADIRU, HANDBOOK OF INDUSTRIAL AND SYSTEMS ENGINEERING 9.1 (2005).
[30] Clifton A. Ericson, *A Short History of System Safety*, 42(6) JOURNAL OF SYSTEM SAFETY (2006), http://www.system-safety.org/ejss/past/novdec2006ejss/clifs.php.
[31] *Id.*

University of Southern California, a master's degree program was offered for the first time in Aerospace Operations Management, from which specific system safety graduate courses were developed. And, in 1965, the first official System Safety Conference was held in Seattle, Washington, signaling the full recognition and institutionalization of system safety.

The adoption of system safety meant that the "fly-fix-fly" method was replaced by the "identify-analyze-control" method.[32] By the end of the 1980s, the number of accidents was drastically reduced and stood at two per every hundred thousand flight-hours. However, the cost of a single plane rose to such an extent that the financial loss from these two accidents was higher than the financial loss from ten such accidents in the 1940s.

D. Classifying required levels of safety according to levels of danger

The need to prevent accidents before they occur has also been made clear with regard to medical devices. For example, the Dalkon Shield, an intrauterine contraceptive device, was introduced on the worldwide market in 1970. Touted as a safe and effective contraceptive, its use resulted in a high percentage of inadvertent pregnancies, serious infections, and, in some cases, death.[33] Because it was considered a medical device instead of a drug, it did not require FDA approval before it was put on the market.[34] In the early 1970s, several other medical devices, such as catheters, artificial heart valves, defibrillators, and pacemakers, attracted the attention of consumers, the FDA, and the US Congress as possible health risks.[35] Because the FDA had limited authority to prevent such risks from materializing, Congress enacted the Medical Device Amendments of 1976, which required that the FDA review medical devices before allowing them to be marketed.[36]

---

[32] USAF HANDBOOK, *supra* note 18, at 3.
[33] Medtronic, Inc. v. Lohr, 518 U.S. 470, 476 (1996).
[34] Carol T. Rieger, *The Judicial Councils Reform and Judicial Conduct and Disability Act: Will Judges Judge Judges?*, 37 EMORY L.J. 45, 62 (1988).
[35] *Medtronic, Inc.*, 518 U.S. at 476.
[36] Lana Steven, *Tenth Circuit Survey: Torts: Products Liability*, 75 DENV. U.L. REV. 1105, 1106 (1998).

US federal law and FDA regulations make a distinction between three classes of medical devices intended for human use, according to the level of danger posed to the public and the required degree of control: Class I, Class II, and Class III devices.[37]

Class I devices are subject to the least regulatory control since they present minimal potential harm to users and are generally simpler in design than Class II or Class III devices.[38] Elastic bandages, examination gloves, and hand-held surgical instruments are examples of Class I devices. These devices are only subject to general controls, which include: registration of companies such as manufacturers, distributors, repackagers and relabelers; medical device listing with the FDA of devices to be marketed; the manufacture of devices in accordance with good manufacturing practices (GMP); the labeling of devices in accordance with labeling regulations; and the submission of a "premarket notification" before a device is marketed.[39]

Devices categorized as Class II pose medium risks[40] and cannot be classified as Class I "because the general controls by themselves are insufficient to provide reasonable assurance of the safety and effectiveness of the device, and for which there is sufficient information to establish special controls to provide such assurance …".[41] Some examples of Class II devices are powered wheelchairs, infusion pumps, and surgical drapes. The special controls required of these devices may include special labeling requirements, mandatory performance standards, and postmarket surveillance.[42]

The strictest regulatory category is Class III.[43] For such devices there is insufficient information to guarantee safety and effectiveness through general and special controls alone. These devices are generally used to support or sustain human life and are of substantial importance in preventing harm to a person's health. A prosthetic heart valve is one example of a Class III device. Class III devices require premarket approval.

---

[37] *See*: 21 U.S.C. § 360c, *available at* http://law.justia.com/us/codes/title21/21usc360c.html; and the FDA explanation of device classes *at* http://www.fda.gov/cdrh/devadvice/3132.html (hereinafter: "FDA Device Classes"). *See also* John Y. Chai, *Medical Device Regulation in the United States and the European Union: A Comparative Study*, 55 FOOD DRUG L.J. 57, 57-59 (2000).
[38] *See* FDA Device Classes, *supra* note 37 ("Class I – General Controls").
[39] *Id.*
[40] Chai, *supra* note 37, at 58.
[41] 21 U.S.C. § 360c(a)(1)(B), *available at* link found in *supra* note 37.
[42] *See* FDA Device Classes, *supra* note 37 ("Class II – Special Controls").
[43] *See* FDA Device Classes, *supra* note 37 ("Class III – Premarket Approval").

Although Class III devices are subject to the strictest control system because they pose a significant danger to human life, even Class I devices must meet strict quality assurance requirements, such as good manufacturing practices.[44]

Not only are forensic devices (such as a breathalyzer or the software used to create a DNA profile, which could provide evidence leading to a murder conviction and a death sentence) not required to meet the standards of Class III medical devices (which pose a danger to human life, such as a prosthetic heart valve), but they are not even required to meet the requirements for Class I devices, such as elastic bandages.[45] In fact, we have found no statutory requirements whatsoever regarding the development of forensic equipment.[46] We propose that the pre-market approval applied to Class III devices also be required with regard to forensic equipment.

E. System safety engineering

System safety engineering is defined as "an engineering discipline requiring specialized professional knowledge and skills in applying scientific and engineering principles, criteria, and techniques to identify and eliminate hazards, or reduce the associated risk."[47] In other words, "[t]he objective of system safety is to achieve

---

[44] *See* Good Manufacturing Practices (GMP) / Quality System (QS) Regulation, at the FDA website: http://www.fda.gov/cdrh/devadvice/32.html. And *see also* Medical Devices / Current Good Manufacturing Practice (CGMP) / Final Rule, 61 Fed. Reg. 52601, 52606 (Oct. 7, 1996) (codified at 21 C.F.R. pts. 808, 812 and 820 (1996)) (hereinafter: "CGMP Final Rule"), *at* http://www.fda.gov/cdrh/fr1007ap.pdf.

[45] Indeed, it is possible that a particular product, which is not an automatic software-based product, will be granted exemptions from the established requirements (such as the exemption from design control requirements in 21 C.F.R. § 820.30); however, devices that produce forensic evidence do not even need an exemption, since the standards for safety-critical systems do not apply to them from the outset.

[46] No proposals for regulating DNA testing deal with the development of the device and software by the manufacturer; instead, they mostly concern instructions for collecting data in the field and for the tests performed by laboratories. *See*, for example: DNA Advisory Board, *Quality Assurance Standards for Convicted Offender DNA Databasing Laboratories,* 2(3) FORENSIC SCI. COMM. (2000), FBI website: http://www.fbi.gov/hq/lab/fsc/backissu/july2000/codis1a.htm; NATIONAL RESEARCH COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE 85 (1996); Technical Working Group on DNA Analysis Methods (TWGDAM) & California Ass'n of Criminalists Ad Hoc Comm on DNA, *Quality Assurance Guidelines for a Quality Assurance Program for DNA Analysis*, 18 CRIME LABORATORY DIG. 44, 52 (1991). We have also found no such requirements in the regulations of different states regarding DNA testing, nor in the American Bar Association's Standards on DNA Evidence, ABA website: http://www.abanet.org/crimjust/standards/dnaevidence.html. *See also* Giannelli, *supra* note 6, at 89-90.

[47] *See* USAF HANDBOOK, *supra* note 18, at vii.

acceptable mishap risk through a systematic approach of hazard analysis, risk assessment, and risk management."[48] This is an extremely broad field.

The primary concern of system safety is the identification, evaluation, elimination, and control of hazards throughout a system's lifetime. Hazards may result from the failure of components but they might also have other causes. A central responsibility of system safety engineers is to evaluate the interfaces between the components of a system – which include humans, hardware, and software, along with the environment – and to determine the effect of component interaction on potentially hazardous system states. This process is known as System Hazard Analysis.[49]

System safety activities begin at the earliest conceptual stages of a project and continue through design, production, testing, operational use, and disposal. One of the ways in which system safety is distinguished from other safety approaches is that its primary emphasis is on the early identification and classification of hazards. This allows action to be taken in order to eliminate or minimize such hazards prior to final design decisions. The main safety system activities (as defined by system safety standards such as MIL-STD-882) include: top-down system hazard analysis; documenting, tracking and resolving hazards; designing in order to eliminate or control hazards and minimize damage; maintaining safety information systems and documentation; and establishing reporting and information channels.[50]

These principles have even been established in regulations.[51] In the matter under discussion, it is important to emphasize that safety engineering is in no way limited just to "black box" testing. Safety engineering entails a series of techniques applied throughout the entire life cycle of a product, including its design, development, testing, manufacture, storage, delivery to the client, and maintenance. The FDA credo is as follows:

> FDA believes that because of the complexity of many components used in medical devices, their adequacy cannot always be assured through inspection and testing at the finished device manufacturer. This is

[48] DOD STANDARD PRACTICE, *supra* note 18, at 3.
[49] Leveson et al., *supra* note 20, at 3.
[50] *Id.* at 3.
[51] For example, *see* 21 C.F.R. 820 ("Quality System Regulation"), *available at* http://law.justia.com/us/cfr/title21/21-8.0.1.1.12.html.

> especially true of software and software-related components, such as microprocessors and microcircuits. Quality must be designed and built into components through the application of proper quality systems.[52]

Thus, regarding airplanes, it has been said that "[s]afety must be designed and built into airplanes, just as are performance, stability, and structural integrity. A safety group must be just as important a part of a manufacturer's organization as a stress, aerodynamics, or a weights group…"[53]

This is also the case regarding the development of software.[54] One of the FDA principles for software validation is that "testing alone cannot verify that software is complete and correct."[55] And, thus, the FDA has stated as follows:

> Software quality assurance needs to focus on preventing the introduction of defects into the software development process and not on trying to "test quality into" the software code after it is written. Software testing is very limited in its ability to surface all latent defects in software code. For example, the complexity of most software prevents it from being exhaustively tested. Software testing is a necessary activity. **However, in most cases software testing by itself is not sufficient to establish confidence that the software is fit for its intended use**.[56]

These principles are well accepted in the field of computer science. "Testing can only prove the existence of defects, not their absence…. If quality was not present

---

[52] CGMP Final Rule, supra note 44, at 52606.
[53] USAF HANDBOOK, *supra* note 18, at 3.
[54] *See*, for example: JOINT SERVICES SOFTWARE SAFETY COMMITTEE, SOFTWARE SYSTEM SAFETY HANDBOOK (1999), http://www.system-safety.org/Documents/Software_System_Safety_Handbook.pdf; NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, SOFTWARE SAFETY GUIDEBOOK (2004), http://www.hq.nasa.gov/office/codeq/doctree/871913.pdf. Regarding FDA requirements for software safety, see Paul A. Mathew, *The Next Wave: Federal Regulatory, Intellectual Property, and Tort Liability Considerations for Medical Device Software*, 2 J. MARSHALL REV. INTELL. PROP. L. 259, 264-75 (2003).
[55] *See* Mathew, *supra* note 54, at 271.
[56] CENTER FOR DEVICES AND RADIOLOGICAL HEALTH, GENERAL PRINCIPLES OF SOFTWARE VALIDATION; FINAL GUIDANCE FOR INDUSTRY AND FDA STAFF 11 (2002), http://www.fda.gov/cdrh/comp/guidance/938.pdf.

in the requirements analysis, design, or implementation, testing cannot put it there."[57] The way to achieve software quality is through proper software design and the implementation of suitable processes, not by "assuring"' and testing the product. There is a necessary relationship between product quality and process.[58]

An example of the differences between safe and unsafe software may be found in the concept of "cyclomatic complexity."[59] Cyclomatic complexity refers to a certain measurement of the number of possible logic paths in running the software. Software that is not developed in accordance with standards of quality assurance could be characterized by high cyclomatic complexity. This leads to an inability to test all paths and a high likelihood of software error.[60] The Federal Aviation Administration (FAA) has made it clear that, "[a]s a goal, software complexity should be minimized to reduce likelihood of errors. Complex software also is more likely to be unstable, or suffer from unpredictable behavior. Modularity is a useful technique to reduce complexity."[61]

Finally, with regard to medical devices, reporting requirements exist that are designed to reveal and rectify defects in a product also after it has been approved for sale.[62]

III. A test case: the current attitude of the law towards forensic evidence obtained from breathalyzers

Breathalyzers are used in the United States and throughout the world to prove drunken driving.[63] The very fact that it is deemed sufficient to measure the level of alcohol in

[57] Norman Hines, *The Problem with Testing*, CROSSTALK – THE JOURNAL OF DEFENSE SOFTWARE ENGINEERING 27, 27-28 (July 2001), *available at* http://www.stsc.hill.af.mil/crosstalk/2001/07/hines.html.
[58] RON S. KENETT & EMANUEL R. BAKER, SOFTWARE PROCESS QUALITY: MANAGEMENT AND CONTROL 1.3 (1999).
[59] Thomas J. Mccabe, *A Complexity Measure*, IEEE. TRANS. SOFTWARE ENG., Vol. SE-2, No. 4. 308 (1976).
[60] FAA HANDBOOK, *supra* note 18, Appendix J (Software Safety), at J-17.
[61] *Id.* at J-23.
[62] FDA information regarding Medical Device Reporting (MDR), *at* http://www.fda.gov/cdrh/devadvice/351.html.
[63] E. John Wherry, Jr., *The Rush to Convict DWI Offenders: The Unintended Constitutional Consequences*, 19 U. DAYTON L. REV. 429, 434 (1994); Boaz Sangero & Mordechai Halpert, *The Danger of a Conviction Based on a Breathalyzer Test* (forthcoming, HA'PRAKLIT) (in Hebrew).

an expired breath, instead of measuring the concentration of alcohol in the blood, already means that a concession is being made regarding the accuracy of the findings.[64]

Moreover, there are numerous possibilities for error during the course of the test, such as: measurement during the absorptive state;[65] the "mouth alcohol" phenomenon;[66] the presence of other materials, like acetone, which could be identified by the device as the ethanol relevant to drunkenness;[67] etc.[68]

National Highway Traffic Safety Administration (NHTSA) regulations, which deal with the regulation and approval of breathalyzers, do not include requirements for the development process itself. For example: they establish model specifications;[69] they regulate the tests to be performed on breathalyzers in order to check their operation and accuracy;[70] and they establish model specifications for calibrating units, in particular.[71] However, they do not contain any instructions whatsoever regarding the manner in which the device is to be developed. The regulations only require that the final product pass a series of tests – that is to say, the device is tested as a "black box."[72] There is no requirement for the manufacturer to meet any quality assurance standards whatsoever. There is no requirement regarding the manner in which the device's software is to be developed. There is no reporting requirement and there are no statutory sanctions against manufacturers, as is the case for medical devices.

---

[64] Michael P. Hlastala, *Physiological Errors Associated with Alcohol Breath Testing*, 9(6) THE CHAMPION 16, 19 (1985); State. v. Mcginley, 229 N.J. Super. 191, 550 A. 2d 1305, 1310 (New Jersey Law Division 1988.

[65] Gerald Simpson, *Accuracy and Precision of Breath Alcohol Measurements for Subjects in the Absorptive State*, 33(6) CLIN. CHEM. 753, 753 (1987). The absorptive stage of alcohol consumption is the stage during which the concentration of alcohol in the blood rises until it reaches a peak. This stage occurs from the start of drinking up until a given period of time following its conclusion, which varies from person to person. The average blood-breath alcohol ratio during the absorptive stage is lower than the average ratio during the stage following absorption (2100). In the absorptive stage, its value could be as low as 630. Therefore, during the absorptive stage, a high concentration of alcohol in a breath does not necessarily prove that the concentration of alcohol in the blood is above the permitted level.

[66] Michael P. Hlastala & Wayne J.E. Lamm, *The Slope Detector Does Not Always Detect the Presence of Mouth Alcohol*, 30 THE CHAMPION 57 (2006). The presence of alcohol in the mouth would cause a breathalyzer to check the concentration of alcohol in the mouth instead of the concentration of alcohol in the lungs. Alcohol in the mouth could be present from sipping a permitted amount of alcohol or from the use of breath fresheners and even the consumption of certain foods.

[67] Leonard Stamm, *The Top 20 Myths of Breath, Blood, and Urine Tests – Part 2*, 29 THE CHAMPION 44, 45-46 (2005).

[68] *Id. See also*: Leonard Stamm, *The Top 20 Myths of Breath, Blood, and Urine Tests – Part 1*, 29 THE CHAMPION 20 (2005); Sangero & Halpert, *supra* note 63.

[69] 58 Fed. Reg. 48705; 72 Fed. Reg. 34742.

[70] 72 Fed. Reg. 34742 § 4.

[71] 58 Fed. Reg. 48705, § 4; 72 Fed. Reg. 34742, § 3.

[72] 58 Fed. Reg. 48705, § 4; 72 Fed. Reg. 34742, § 3.

A recent case in the New Jersey Supreme Court dealt with the reliability of the Draeger Alcotest 7110 MKIII-C.[73] In a special proceeding conducted by retired Judge Michael P. King serving as a Special Master, thirteen experts were heard on the subject, on behalf of both the State of New Jersey as the plaintiff-appellant (experts who included the manufacturer's engineers) and the defendants (including experts who had written major articles on this subject).[74] Some of the state's witnesses expressed great confidence in the device and in the "black box" testing method performed on it.[75]

Judge King found this testing method to be scientifically reliable.[76] Despite the fact that one of the state's witnesses expressed confidence in the Draeger breathalyzer and its software, he also stated that his confidence would be higher if an independent body had tested the software.[77] Moreover, he stated that the best way to assure the quality of the software is by adopting standardized processes at the time of its development.[78] And, indeed, in an uncommon step, the defense was given a limited opportunity to examine the device's software.[79] Thus, it was decided that the manufacturer of the device would turn the software over to an independent software house, so that it could be tested for the defense. The manufacturer, Draeger, also sent the software to be tested by an independent software house of its own.[80]

The defense's software house found many defects in the device's software. The clearest finding was that the software would not meet the standards of the US government, the US military, the FAA, or the FDA, nor commercial standards designed to ensure public safety.[81] This finding was not refuted, and was even

---

[73] *See* State v. Chun, 191 N.J. 308, 923 A.2d 226, 2007 N.J. LEXIS 579 (2007).
[74] State v. Chun, 2007 N.J. LEXIS 39.
[75] *Id.* at *226-27, *235.
[76] *Id.* at *278.
[77] *Id.* at *233.
[78] *Id.* at *232, *235 (stating that he "would want to know if the software was certified in conformance with ISO 9000 and 9003").
[79] The order for analysis of the software was issued after publication of the aforesaid conclusions in State v. Chun, 2007 N.J. LEXIS 39. See this order in State v. Chun, 191 N.J. 308, 923 A.2d 226, 2007 N.J. LEXIS 579 (2007).
[80] *See*: the Base One Technologies report, on behalf of the defendants, *at* http://www.dwi-nj.com/pdf/AlcoTestReportBaseOne82707v2.pdf; the SysTests Labs report, on behalf of the manufacturer, *at* http://www.dwi-nj.com/pdf/DSDIStaticCodeReviewAnalysisFindings.pdf.
[81] See the Base One Technologies report, *supra* note 80, at 3; and *see id.* at 12 ("The program presented shows ample evidence of incomplete design, incomplete verification of design, and incomplete 'white box' and 'black box' testing. Therefore the software has to be considered unreliable and untested, and in several cases it does not meet stated requirements.").

supported, by the report of the software house that tested the device on behalf of the manufacturer.[82]

However, in Judge King's Supplemental Findings and Conclusions of Remand Court, which followed the first report and the software testing, no significance was found in the fact that the software was not written in compliance with any standards.[83] In this matter, Judge King based his conclusion on the testimony of the state's witnesses, which he summarized as follows:[84]

> Geller: He did not know if there were any industry standards which governed source code review. In Geller's opinion, quality software could be developed without standards and conversely, software could meet standards but still be of questionable quality. He was not personally familiar with any of the standards cited in Base One's report nor did he know if Draeger applied any standards to the Alcotest source code.
>
> Dee: He did not agree with Base One's criticism of the Alcotest's lack of standards. In his opinion, such standards usually referred to the design and documentation of the code, and rarely addressed the code's performance. He was unaware of any standard against which the United States evaluated software. Moreover, he objected to Base One's reference to standards without stating which specific provisions were violated. He said it was possible to fully test the source code given the singular or specialized function of this application.

---

[82] See the SysTest Labs report, on behalf of Draeger, *supra* note 80, at 3 ("It was recognized that the examined source code is highly complex. Industry best practices dictate that cyclomatic complexity, a standard measure of source code complexity indicative of both understandability and maintainability; have a value no greater than 10. Upon review the NJ v.3.11 source code was found to include 81 modules with cyclomatic complexity indices in excess of 10 and three modules with indices in excess of 100."). The software house even found a significant error, *id.* at 3-4 ("During tests in which the subject is required to provide three sufficient breath samples, and the third sample is within tolerance with each of the other two samples, and the lowest of the six recorded result values is the second breath sample's electrochemical test result, the 'Reported Breath Test Result' is invalidated by a buffer overflow error, although the measured alcohol concentration values are correctly retained and reported in the Alcohol Influence Report.").

[83] State v. Chun, Supplemental Findings and Conclusions of Remand Court (Nov. 14, 2007), *at* http://www.judiciary.state.nj.us/pressrel/supplemental_opinion.pdf (hereinafter: "*Chun - Supplemental Findings*").

[84] *Id.* at 90-91. This summary should enable the reader to better understand the "black box" approach, which, unfortunately, still prevails in the field of forensic evidence.

Shaffer: He was unaware of any single industry standard for software development. He referred to "industry standards" as collections of techniques and common-sense wisdom which had proven effective over time. Shaffer did not agree with Base One's assertion that the failure to use industry coding standards prevented the testing of critical paths in the Alcotest's software including 3200 lines of code designed to make decisions.

Because the Alcotest in the United States was highly configured to meet the requirements of specific applications, all of the 3200 lines of decision code – as calculated by Base One – were not relevant. Shaffer also said there were many unused or uncalled modules or sections of code by design. Shaffer did say standard style would be helpful but was not necessary.

Based on the testimony of these three witnesses, which he found persuasive, Judge King made no recommendation regarding the need for standards.[85] He then expressed a similar view regarding the matter of "cyclomatic complexity," also based on the testimony of the state's expert witnesses, summarized as follows:

Geller: He relied on the cyclomatic complexity metric developed by Thomas McCabe in 1976 to measure the number of potential paths through the code. Because high complexity increases the risk of inherent defects, coding guidelines recommend keeping the cyclomatic complexity of functions under ten, and or even seven. The SysTest report identified more than eighty-one modules in excess of ten and three in excess of a hundred. While the report recommended restructuring the code to make it less complex, Geller said the complexity indices did not influence the instrument's accuracy. Nor did excessive complexity cause failures in the interfaces between software and hardware. However, the higher complexity made the code more difficult to understand and maintain,

[85] *Id.* at 91.

placing an increased burden on the programmers who worked with the software.[86]

Recommendation: None, because this goes to style and not the accuracy of the Alcotest. We accept Geller's and Dee's testimony as persuasive that the Alcotest performs accurately at this level of complexity.[87]

The New Jersey Supreme Court accepted Judge King's conclusions that the device was reliable, subject to compliance with certain users instructions.[88] The Court saw no need to require that the software comply with any specific standards.[89] It accepted the opinion of the state's experts that the criticism of the device's software by defense experts was solely related to matters of style and theory.

Interestingly, Draeger had appointed no one to be responsible for assuring the quality of the software, as required in the development of safety-critical systems.[90] Shaffer, the programmer, testified that he was unaware of a single industry standard and was unfamiliar with the ISO 9000 standards for software and that, instead, had adopted "common sense" standards learned from his own experience.[91]

This mode of software development is risky. And, indeed, it was made clear that, when implementing software changes requested by the State of New Jersey, Shaffer himself had created a buffer overflow defect.[92]

This error only occurred under very limited circumstances where the first two breath tests were out of tolerance, the subject provided a third breath sample within tolerance of each of the other two samples, and the lowest of the six recorded test

---

[86] *Id.* at 91-92.
[87] *Id.* at 92.
[88] State v. Chun, 194 N.J. 54, 943 A2d. 114, 2008 N.J. LEXIS 133, ***167-77.
[89] *Id.* at ***135-36:

> Our evaluation of the exhaustive record relating to the source code leaves us confident that its errors have been revealed. Based on that record, we do not share defendants' larger concerns that it is likely to generate inaccurate results simply because, from a source code writer's viewpoint, it is complex or prolix. There being no evidence in the record that these asserted shortcomings are anything more than stylistic, theoretical challenges, we decline defendants' invitation to require that the firmware comply with any specific programming standards as unnecessary at this time.

[90] *Chun - Supplemental Findings, supra* note 83, at 65.
[91] *Id.* at 65-66.
[92] *Id.* at 73.

results was the second breath sample's electrochemical (EC) test result.[93] This defect was not discovered in field tests since the circumstances under which it occurs are not very common.[94] The software error was also not detected by the defense's software house; instead, it was detected by Draeger's independent software house.

This scenario is consistent with all of the concerns that we have described regarding an unsafe manner of development, and it is certainly possible that other software errors remained undetected. If this scenario was the subject of a tort claim, it is possible that Draeger would be found negligent because it had not shown sufficient care in developing its products.[95] However, software that has been developed negligently could, under certain circumstances – particularly when it has the power to seal a person's fate in a criminal trial – be deemed to yield admissible evidence carrying significant weight, even though it might cause errors leading to wrongful convictions.

In the next section we argue that devices used to produce forensic evidence should also be regarded as safety-critical systems. If we are correct, then Judge King's conclusions about the breathalyzer, as well as his willingness to treat the device as a "black box," is mistaken. So too is the New Jersey Supreme Court's failure to require that Draeger comply with some type of standards in coding the software.[96] Therefore, it is necessary to carefully examine the manner in which the device is developed.

IV. From a plane crash to the conviction of an innocent person – why is it necessary to develop forensic equipment as a safety-critical system?

A. Classifying forensic equipment as a safety-critical system

Daryl Mack was convicted of murder based on a single piece of DNA evidence obtained by running a sample through a database.[97] He was sentenced to death and executed on April 26, 2006.[98] It is possible, even if the likelihood is low, that

[93] *Id.* at 10.
[94] *Id.* at 30, 32, 85-86.
[95] *See* Mathew, *supra* note 54, at 290-97 (discussing negligence law, medical device software, and the law of product liability).
[96] *See* State v. Chun, 2007 N.J. LEXIS 39.
[97] Mack v. State, 75 P.3d 803 (Nev. 2003).
[98] *See* http://www.clarkprosecutor.org/html/death/US/mack1019.htm.

computer software relied upon for comparing DNA will also produce erroneous findings, just as it is possible that a testing error has occurred in the laboratory.[99] This is only one example of how forensic evidence could possibly endanger the life of an innocent person. Therefore, a device used to produce forensic evidence is a safety-critical system in the most fundamental sense, in that it endangers human life. Research clearly shows that erroneous scientific evidence does occasionally lead to the conviction of innocent persons.[100] In our opinion, this is no less of a catastrophe than the failure of a car's brakes. In any event, we have found no mandatory regulation – for any forensic evidence whatsoever – regarding the development of software or devices in accordance with safety-critical standards, although such regulations are so well accepted in other matters of life-and-death.

Why should strict methods for control, development, and testing – deemed so essential for protecting public safety when it comes to safety-critical systems – not be considered equally necessary in the realm of forensic science in order to ensure that we do not convict an innocent person based on erroneous scientific evidence. Why should manufacturers of equipment peculiar to crime laboratories be the only manufacturers not required by the law to meet any standards of quality assurance.

B. Why "black box" testing is not enough

As long as we believe that a device is functioning properly, it might seem sufficient to test its performance as a "black box," without bothering to try to understand its design or development. Why is it so important for us to carefully test the device, to verify that it has been developed in accordance with strict standards of safety, and to ensure that its software will not fail, in the future, under given circumstances?

Indeed, testing is also very important in safety engineering.[101] However, it is not enough that the authorities just examine the forensic end product. They must require manufacturers to develop their products in accordance with rules of safety engineering and they must monitor this development in order to ensure that this has in fact been done.

---

[99] On the possibility of error regarding DNA evidence, see Sangero & Halpert, *supra* note 3, at 73-76.
[100] SCHECK ET AL., *supra* note 5, at 158-71; William C. Thompson et al., *How the Probability of a False Positive Affects the Value of DNA Evidence*, 48 J. FORENSIC SCI. 47, 47-48 (2003).
[101] USAF HANDBOOK, *supra* note 18 at 136-40.

First, testing cannot give an indication of the device's performance under all possible field conditions, and no number of tests would be sufficient.[102] Would someone be willing to fly in an airplane that was not developed in accordance with safety-critical standards and has not passed accepted tests for quality assurance – including its software – essentially taking a chance based solely on the fact that its prototype has flown 1,000 times without crashing? Perhaps it will be on flight number 1,001 that a failure will occur that could have been discovered in advance if only it had been developed according to proper rules of quality assurance and the accepted standards in this field?

The flaw in this approach also emerges from the 1986 Space Shuttle Challenger accident.[103] Prof. Richard Feynman, Nobel prizewinner in physics and a member of the presidential commission investigating the disaster, in describing the attitude of senior NASA officials, has said:

> The argument that the same risk was flown before without failure is often accepted as an argument for the safety of accepting it again. Because of this, obvious weaknesses are accepted again and again, sometimes without a sufficiently serious attempt to remedy them, or to delay a flight because of their continued presence.
>
> …
>
> The fact that this danger did not lead to a catastrophe before is no guarantee that it will not the next time, unless it is completely understood. When playing Russian roulette the fact that the first shot got off safely is little comfort for the next.[104]

[102] Regarding software, see Cem Kaner, *The Impossibility of Complete Testing*, 4(4) SOFTWARE QA MAGAZINE 28 (1997), http://www.kaner.com/pdfs/imposs.pdf. There are too many inputs to a program and too many combinations of inputs to a program in order to test them all. There are too many outputs to a program. Therefore, it is impossible to test all of the logic paths through a program – they are too numerous. It is impossible to test all of the other risks of software failure, such as user interface design errors or incomplete requirements analyses. Kaner gives an example of a program that was estimated to require 10^{100} tests in order to check all possible situations. This is an astronomical figure and it is clearly impossible to conduct so many tests.

[103] *See* REPORT OF THE PRESIDENTIAL COMMISSION ON THE SPACE SHUTTLE CHALLENGER ACCIDENT, Wash. D. C (1986). http://science.ksc.nasa.gov/shuttle/missions/51-l/docs/rogers-commission/table-of-contents.html.

[104] Feynman, *supra* note 1.

Second, as it is known in the field of safety engineering, testing the final product alone leads to flaws being corrected only after they have already occurred.[105] In the 1940s, fighter planes logged an average of ten thousand flight hours before the occurrence of a crash.[106] According to the approach of the New Jersey Supreme Court, perhaps each one of these ten thousand flight hours is to be deemed a test that proves the reliability of the plane, since, if the plane had a problem, it should have crashed?[107] However, despite the immense number of (cumulative) flight hours without an accident, it became clear that the planes still had defects. These defects were repaired, but only after a crash had already occurred.[108] Moreover, in the world of criminal law, the danger is that even a "fly-fix-fly" method such as this would not be implemented. When the manufacturer of equipment used to produce forensic evidence, such as a breathalyzer, markets a device which, because of a design, software, or other type of defect, occasionally yields erroneous results, the chances that this will be discovered by a court of law are small. Whereas a car with a safety flaw will be involved, sooner or later, in an accident that exposes the defect, this is not the case with forensic equipment. If an innocent person is convicted based on erroneous scientific evidence – and the law does allow for a conviction based on a single piece of evidence[109] – the defendant's claims of innocence will not be considered a refutation of the reliability of the forensic device, and the chances of proving that there was a testing error are slim. Moreover, many people perceive a conviction itself as further confirmation of their faith in the device's reliability. Therefore, with forensic equipment – even more than with devices in other fields, where defects are likely to be discovered – if errors are to be prevented (because it is unlikely they will later be detected) great caution must be taken in the design and manufacturing stages in order to minimize the possibility of error. In effect, this is the only stage when it is reasonably possible to discover and to avoid potential defects.

---

[105] See the discussion of the "fly-fix-fly" method, at text accompanying *supra* notes 21 through 32.

[106] USAF HANDBOOK, *supra* note 18 at 2.

[107] State v. Chun, 194 N.J. 54, 943 A2d. 114, 2008 N.J. LEXIS 133, ***121-22:

> First, in "black box" testing, the machine performed accurately by demonstrating the ability to identify the percentage of alcohol in known solutions within the applicable tolerance parameters. Were there a fundamental defect in the source code, one would expect that the machine would not be able to perform in this fashion.

[108] As we have said, this method was abandoned. The goal of safety engineering is to prevent the mishap *prior* to its occurrence.

[109] Sangero & Halpert, *supra* note 3 at 62; See the case of Daryl Mack, *supra* notes 97 and 98, and accompanying text.

An excellent example of what we are talking about is the legal proceeding conducted in New Jersey concerning the breathalyzer. The State of New Jersey, the plaintiff-appellant in the case, argued that its tests of the device as a "black box" showed it to be reliable (i.e., "fly").[110] Only within the framework of the legal proceeding, when – in an uncommon step – the device's software was tested, did the manufacturer (Draeger) discover a "buffer overflow" error that occurs under very specific circumstances.[111] In retrospect, it became clear that the particular situation in which the error occurs (when a third breath test was conducted) simply did not arise in the field and was therefore not tested and discovered.[112] The judgment rendered by the New Jersey Supreme Court established specific repair instructions (i.e., "fix").[113] Since it was not proven that there were additional flaws in the device, the Court rejected the complaints by defense experts that the software cannot be tested and that, consequently, there might be additional flaws in the device.[114] The device, subject to certain instructions, was approved for use (i.e., "fly" until another flaw is discovered).[115] However, as we have said, there is a much lower probability that this flaw would be discovered in the legal realm than in other areas of our daily lives. This is a clear example of why it is wrong to suffice with a test of the final product, since there are many possible combinations and it is impractical to test them all; and it demonstrates that the courts apply a "fly-fix-fly" method when dealing with forensic evidence, a method that was abandoned sixty years ago in other fields. In our opinion, efforts must already be made during the development of the device to achieve "first-time safety."[116]

Third, testing alone will not make the device more reliable. It is not enough to learn through testing that the device is liable to yield erroneous results, but rather there is a need to take steps in order to make it more reliable. In the history of forensic evidence, we have seen many cases where scientific evidence has been presented in

[110] State v. Chun, 2007 N.J. LEXIS 39.
[111] *Chun - Supplemental Findings*, *supra* note 83, at 73.
[112] *Id.* at 30, 32, 85-86.
[113] State v. Chun, 194 N.J. 54, 943 A2d. 114, 2008 N.J. LEXIS 133, ***125-32 and the appendices to the judgment.
[114] *Id.* at 135-36.
[115] *Id.* at 168-77.
[116] BADIRU, *supra* note 29, at 9.2.

court and prosecution experts claim it to be infallible.[117] In retrospect, with the revelations of numerous wrongful convictions, it appears that this was not the case.[118]

Even laboratory proficiency testing proves that there are errors in all types of forensic evidence. The first laboratory proficiency tests were conducted in 1978.[119] Such tests are also performed in fields unrelated to forensic evidence, such as in the oversight of medical laboratories.[120] Proficiency testing in certain fields, such as DNA evidence, has been anchored in legislation.[121] There is a proposal to conduct "blind" proficiency tests (conducted without the knowledge of those being tested) for all forensic laboratories.[122] *Such tests have revealed numerous errors in forensic evidence, even with evidence that has been considered reliable for over a hundred years (such as fingerprint evidence).*[123] Therefore, we are very far from a situation in which testing the final product as a "black box" reveals that there are no errors in

[117] Thus, for instance, at the inception of DNA evidence, experts testified in court that a false positive was impossible. This was proven to be incorrect, *See* Thompson et al., *supra* note 6, at 47-48. Even with regard to fingerprint evidence, for a hundred years it was argued that errors were not possible, but this too was proven to be incorrect. *See* Sangero & Halpert, *supra* note 3 at 63-71. In one case, a prosecution expert testified that the GC/MS drug test is like a "Mercedes" and the "gold standard" of drug tests. However, the defense pointed out unexplained errors in the tests of experts from the same laboratory. *See* United States v. Richard J. Israel Jr., 60 M.J. 485, 486 (C.A.A.F. 2005), www.armfor.uscourts.gov/opinions/2005Term/04-0217.htm. Moreover, many defendants have been convicted on the basis of some odd "tests" – such as a microscopic examination of hair. Today this is not considered science at all but rather "junk science." *See* SCHECK ET AL., *supra* note 5, at 158-71 (Chapter 7 – "Junk Science").

[118] Regarding wrongful convictions and false positives, in actual cases, as the result of an erroneous DNA test, see Sangero & Halpert, *supra* note 3 at 74-75. Regarding 22 cases where, during legal proceedings, mistakes were discovered in fingerprint evidence, see Simon A. Cole, *More than Zero: Accounting for Error in Latent Print Identification*, 95 J. CRIM. L. & CRIMINOLOGY 985, 1001-16 (2005). Mistakes in forensic evidence are the second most significant factor leading to wrongful convictions (following only mistaken eyewitness testimony). *See* Brandon L. Garrett, *Judging Innocence*, 108 COLUM. L. REV. 55, 76 (2008).

[119] Giannelli, *supra* note 6, at 72.

[120] *See, e.g.*, 42 C.F.R. § 493.1236.

[121] Giannelli, *supra* note 6, at 82-84. And see, for example, regarding drug testing in the workplace in Florida, FLA. STAT. ANN. § 112.0455(13)(b) (5).

[122] Lauri Constantine et al., *Model Prevention and Remedy of Erroneous Convictions Act*, 33 ARIZ. ST. L.J. 665, 676, 699-701 (2001).

[123] Saks & Koehler, *supra* note 6, at 895. Regarding the results of proficiency tests for DNA laboratories, see Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence*, 34 JURIMETRICS J. 21, 26 (1993). *See also* Koehler et al., *supra* note 6, at 206-11. Regarding drug tests, see the results of the proficiency tests conducted for laboratories in Italy, which performs sophisticated reliability tests, such as GC/MS. The best result was obtained with regard to methadone – 0.2% false positive. The worse result was obtained with regard to opiates – 5.0(!) false positive. The highest percentage of false negatives reached 30.7%. Santo D. Ferrara et al., *Proficiency Testing for Psychoactive Substances In Italy*, 113 INT'L J. LEGAL MED. 50, 52 (1999).

forensic evidence.[124] We know today that errors occur in all types of forensic evidence. And, despite the discovery of these errors, the courts continue to rely on scientific evidence, even as the sole evidence for a conviction,[125] while (apparently) treating the error as a matter of fate.[126] Therefore, it is not enough to know that there are errors in the testing, but rather there is a need to make sure that the evidence will be more reliable. Although we strongly support laboratory proficiency testing, and even its improvement, we believe that this is not enough – a way must be found to make forensic evidence more reliable. Safety engineering could not only help to improve the devices, but also to prevent human error.[127]

Furthermore, the discovery of flaws in the tests themselves does not necessarily mean that we will always be able to locate their source. And it does not mean that we will know how to correct them or how to identify them in the legal process. Testing flaws are liable to be classified – in general terms – as false positives, while the exact source of the error is not discovered. For example, a software failure is liable to cause the breath alcohol concentration to appear to be higher than the permitted level, whereas the blood alcohol concentration is actually below the permitted level. This error will be classified as being related to the blood-breath ratio, whereas, if the true source of the error is a software problem, the reason for the flaw will not be discovered.[128] Thus, in the case of DNA evidence, even when it has been

[124] Also regarding the tests for drunken driving discussed in the New Jersey case, from the studies where errors had occurred it became clear that this was possible in few cases as a result of the blood-breath ratio. *See* State v. Chun, 2007 N.J. LEXIS 39, *166.

[125] *See* Sangero & Halpert, *supra* note 3.

[126] Regarding DNA evidence, see Armstead v. State, 673 A.2d 221, 245 (Md. 1996). The majority opinion in this case approved the presentation of the evidence with the reasoning that the jury in this case had the information about the errors discovered in the proficiency tests for the laboratory in question and that the defense had the ability to raise this same question on cross examination. Also in the matter of tests for drunken driving, the New Jersey Supreme Court was aware of the fact that there would be individuals for whom the blood-breath ratio would be less than a value of 2100: State v. Chun, 194 N.J. 54, 943 A2d. 114, 2008 N.J. LEXIS 133. Nonetheless, it saw fit to allow the evidence (*id.* at 72-71, and *see infra* note 128).

[127] DICK SAWYER, DO IT BY DESIGN - AN INTRODUCTION TO HUMAN FACTORS IN MEDICAL DEVICES (FDA, Dec. 1996), http://www.fda.gov/cdrh/humfac/doitpdf.pdf ("The purpose of this primer is to encourage manufacturers to improve the safety of medical devices and equipment by reducing the likelihood of user error. This can be accomplished by the systematic and careful design of the user interface, i.e., the hardware and software features that define the interaction between users and equipment").

[128] In the New Jersey case (State v. Chun, 2007 N.J. LEXIS, 39, *166), the state presented the findings of a study that had been conducted in New Zealand: A. R. Gainsford, et al., *A Large-Scale Study of the Relationship Between Blood and Breath Alcohol Concentrations in New Zealand Drinking Drivers*, 51 J. FORENSIC SCI. 173 (2006). This study examined the blood-breath ratio of persons who had failed a breath test and also requested a blood test. The results indeed show that the blood-breath ratio of most of those tested was above the accepted threshold of 2100. However, there were also those who were

discovered, in retrospect, that a testing error did actually occur, the exact source of the error was not identified.[129]

C. The expected benefit to society from the development of forensic equipment as a safety-critical system; and a cost-affective argument

A device that is developed in accordance with safety-critical standards may be expected to yield more precise and more reliable scientific evidence. This means, first of all, that there will be *fewer false positives*, i.e., cases where a suspect or defendant has not committed the crime but the test erroneously indicates that he was involved. Secondly, an improved device will also yield *fewer false negatives*, i.e., cases where a person has committed the ascribed offense but the test erroneously indicates that he was not involved. This would lead to better and more efficient enforcement of the criminal law.

One possible question relates to the cost of the required safety program and whether or not it is economically worthwhile. Financial considerations are also taken into account regarding aircraft safety.[130] For example, the general cost of the safety program in the manufacture of the F-14 fighter plane was approximately five million dollars spread out over ten years of development.[131] This was about a third of the cost

---

measured to have a ratio of 688, i.e., less than a third of the threshold (*id.* at 176). Such persons, whose level of alcohol is a third of that permitted, are liable to fail a breath test. How is it possible to know from the tests that the source of the error is indeed a very low blood-breath ratio and not a software defect or some other flaw in the device? That is to say, testing does not necessarily reveal the exact source of the error. Furthermore, one of the state's witnesses, Gullberg, described the findings of another study in which 5 or 6 out of 793 false positives were discovered (State v. Chun, 2007 N.J. LEXIS, 39, *265), however, the source of the error was unknown. Such errors are automatically classified as a blood-breath ratio problem. And, once again, it should be asked, how do we know that the source of error is not a defect in the device itself?

[129] A commission of inquiry established in New Zealand to investigate the erroneous match between the DNA of a person unrelated to the crime and that from a sample collected at the crime scene, failed to find the exact source of error but was satisfied to determine that it was some sort of contamination that had occurred in the initial stages of the criminal investigation. *See* Sangero & Halpert, *supra* note 3, at 75 and n178.

[130] USAF HANDBOOK, *supra* note 18, at 2.

[131] Safety is also possible to achieve efficiently. *Id.*: "A really large program (e.g., B-1B) might have 30-40 government and contractor people involved at a peak period. Most programs need only one or two system safety personnel in the government program office and four or five at the peak of the contractor's effort. One person can monitor several subsystem programs simultaneously. Clearly, the saving of just one aircraft by a system safety program pays for that program many times over."

of a single plane. Therefore, it was enough to prevent the crash of one plane in order to make it very worthwhile from an economic perspective.[132]

In the matter under discussion, apart from the enormous suffering caused by the wrongful conviction of an innocent person, it is possible to try and assess the direct financial loss to society at large. The cost of putting the wrong person on trial is tremendous.[133] Furthermore, the extent of financial harm caused to a person wrongfully convicted and incarcerated can be deduced from the damages awarded in civil suits in such cases.[134] James Newsome was awarded fifteen million dollars for having served fifteen years imprisonment on a wrongful conviction.[135] Eric R. Sarsfield was awarded $13,655,940 for ten years served on a wrongful conviction.[136] Therefore, the economic harm caused by a wrongful conviction and incarceration could reach many millions for a single case.[137] Furthermore, even if an amount equivalent to that invested in the safety system of the F-14 fighter plane would be invested in the safety, development, and manufacture of a forensic device, then this would be economically worthwhile if, in doing so, the wrongful conviction and prolonged incarceration of one person would be prevented. In one case, termination from employment as a result of a false positive in a drug test led to an award of about four hundred thousand dollars – without even a single day of incarceration.[138] This

---

[132] *Id.*

[133] An assessment of the cost of trying capital cases, conducted in Kansas, put this at over a million dollars per case. In one of the cases, it even exceeded two million dollars. *See* KANSAS LEGISLATIVE DIVISION OF POST AUDIT, PERFORMANCE AUDIT REPORT: COSTS INCURRED FOR DEATH PENALTY CASES – A K-GOAL AUDIT OF THE DEPARTMENT OF CORRECTIONS 10-12 (Dec. 2003), http://www.kslegislature.org/postaudit/audits_perform/04pa03a.pdf. See also the calculations of trial costs in David Wippman, *Notes and Comment: The Costs of International Justice*, 100 AMER. J. INT'L L. 861, 863-64 n16 (2006). According to media reports, some trials cost much more. For example, the O.J. Simpson trial was reported to have cost nine million dollars: http://findarticles.com/p/articles/mi_m1355/is_n7_v89/ai_17968684.

[134] Unfortunately, many of those who have been wrongfully convicted are not awarded full compensation for their damages because they have not proven sufficient liability on the part of the authorities. *See* Garrett, *supra* note 118, at 176. In the present article, we are not focusing on entitlement to damages, but rather on the assessment of its extent.

[135] Newsome v. McCabe, 319 F.3d 301, 303 (2003).

[136] Sarsfield v. City of Marlborough, 2007 U.S. Dist. LEXIS 5445. Regarding this case, see Elaine Thompson, *Ex-inmate awarded $13.6M; Innocent man was jailed for rape,* TELEGRAM & GAZETTE (MA), Oct. 7, 2006, at the National Association of Criminal Defense Lawyers website: http://www.nacdl.org/public.nsf/defenseupdates/innocence012.

[137] For additional cases where millions have been awarded, see Brandon L. Garrett, *Innocence, Harmless Error, and Federal Wrongful Conviction Law*, 2005 WIS. L. REV. 35, 43-44 (2005). In other countries as well, high sums have been awarded. For example, David Milgaard, from Canada, was awarded ten million dollars in 1999 for 23 years of wrongful imprisonment: *David Milgaard: Timeline*, CBS NEWS ONLINE (updated Dec. 2006), *at* http://www.cbc.ca/news/background/milgaard/.

[138] Ishikawa v. Delta Airlines 343 F.3d 1,129, 1,131 (9th Cir. 2003).

also indicates the magnitude of economic harm caused to people who are wrongfully convicted on the basis of an erroneous breath test. Even if a person is not sent to jail, this could lead to the loss of an occupation, like that of someone who makes a living as a driver.

Beyond a comparison between the cost of safety and the economic harm caused by a mishap, it can also be shown that the expense required to achieve safety is low in comparison to the scale and overall cost of the project. For example, the use of DNA evidence in the United States has been examined. As of October 2007, over five million DNA profiles of convicted offenders and close to 195,000 profiles from crime scene evidence have been stored in the American National DNA Index System.[139] The cost of testing a sample from a crime scene, not including expert testimony in court, varies from state to state and from laboratory to laboratory[140] and ranges from $425 to $1,720 per test.[141] A DNA test for a convicted offender is much cheaper and could cost between $25 and $110.[142] Even without making an exact calculation and without knowing the precise data it appears that the use of DNA evidence is a large-scale project with an overall cost of at least hundreds of millions of dollars. Therefore, even an investment of several million dollars in safety (the sum required for the safety program of the F-14 fighter plane) would be relatively small given the scale of the project.

Up to this point, we have been dealing with the cost of false positives. An examination of false negatives shows the tremendous benefit to society from an increased level of precision in testing, which would lead to more effective law enforcement. For example, a much greater benefit could be obtained from the DNA database, since fewer offenders would evade punishment as a result of a false negative in the testing.

Finally, we should note a significant difference between our proposal to invest in raising the level of precision in testing, and other proposals designed to prevent the conviction of innocent persons, e.g., the proposal to require additional evidence, such

[139] For this data on the FBI website, see http://www.fbi.gov/hq/lab/codis/national.htm (last visited July 17, 2008).
[140] N.C. OFFICE OF STATE BUDGET & MGMT., N.C. DEP'T OF JUSTICE, COST STUDY OF DNA TESTING AND ANALYSIS 7-8 (2006), www.osbm.state.nc.us/files/pdf_files/3-1-2006FinalDNAReport.pdf.
[141] *Id.* at 8.
[142] *Id.*

as "strong corroboration" to a confession.[143] One of the arguments against such a proposals is that, despite the great benefit of requiring "strong corroboration" to a confession, society is liable to be harmed by the acquittal of guilty persons whose confessions cannot be corroborated. However, our proposal to improve the accuracy of scientific evidence is not likely to cause such harm; the opposite is true: more correct identifications will be made and more crimes will be solved.

D. A response to a possible "case-specific" argument

It is a common mistake to think that, although the accuracy of forensic evidence is unknown, it is still possible to trust a court of law to determine – by examining the evidence and expert testimony – whether or not, in the specific case before it, all procedures have been followed so that it can be established that no error has been made.[144] As we have already shown in detail elsewhere,[145] this is just an illusion. Thus, for example, even in cases where it is determined that an error has been made, the source of the error is not always successfully identified.[146] Errors occur even when proper testing procedures have been followed and even when experts testify that all steps were taken to avoid error.[147]

Obtaining forensic evidence, such as DNA, is sometimes described as if the person conducting the test only needs to follow instructions and press a button in order to get a result.[148] Manufacturers are not enthusiastic about providing information beyond said testing instructions.[149] Defense efforts to obtain developmental validation data and primer sequences from DNA kit makers are met by the refusal of manufacturers.[150] In many cases, courts reject discovery motions by the defense for this information, for various reasons, such as the argument that these are commercial secrets.[151] However, since the development of these kits is not subject to

[143] Which, in our opinion is a necessary requirement. *See* Sangero, *supra* note 5.
[144] Sangero & Halpert, *supra* note 3, at 56-59.
[145] *Id.*
[146] *Id.* at 75.
[147] *Id.* at 74.
[148] Mellon, *supra* note 2, at 1098.
[149] *Id.*
[150] *Id.* at 1108-23.
[151] *Id.* at 1112-22.

oversight by the authorities, in contrast to the case with safety-critical systems, this is a serious problem and, if there is a defect, no one will be able to detect it.

E. The moral argument

The problem with the existing situation is demonstrated if we take the breathalyzer, which we have discussed above,[152] as an example. Safety-critical standards have never been applied to this device, just as they have never been applied to any other device producing forensic evidence. Let us assume that there was a need for this device in the medical field. As we have seen, the breathalyzer does not meet FDA standards – and, in effect, there has been no attempt to make it compatible with these standards – so it would not be approved for use in the medical field, even if its use was not liable to endanger human life. However, its use is accepted without reservation by the legal system and people are convicted of criminal offenses based on its findings.

Moreover, we are not talking about a field where the government already invests reasonable resources to ensure product safety and we are demanding that it invest further resources. The reality is that the government currently invests nothing to monitor the development of such devices. In addition, even when imperfect, non-scientific evidence is admitted in a criminal trial, this is because there is no other choice. Thus, for example, the reliability of eyewitness testimony depends on circumstances such as whether or not the incident occurred during the light of day and whether or not the witness had a good angle of view. These are circumstances over which we have no control. However, when it is the government that introduces the evidence – as is the case with forensic evidence – and it has the ability to ensure that the evidence will be more accurate, then it has a moral duty to do all that is reasonably possible so that the device producing the evidence will be more precise and not lead to the conviction of innocent persons.

This is not a matter of fate and it is not an existing danger that the government finds it hard to locate the resources to minimize (such as investing in the infrastructure of highways in order to reduce the number of accidents). This is a device that the government itself develops (whether directly or indirectly – by ordering the device

[152] *Supra* Part III.

from the manufacturer, by approving the device, or by purchasing the device for operation by the police), and its entire objective is to prove a person's guilt. This is the case with a device used to detect drunkenness, a device used to detect speeding, etc. It is inconceivable that the government would be frugal regarding such devices, which determine the fate of individuals, by conceding the highest degree of accuracy. Such a device has the power to negate an individual's liberty and it certainly has the power to stigmatize a person. Therefore, the government has a moral duty to do all that is reasonably possible and to invest the necessary resources to ensure that the device is of the highest degree of accuracy.

Such a duty may also be deduced from various theories dealing with the social contract between the state and its citizens: we never agreed that the state would impose criminal responsibility and punishment on us based on evidence that is not sufficiently reliable, when the state has the ability to improve its reliability, but fails to do so.[153]

In our opinion, until a device can be developed that meets strict standards of safety in order to avoid errors, the use of existing devices to obtain evidence for proving guilt in a criminal trial should be halted. A person's freedom is no less important than the values protected by strict standards of quality assurance intended for safety-critical systems.

V. EPILOGUE: A call for legislative reform applying safety-critical standards to forensic evidence

A device used to produce forensic evidence could also lead to a catastrophe – the conviction of an innocent person is most certainly a catastrophe. There is no reason why safety-critical standards, which have been developed as a result of technological progress and which are based on considerable experience, should continue to be ignored in the field of forensic evidence, leaving this field to lag behind. Not only is this situation illogical and embarrassing – it is also morally wrong and demands an immediate change.

[153] *See* Rinat Kitai, *Protecting the Guilty*, 6 BUFF. CRIM. L. REV. 1163, 1172-79, 1186-87 (2003).

We call on lawmakers to enact legislation to monitor and regulate the development of forensic equipment designed for use by the legal system. A manufacturer who wants to sell a forensic device for use in the legal process should be required to obtain approval in a similar manner as that required of the manufacturers of medical devices. We also call on lawmakers to establish that a prerequisite for the admissibility of evidence obtained through a forensic device intended for use in the legal process – at least when such evidence is to be used in a criminal trial – is that the forensic device be developed and monitored according to the same principles laid down for safety-critical systems.