

Winter 2010

Learning from Error in American Criminal Justice

James M. Doyle

Follow this and additional works at: <https://scholarlycommons.law.northwestern.edu/jclc>

 Part of the [Criminal Law Commons](#), [Criminology Commons](#), and the [Criminology and Criminal Justice Commons](#)

Recommended Citation

James M. Doyle, Learning from Error in American Criminal Justice, 100 J. Crim. L. & Criminology 109 (2010)

This Criminal Law is brought to you for free and open access by Northwestern University School of Law Scholarly Commons. It has been accepted for inclusion in Journal of Criminal Law and Criminology by an authorized editor of Northwestern University School of Law Scholarly Commons.

LEARNING FROM ERROR IN AMERICAN CRIMINAL JUSTICE

JAMES M. DOYLE*

While an operator error may be the proximate “cause” of the accident, the root causes were often present within the system for a long time. The operator has, in a real sense, been “set up” to fail by poor design, faulty maintenance, or erroneous management decisions.

—Lucian L. Leape¹

[V]irtually every system we will examine places “operator error” high on its list of causal factors—generally about 60 to 80 percent of accidents are attributed to this factor. But if, as we shall see time and again, the operator is confronted by unexpected and usually mysterious interactions among failures, saying that he should have zigged instead of zagged is possible only after the fact. Before the accident no one could know what was going on and what should have been done.

—Charles Perrow²

Wrongful convictions and other criminal justice system errors can be seen as “organizational accidents” in which small mistakes (no one of which would suffice to cause the event) combine with each other and with latent defects in the criminal justice system to create disasters. Employing this conception of error in a consistent routine of examination of wrongful convictions, near misses, and other errors can increase the impact of the lessons of error, mitigate the fragmentation of the criminal justice system, and lay the foundation, as it has in medicine and aviation, for the creation of a “culture of safety.”

* Consulting Director, Center for Modern Forensic Practice, John Jay College of Criminal Justice, City University of New York; Of Counsel, Carney & Bassil, Boston, Massachusetts. B.A. Trinity College; J.D. Northwestern University School of Law; LL.M. Georgetown University Law Center. Contact: jdoyle@jjay.cuny.edu.

¹ Lucian L. Leape, *Error in Medicine*, 272 JAMA 1851, 1854 (1994).

² CHARLES PERROW, *NORMAL ACCIDENTS* 9 (1984).

I. INTRODUCTION

American criminal justice practitioners seem increasingly ready to face their errors and learn from them. Are they edging toward a new orientation? Is there potential for a cultural change? Could criminal justice practitioners adopt some version of the quality reform initiatives that have reshaped other high-risk fields such as aviation and medicine? Can the criminal justice system embrace “a theory of work, which conceptualize[s] the continual improvement of quality as intrinsic to the work itself”?³ Is it possible that the current era, defined by episodic patches motivated by high-profile tragedies, will be replaced by a new period, dedicated to the sustained *practice* of learning from error?

Take, as a starting point, the U.S. Department of Justice’s landmark 1998 study of the first twenty-eight wrongful convictions exposed by DNA testing, *Convicted By Juries, Exonerated By Science*.⁴ Among criminal justice practitioners and policy-makers, this “Green Book” quickly became the most talked-about publication of the year. Every veteran practitioner had made or seen mistakes. Still, for most, this recognition of fallibility was a nagging ache that they learned to live with, not a sharp pain that provoked them to action. Suddenly, practitioners confronted concrete, specific, and irrefutable proof of tragic errors.

The errors identified in the Green Book came from the sort of bread-and-butter cases that everyone had handled and would handle again, not from arcane borderland specialties. Innocent men who were convicted by the testimony of sincere eyewitnesses, our oldest form of evidence, dominated the list of exonerated prisoners. Twenty-four of the twenty-eight cases involved misidentifications. By the time Barry Scheck, Peter Neufeld, and Jim Dwyer published *Actual Innocence* in 2000, the roster was up to 53 innocent defendants imprisoned for an aggregate 197 years by 77 mistaken eyewitnesses.⁵ Eight of the exonerated had been sentenced to death. Fifty-three actual rapists and murderers had been left free to find additional victims.⁶ The current exoneration count stands at over two

³ CHARLES KENNEY, *THE BEST PRACTICE: HOW THE NEW QUALITY MOVEMENT IS TRANSFORMING MEDICINE* 30 (2008).

⁴ EDWARD CONNORS ET AL., *NAT’L INST. JUST., CONVICTED BY JURIES, EXONERATED BY SCIENCE: CASE STUDIES IN THE USE OF DNA EVIDENCE TO ESTABLISH INNOCENCE AFTER TRIAL* (1996). The “Green Book” is certainly only a starting place. A genuinely heroic body of scholarship has marshaled and categorized the exoneration cases. See Brandon L. Garrett, *Judging Innocence*, 108 COLUM. L. REV. 55 (2008); Samuel R. Gross et al., *Exonerations in the United States 1989 Through 2003*, 95 J. CRIM. L. & CRIMINOLOGY 523 (2005)..

⁵ BARRY SCHECK, PETER NEUFELD & JIM DWYER, *ACTUAL INNOCENCE: FIVE DAYS TO EXECUTION AND OTHER DISPATCHES FROM THE WRONGLY CONVICTED* 264 (2000).

⁶ *Id.*

hundred; eyewitness misidentification cases still provide the largest element of the list.⁷

There were efforts to dismiss the Green Book and later Innocence Project compilations as catalogues of freakish mishaps: at best noise in the data; at worst, bleeding-heart propaganda. This argument (made by Justice Scalia, among others)⁸ gained very little traction with the public, in part because every time it was put forward, the Innocence Project exposed another horrifying wrongful conviction.

The criminal justice system's front-line practitioners—the people who actually do the work on the streets and in the courts—showed little interest in the comfort that the system's apologists tried to offer them. The criminal practitioners were drowning in heavy caseloads, so they knew that even very low *rates* of error would still result in a very high absolute number of tragedies. More importantly, practitioners felt that the rarefied utilitarian calculations that absorbed Justice Scalia were beside the point. For practitioners, avoiding errors was a matter of professionalism, workmanship, and, ultimately, self-respect; it was not a matter of social policy.⁹ The front-line troops accepted the Green Book as a call to action: one error was too many. Dozens of jurisdictions, acting independently of one another, mobilized efforts to address the problems identified in the Green Book.

The initial leadership came from different players in different places. Attorney General Janet Reno, who decided that the Green Book would include commentary from the full spectrum of criminal justice system actors, provided an influential template. Under the auspices of the National Institute of Justice, she convened mixed “Technical Working Groups,” which brought together diverse stakeholders to hammer out and publicize new criminal justice “best practices.” These groups addressed crime scene investigations, death investigations, and eyewitness evidence, among other

⁷ Garrett, *supra* note 4, at 78-80.

⁸ See *Kansas v. Marsh*, 548 U.S. 163, 199 (2006) (“Like other human institutions, courts and juries are not perfect. One cannot have a system of criminal punishment without accepting the possibility that someone will be punished mistakenly. That is a truism, not a revelation. But with regard to the punishment of death in the current American system, that possibility has been reduced to an insignificant minimum.”) (Scalia, J., concurring). See generally Joshua Marquis, *The Myth of Innocence*, 95 J. CRIM. L. & CRIMINOLOGY 501 (2005) (arguing that incidence of wrongful convictions is overstated in advocates’ public accounts).

⁹ Kenneth Patenaude, *Improving Eyewitness Identifications*, LAW ENFORCEMENT TECH., Oct. 2008, at 178 (explaining the importance of heeding exoneration lessons).

topics.¹⁰ Peter Neufeld and Barry Scheck, the co-founders of the Innocence Project, who had been among Reno's Green Book commentators, called for a learning-from-error initiative.¹¹ In North Carolina, the first impetus came from the conservative Republican chief justice of the North Carolina Supreme Court.¹² In Boston, it came from the elected district attorney;¹³ in Illinois, from Northwestern University's Center on Wrongful Convictions and the Governor's Commission on Capital Punishment;¹⁴ and in New Jersey, from a Republican attorney general.¹⁵ Every time judges, cops, prosecutors, or Innocence Network lawyers took steps forward, they found allies from all points of the criminal justice system, often among their courtroom adversaries.

The success of these novel and diverse groups often took their participants by surprise—their members were not accustomed to working together. But it has turned out that no one became a police officer or lawyer to participate in putting the innocent in prison so that the guilty could remain free. Battle-scarred detectives and unfledged public defenders had something in common: they took the phrase “good enough for government work” as an insult and a goad when it was applied to their own professional lives.¹⁶

¹⁰ NAT'L INST. JUST., CRIME SCENE INVESTIGATION: A GUIDE FOR LAW ENFORCEMENT (2000); NAT'L INST. JUST., EYEWITNESS EVIDENCE: A GUIDE FOR LAW ENFORCEMENT (1999); NAT'L INST. JUST., DEATH INVESTIGATIONS: A GUIDE FOR THE SCENE INVESTIGATOR (1999).

¹¹ Barry C. Scheck & Peter J. Neufeld, *Toward the Formation of "Innocence Commissions" in America*, 86 JUDICATURE 98, 99 (2002).

¹² Christine C. Mumma, *The North Carolina Actual Innocence Commission: Uncommon Perspectives Joined by a Common Cause*, 52 DRAKE L. REV. 647, 648-49 (2004); Matthew Eiseley, *Better ID Sought in Criminal Inquiries*, NEWS & OBSERVER (Charlotte, N.C.), Sept. 13, 2003, at B1. The North Carolina Actual Innocence Commission also serves as a remedial vehicle. See Jerome M. Maiatico, Note, *All Eyes on Us: A Comparative Critique of the North Carolina Innocence Inquiry Commission*, 56 DUKE L.J. 1345, 1356 (2007).

¹³ Daniel F. Conley, *Our Duty to Free the Wrongly Convicted*, BOSTON GLOBE, Mar. 19, 2004, at A14. See generally Garrett, *supra* note 4, at 56; Press Release, Report of the Task Force on Eyewitness Evidence to Daniel F. Conley, District Attorney, Suffolk County, and Kathleen M. O'Toole, Commissioner, Boston Police Department (July, 2004), available at http://www.innocenceproject.org/docs/Suffolk_eyewitness.pdf.

¹⁴ STATE OF ILL., REPORT OF THE GOVERNOR'S COMMISSION ON CAPITAL PUNISHMENT (2002), available at http://www.idoc.state.il.us/ccp/ccp/reports/commission_report/summary_recommendations.pdf. See generally SCOTT TUROW, *ULTIMATE PUNISHMENT* (2003) (reviewing lessons of service on Illinois Governor's Commission on Capital Punishment).

¹⁵ Letter from John J. Farmer, Jr., New Jersey Attorney-General, to Various Recipients, *Attorney General Guidelines for Preparing and Conducting Photo and Live Lineup Identification Procedures* (Apr. 18, 2001).

¹⁶ Patenaude, *supra* note 9; see also JAMES DOYLE, *TRUE WITNESS: COPS, COURTS, SCIENCE AND THE BATTLE AGAINST MISIDENTIFICATION* 169-87 (discussing participant's account of Technical Working Group discussions).

The post-exoneration landscape began to be marked by examples of tentative cooperation. To call this development a movement¹⁷ captures some of its momentum, but the term obscures the fact that these initiatives arose organically from largely uncoordinated local efforts. They were spurred by local law enforcement, the local bar, or the local judiciary, often in response to local journalists' coverage of exonerations.¹⁸ By now, these scattered alliances have produced a substantial body of achievements affecting the way crimes are investigated and adjudicated. Legislation, collections of best practices, binding administrative guidelines, non-binding guides, and enhanced training programs have been created or are underway on national, state, and local levels. They can be expected to multiply.

The purpose of this Article is not to catalogue these efforts or to consider their merits; it is to examine them as precursors and to ask toward what goal these first steps could lead. In examining that question, this Article will mobilize the experiences of medicine and aviation in blazing their own trails toward a culture of safety to illuminate what may develop from the criminal justice system's efforts to learn from error. By recognizing some of the self-imposed limits of the current approach, we can reveal the potential for broader and deeper change.

II. THE WRONG MAN: EYEWITNESS "SYSTEM VARIABLES"

The effort to integrate the science of perception and memory into eyewitness evidence collection (in interviews, photo arrays, and lineups) is the reform initiative that has moved forward most rapidly and in the greatest number of jurisdictions in the aftermath of the Green Book's initial shock. From a distance, the implementation of new eyewitness procedures presents a simple story. By an accident of history (this version goes), when the Green Book's DNA exoneration catalogue was released, there was a body of psychological science on eyewitness performance ready to be taken down from its shelf and immediately employed. In fact, the history of the eyewitness reforms is both more complex and more instructive. It shows that psychologists were capable of making important science-based recommendations and that members of the criminal justice system were willing to embrace the recommendations.

¹⁷ Garrett, *supra* note 4, at 57-58. An important element of the wave of reform, the electronic recording of interrogations, is described in Thomas P. Sullivan, *Electronic Recording of Custodial Interrogations: Everybody Wins*, 95 J. CRIM. L. & CRIMINOLOGY 1127 (2005).

¹⁸ See, e.g., JON B. GOULD, *THE INNOCENCE COMMISSION: PREVENTING WRONGFUL CONVICTIONS AND RESTORING THE CRIMINAL JUSTICE SYSTEM* (2008) (providing a narrative of the development of state innocence commission).

Legal skepticism about eyewitness reliability is immemorial and, at least since the publication over one hundred years ago of Hugo Münsterberg's *On the Witness Stand*,¹⁹ has been reinforced by psychological findings about the process of human perception, memory storage, and recall. Until the early 1970s, practitioners in the fields of law and psychology pursued their concerns about eyewitness reliability along parallel paths. In Warren Court identification decisions such as *Wade*, *Stovall*, and *Neil*, the legal system focused on suggestive police misconduct as the source of eyewitness error.²⁰ Psychologists continued to pursue research following the general lines that Münsterberg and his contemporaries had laid down: they rejected the model of a simple, photographic, and permanent memory capacity in humans, and explored memory's complexities and frailties.

But in the early 1970s the two fields were steered into a collision by Robert Buckhout, a psychology professor at Brooklyn College and dedicated *provocateur*. Buckhout demanded a place for eyewitness psychology in the courts. He staged demonstrations of eyewitness unreliability. For example, he created a televised crime and lineup that placed viewers in the role of eyewitness.²¹ The viewers called in with their picks for the perpetrator, and the results showed fewer correct answers than random guessing would have achieved.²² More importantly, Buckhout testified in court.²³ His expert testimony in the Angela Davis murder trial (a media circus of its era) was widely regarded as having led to Davis's acquittal in the face of what had seemed to be an overwhelming prosecution case.²⁴ Buckhout reopened the issue when he built on that success with a survey article in *Scientific American*²⁵ and began an energetic campaign of encouraging defense lawyers to raise psychological findings in their eyewitness cases. Buckhout had created a market for eyewitness psychology among desperate defense lawyers, and they clamored for more.

¹⁹ HUGO MÜNSTERBERG, *ON THE WITNESS STAND: ESSAYS ON PSYCHOLOGY AND CRIME* (1908).

²⁰ *Neil v. Biggers*, 409 U.S. 188 (1972); *United States v. Wade*, 388 U.S. 218 (1967); *Stovall v. Denno*, 388 U.S. 293 (1967). See generally Felice J. Levine & June Louin Tapp, *The Psychology of Criminal Identification: The Gap from Wade to Kirby*, 121 U. PA. L. REV. 1079 (1973) (collecting then extant psychological studies on eyewitness performance).

²¹ Robert Buckhout, *Nearly 2,000 Witnesses Can Be Wrong*, 16 BULL. PSYCHONOMIC SOC'Y 307 (1980). On Buckhout and his era and the history of the law/psychology interactions generally, see DOYLE, *supra* note 16.

²² *Id.*

²³ Buckhout, *supra* note 22.

²⁴ MARY TIMOTHY, *JURY WOMAN: THE STORY OF THE TRIAL OF ANGELA DAVIS* 220-21 (1976).

²⁵ Robert Buckhout, *Eyewitness Testimony*, 231 SCI. AM. 23 (1974).

It seemed to many—and especially to the prosecutors Buckhout gleefully taunted—that if Buckhout was given his way he would eliminate eyewitness evidence altogether.

At this point Elizabeth Loftus entered the scene.²⁶ Where Buckhout had staged dramatic, media-friendly demonstrations, Loftus devised imaginative and technically scrupulous laboratory experiments. Her inquiry into the malleability of memory had a number of far-reaching impacts. First, by applying rigorous scientific procedure to the eyewitness inquiry, she won entry for the topic into the most prestigious peer-reviewed journals. Young research psychologists saw how the era's idealistic urge to be socially relevant—the desire to “give psychology away”²⁷—could be combined with an academic career; the volume of published research into eyewitness questions exploded.²⁸ But beyond that, Loftus's experiments showed that memory evidence was “trace evidence.” Like blood or semen found at a crime scene, memory evidence was difficult to recover, easy to contaminate, and, even worse, once exposed to contamination, impossible to take back into a laboratory to test for whether the contaminates had taken effect.

Lawyers and psychologists settled into a dialogue tightly focused on the issue of whether Loftus and her colleagues should be permitted to testify as expert witnesses in court. This dialogue often resembled an exchange of mortar fire between opposing trenches. One or two appellate courts required the admission of the testimony,²⁹ others ordered its exclusion, but most required trial judges to exercise their discretion in deciding to admit or exclude expert testimony on the vulnerabilities of eyewitness memory.³⁰ Everyone settled in to grind out the argument in case-by-case adversarial combat.

²⁶ ELIZABETH F. LOFTUS & KATHERINE KETCHAM, WITNESS FOR THE DEFENSE: THE EYEWITNESS AND THE EXPERT WHO PUT MEMORY ON TRIAL (1991); *see also* DOYLE, *supra* note 16, at 83-100.

²⁷ George A. Miller, *Psychology as a Means of Promoting Human Welfare*, 24 AM. PSYCHOLOGIST 1063, 1071 (1969).

²⁸ BRIAN L. CUTLER & STEVEN D. PENROD, MISTAKEN IDENTIFICATION: THE EYEWITNESS, PSYCHOLOGY, AND THE LAW 67-70 (1995).

²⁹ *See, e.g.*, State v. Chapple, 660 P.2d 1208 (Ariz. 1983) (en banc); People v. McDonald, 690 P.2d 709 (Cal. 1984).

³⁰ *See, e.g.*, United States v. Brownlee, 454 F.3d 131 (3d Cir. 2006). A gargantuan body of law review commentary—beginning with Fredric D. Woocher, *Did Your Eyes Deceive You? Expert Psychological Testimony on the Unreliability of Eyewitness Identification*, 29 STAN. L. REV. 969 (1977), and continuing through Jules Epstein, *The Great Engine that Couldn't: Science, Mistaken Identification, and the Limits of Cross-Examination*, 36 STETSON L. REV. 727 (2007), and beyond—addresses the expert witness issue.

The deck was reshuffled by Gary Wells. Wells argued that Buckhout, Loftus, and the psychologists who had followed their banner into courtroom battles had chosen the wrong point of entry.³¹ To begin with, Wells noted that the results of the Loftus experiments were statistical: they could tell you what happened eight times out of ten, but they could not tell you whether *this* case was one of the eight. Besides, when it came to courtroom relevance, Loftus was a prisoner of her own scientific scrupulousness. Her meticulous studies required that she isolate and manipulate one factor (stress, post-event information, duration, darkness, race) in an eyewitness encounter while controlling all others. But actual crime events were composed of hundreds of factors. Science had no mechanism for aggregating these separate factors and accounting for potential interactions. Even if one artificially limited an event to twenty factors with either high or low influence, there were over a million potential outcomes. Experimental psychological findings—*especially* the best and most careful of them—make very clumsy tools for use in courtroom attempts to catch eyewitness errors in specific cases *after* they have been made.

Wells suggested that psychologists should work to prevent errors (or at least reduce the error rate) rather than claim the ability to diagnose errors from the witness stand.³² Moreover, he argued that psychologists should acknowledge that many experimental variables were not improvable in real crime environments. Factors like lighting, duration, race, age, or presence of a weapon, which Wells christened “estimator variables,” were the wrong things—or at least not the first things—to address.³³

Wells contended that the initial emphasis should be placed on choices that were under the control of criminal justice system practitioners—factors which he christened “system variables.”³⁴ How did police question eyewitnesses? How were fillers picked for lineups? Who ran the lineups? What instructions were witnesses given at a lineup? What were the impacts of these decisions on reliability? If we knew the answers to those questions, we could shape investigative practice accordingly.

The answers that Wells and his colleagues developed over the following decade led them to propose concrete modifications to police procedure. Four of these changes were recommended in a White Paper of the American Psychology-Law Society: choose lineup fillers to match the verbal description of the perpetrator (not the suspect); instruct witnesses viewing a lineup that the perpetrator may or may not be present; have the

³¹ Gary L. Wells, *Applied Eyewitness Testimony Research: System Variables and Estimator Variables*, 36 J. PERSONALITY & SOC. PSYCH. 1546 (1978).

³² *Id.*

³³ *Id.*

³⁴ *Id.* at 1548.

lineup or photo array conducted by a “blind” administrator who does not know which lineup member is the suspect; and immediately record the witness’s statement of confidence in any choice the witness has made.³⁵ When these four recommendations were supplemented by a fifth—show the lineup members one at a time (sequentially) rather than in the traditional line (simultaneously)³⁶—the new “double-blind, sequential” lineup protocol was created. Laboratory results indicated that the new double-blind, sequential technique would—by eliminating unconscious police influence and muting the witnesses’ natural tendency to resort to a relative judgment choice of “looks-most-like” innocents from arrays and lineups that did not include the actual perpetrator—produce results twice as reliable as the traditional simultaneous method.³⁷

By the end of the 1990s, scattered jurisdictions wracked by eyewitness exonerations moved to adopt versions of Wells’s system-variable program and employed double-blind, sequential identification techniques. First, New Jersey’s attorney general ordered statewide adoption of the reforms.³⁸ The National Institute of Justice published the *Guide for Law Enforcement*, which recommended best practices. The Suffolk County (greater Boston area) District Attorney convened an eyewitness task force,³⁹ and then he followed New Jersey’s lead and instituted double-blind, sequential procedures. The Hennepin County (Minneapolis) State’s Attorney⁴⁰ and the Wisconsin Attorney General⁴¹ did the same.

Science—by means of DNA exonerations—confronted the criminal justice system with errors; the criminal justice system responded by applying the lessons of psychological science. In a variety of jurisdictions

³⁵ Gary L. Wells et al., *Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads*, 22 LAW & HUM. BEHAV. 603, 627-35 (1998).

³⁶ Of the elements of the reformers’ program, the superiority of the “sequential” technique is the most contested. See generally Roy S. Malpass, *A Policy Evaluation of Simultaneous and Sequential Lineups*, 12 PSYCH. PUB. POL’Y & L. 394 (2006). But see, e.g., Nancy Steblay et al., *Eyewitness Accuracy Rates in Sequential and Simultaneous Lineup Presentations: A Meta-Analytic Comparison*, 25 LAW & HUM. BEHAV. 459 (2001). For a study indicating that the “sequential” technique is superior for some lineups but not for others, see Michael R. Leippe, Donna Eisenstadt & Shannon M. Rauch, *Cueing Confidence in Eyewitness Identifications: Influence of Biased Lineup Instructions and Pre-Identification Memory Feedback Under Varying Lineup Conditions*, 33 LAW & HUM. BEHAV. 194 (2009).

³⁷ Steblay et al., *supra* note 36.

³⁸ DOYLE, *supra* note 16, at 189-97.

³⁹ Press Release, *supra* note 13.

⁴⁰ Amy Klobuchar et al., *Improving Eyewitness Identifications: Hennepin County’s Blind Sequential Lineup Pilot Project*, 4 CARDOZO PUB. L. POL’Y & ETHICS J. 381 (2006).

⁴¹ OFFICE OF THE ATTORNEY GENERAL, WIS. DEP’T OF JUSTICE, MODEL POLICY AND PROCEDURE FOR EYEWITNESS IDENTIFICATION (2005), available at <http://www.thejusticeproject.org/wp-content/uploads/eyewitnesspublic1.pdf>.

and contexts, the criminal justice system gathered its divergent practitioners—prosecutors, judges, defenders, scientists, and police—and they cooperatively addressed its policies and practices. Longstanding traditional lineup routines were abandoned in favor of the science-based reforms. It proved that it is not inevitable that when confronted with errors, the system’s practitioners will simply shrug and hope things get better on their own.

Still, was all of this anything more than patching holes in the blacktop?

III. THE WRONG MAN MEETS THE WRONG PATIENT

Elements of the eyewitness reforms could prove to be harbingers of something considerably more interesting than isolated fixes to specific problems: they reveal the potential for a shift from the retrospective, adversary inspection model of quality control toward an ideal of continuous quality improvement. The eyewitness reforms provide an opportunity to exploit the experience of aviation and medicine—two other high-risk enterprises in which even a low rate of error can result in catastrophic costs—in trying to improve criminal justice system quality.

Gary Wells’s decision to focus on prevention of eyewitness error instead of its retrospective diagnosis was more than a natural adaptation to the limitations of psychological science. This commonsense adjustment conceals a potential paradigm shift: jurisdictions that adopted double-blind, sequential lineups overcame the legal system’s deep cultural tradition of treating errors as the work of “bad apples.”

The endemic assumption in the criminal justice system, as in medicine, had always been “good man, good result.”⁴² As Dr. Lucian Leape wrote in his seminal 1994 essay, *Error in Medicine*:

Physicians are expected to function without error, an expectation that physicians translate into the need to be infallible. One result is that physicians, not unlike test pilots, come to view error as a failure of character—you weren’t careful enough, you didn’t try hard enough. This kind of thinking lies behind a common reaction by physicians: “How can there be an error without negligence?”⁴³

In criminal justice culture, “[h]omicide detective,” or “prosecutor,” or “defender,” or “judge” substitutes effortlessly for “physician” in Leape’s analysis. In this conception, any error (or at least any error that couldn’t be disposed of comfortably with “shit happens”) is an *operator* error: some surgeon, or cop, or nurse, or forensic scientist, or lawyer at the site of the event was lazy, or ill-trained, or venal, or careless.

⁴² Donald Berwick, *Continuous Improvement as an Ideal in Healthcare*, 320 NEW ENGL. J. MED. 53, 53-55 (1989).

⁴³ Leape, *supra* note 1, at 1851.

The legal system traditionally blamed eyewitness unreliability on the misconduct of bad apple police officers who might be “rigging” lineups. The Warren Court’s misconduct-oriented opinions were a manifestation of this approach to quality control. The criminal system maintains an elaborate, retrospective, adversarial *inspection* of criminal investigations. Inspection proceeds through grand jury indictment, pretrial discovery, suppression motions, cross-examination, argument, and instruction, all of which are conducted with an error-averse presumption of innocence and reasonable doubt standard as backups at the terminal adversary trial phase of the process. Some people see this criminal process as an obstacle course pointlessly frustrating prosecutors, others as a conveyor belt mindlessly “processing” defendants,⁴⁴ but everyone in the system agrees that mistakes come from bad apples and that litigation at the end of this process is well-designed to find them.

The Green Book did list a number of cases that conformed to the bad apple model. For example, it described the exploits of the rogue forensic scientist, Fred Zain, who fabricated numerous laboratory tests and invented fictitious results.⁴⁵ Better inspection in labs and courtrooms seemed to be an obvious strategy for frustrating the world’s Fred Zains.⁴⁶ But the Green Book’s *eyewitness* exonerations presented cases in which bad apples had played no role. The eyewitness exonerations typically were villainless tragedies in which the witnesses were sincere, and the cops went “by the book” (as the book then stood) and usually were encouraged by some form of apparent corroboration.

The system-variable eyewitness reforms were revolutionary because they uncoupled quality improvement from retrospective inspection. They linked quality to science-based improvements to practice early in the investigative process, upstream from the ultimate inspection stage error.

As it turned out, the absence of villains from the eyewitness exoneration cases opened a shortcut that medical reformers had already struggled to find.⁴⁷ Dr. Donald Berwick, a pioneer in the medical quality movement, slaved for years to develop a system for measuring various

⁴⁴ HERBERT L. PACKER, *THE LIMITS OF THE CRIMINAL SANCTION* 159-65 (1968). See generally Erik Luna, *System Failure*, 42 AM. CRIM. L. REV. 1201 (2005) (suggesting the potential application of “systems theory” in its most general form to criminal justice and exonerations).

⁴⁵ CONNORS ET AL., *supra* note 4, at 18.

⁴⁶ See Brandon L. Garrett, *Aggregation in Criminal Law*, 95 CAL. L. REV. 383, 412-16 (2007); Paul C. Giannelli, *Wrongful Convictions and Forensic Science: The Need to Regulate Crime Labs*, 86 N.C. L. REV. 163, 172-74 (2007).

⁴⁷ See generally KENNEY, *supra* note 3, at 15-47 (recounting the early history of medical quality movement).

aspects of health care performance. He was convinced that one had to measure doctors' performance before one could improve any aspect of practice.⁴⁸ That basic task, he pointed out, was not being carried out. Berwick began to compile comprehensive statistics (on, for example, waiting times and medication errors), and he then tried to persuade the physicians at the Harvard Community Health Plan to pay attention to the inconsistencies that his numbers revealed.⁴⁹ When Berwick proudly exposed his statistical treasures at his first meeting with the doctors he proposed to help, an enraged doctor crumpled Berwick's report, threw it in Berwick's face, and then stalked from the room.⁵⁰

After versions of this experience had been repeated a few times, Berwick realized that any program of measurement was inextricably associated in the doctors' minds with a system of surveillance and post hoc inspection, which had blaming as its sole purpose and public ignominy as its only outcome. Measurement of performance in medicine was so inconsistent because no one saw it as personally advantageous to be measured: having your performance measured could only land you in a world of pain.

It was clear to Berwick that the exposure of imperfections would be a source of personal pain for the objects of his studies.⁵¹ This suffering was bad enough when it involved only distended waiting times for radiological procedures, but it was immeasurably worse when it revealed mistakes that harmed patients. The catastrophic impact of the tort system was obvious. But beyond that, the problem was made more acute by the fact that, from their first days in medical or nursing school, everyone in medicine had been indoctrinated with a code combining total individual responsibility for patients and perfectionism in performance. The in-house tradition of morbidity and mortality (M & M) reviews of surgical accidents might as well have been designed to reinforce a determination to treat every error as an operator error and every operator error as evidence of moral or professional failing in the operator.⁵² In theory, M & M presentations were an opportunity to learn from mistakes; in fact, they were experienced as a dreaded ceremonial warning that perfection was the requirement and that individual responsibility for perfection was indivisible and indispensable.⁵³

⁴⁸ *Id.*

⁴⁹ *Id.*

⁵⁰ *Id.* at 17.

⁵¹ Berwick, *supra* note 42, at 53.

⁵² ATUL GAWANDE, *When Doctors Make Mistakes* (1999), reprinted in *COMPLICATIONS: A SURGEON'S NOTES ON AN IMPERFECT SCIENCE* 47 (2002).

⁵³ *Id.*

The result of this preoccupation with perfectionism was a “cycle of fear”⁵⁴ in which medical professionals ignored or suppressed accounts of errors, thereby undermining efforts to prevent such errors in the future. Presumably practitioners directly involved with an error were scarred by the experience and took its lessons with them, but the lessons were not shared. Everyone agrees that, in the early days of the medical reform campaign, errors were dramatically under-reported.⁵⁵

But the reformers quickly recognized errors as a powerful lever. They knew that although doctors saw things through the lens of the individual physician-patient encounter, and while hospital administrators and risk managers saw things through the statistical lens of the survival of a large functioning enterprise, *everyone* hated error and its costs.⁵⁶ Berwick and his allies made the phrase “[e]very defect is a treasure” their battle-cry.⁵⁷ The National Institute of Medicine published a landmark volume, *To Err is Human: Building a Safer Health Care System*, that hammered on the themes that the study of errors was useful and that the cultural and institutional barriers to studying errors should be attacked.⁵⁸

Berwick and his colleagues believed that one of the things driven underground by medicine’s otherwise admirable ideology of perfectionism and personal responsibility was the crucial fact that most errors were committed by capable people operating within systems that did not account for human imperfections.⁵⁹ Dr. Lucien Leape noted that although “[m]ortality and morbidity conferences, incident reports, risk management activities, and quality assurance committees abound” in hospitals, they all focused on incidents and individuals.⁶⁰ As Leape saw it, “[r]oot causes in the underlying systems were not sought. No one assumed—as an aviation engineer would have—that errors and failures are inevitable and tried to design systems to prevent or absorb them.”⁶¹ Leape explained: “While the

⁵⁴ WILLIAM W. SCHERKENBACH, *THE DEMING ROUTE TO QUALITY AND PRODUCTIVITY: ROAD MAPS AND ROADBLOCKS* (1986).

⁵⁵ *TO ERR IS HUMAN: BUILDING A SAFER HEALTH SYSTEM* (Linda T. Kohn, Janet M. Corrigan & Molla S. Donaldson eds., National Academy Press 2000) [hereinafter *TO ERR IS HUMAN*]; R. Lawton & D. Parker, *Barriers to Incident Reporting in a Healthcare System*, 11 *QUALITY & SAFETY HEALTH CARE* 15 (2002). Critics of the criminal justice system similarly argue that the known wrongful conviction cases represent merely the tip of a much larger iceberg. See generally SCHECK ET AL., *supra* note 5. Its defenders argue that something close to perfection has been attained. See generally Marquis, *supra* note 8.

⁵⁶ KENNEY, *supra* note 3, at 71.

⁵⁷ Berwick, *supra* note 42, at 54.

⁵⁸ *TO ERR IS HUMAN*, *supra* note 55.

⁵⁹ *Id.*

⁶⁰ Leape, *supra* note 1, at 1855.

⁶¹ *Id.*

proximal error leading to an accident is, in fact, usually a ‘human error,’ the causes of that error are often well beyond the individual’s control. All humans err frequently. Systems that rely on error-free performance are doomed to fail.”⁶² Of course, it is an inspiring thing when a surgeon takes responsibility for anything that happened to *his* patient during an operation, but in fact the surgeon had no real power over the lab error or medication mix-up that later took effect in the operating room. This insight cleared the way for the reformers to explore the error-elimination practices of aviation and other high risk enterprises and compare them with the practices of medicine. Leape pointed to the contrasts between a National Transportation Safety Board post-accident review and prevailing health care practice:

For example, if a nurse gives a medication to the wrong patient, a typical response would be exhortation or training in double-checking the identity of both patient and drug before administration. Although it might be noted that the nurse was distracted because of an unusually large case load, it is unlikely that serious attention would be given to evaluating overall work assignments or to determining if large case loads have contributed to other kinds of errors. It is even less likely that questions would be raised about the wisdom of a system for dispensing medications in which safety is contingent on inspection by an individual at the end point of use. Reliance on inspection as a mechanism of quality control was discredited long ago in industry More imaginative solutions could easily be found—if it were recognized that both systems and individuals contribute to the problem.⁶³

Guided by this perception, Leape, Berwick, and their colleagues now harnessed the growing literature on the psychology of accidents and human error developed in aviation and industrial investigations in their crusade for health care quality reform.

Writers such as James Reason,⁶⁴ Charles Perrow,⁶⁵ and Diane Vaughan⁶⁶ derived lessons from the tragic misadventures of humans interacting with advanced technologies. Medical quality pioneers applied these lessons to health care environments where—as in criminal justice—the mix of human beings and complex technologies was dominated by the impact of human performance, not by the reliability of technological innovations. That body of knowledge and experience now offers a new direction for criminal justice if practitioners and policy-makers succeed in mobilizing it in the criminal system.

The medical quality movement’s human error approach to generating the narrative of a case-gone-wrong provides an illuminating contrast to the

⁶² *Id.* at 1852.

⁶³ *Id.*

⁶⁴ JAMES REASON, HUMAN ERROR (1990).

⁶⁵ PERROW, *supra* note 2.

⁶⁶ DIANE VAUGHAN, THE CHALLENGER LAUNCH DECISION: RISKY TECHNOLOGY, CULTURE AND DEVIANCE AT NASA (Univ. of Chi. Press 1996).

criminal justice system's approach. The standard review of a "wrong man" conviction exposed by DNA has been a laconic narrative along the lines of: "The witness picked the wrong guy; we believed him; the jury believed him too."⁶⁷ For example, the Innocence Project posts on its website this account of the Massachusetts case of Denis Maher:

On November 16, 1983, a 28-year-old woman was attacked as she was walking home from work in Lowell, Massachusetts. An unknown man accosted her and tried to engage her in conversation before forcing her into a nearby yard, where he sexually assaulted her. The following evening, a 23-year-old woman was attacked less than one hundred yards away from the site of the first assault.

The second victim had been on her way home from work when she was pushed to the ground by a man wielding a knife. She was able to escape her assailant after a vehement struggle and notified the police. The victim described her attacker as a man wearing a red, hooded sweatshirt and a khaki military-style jacket.

Dennis Maher was stopped and questioned by police on the night of the second attack. He was wearing a red, hooded sweatshirt and a subsequent search of his vehicle turned up an army field jacket, a military issue knife, and a rain slicker. Maher, then a sergeant in the United States Army, was arrested and charged with the two attacks, as well as an unsolved rape that occurred the previous summer in Ayer, Massachusetts. Though their descriptions varied, all three victims identified Maher in photographic lineups.⁶⁸

Contemporary reviews of "wrong patient" events are very different from these "wrong man" criminal justice wrongful conviction narratives.

In an article reviewing the infliction of an invasive procedure on the wrong patient for *Annals of Internal Medicine*, Dr. Mark Chassin and Dr. Elise Becher examined a situation in which at least two bad apples were certainly available: a nurse had mistakenly brought the wrong patient, and an attending physician had failed to introduce himself to the patient at the beginning of the procedure.⁶⁹ But in reviewing the situation, Chassin and Becher followed Leape and Berwick and explicitly invoked the lessons of the human error studies.⁷⁰ In essence, they applied the approach of the interdisciplinary National Transportation Safety Board Go Teams⁷¹ that respond to air disasters:

⁶⁷ On the generally uninformative nature of wrongful conviction accounts, see Gross et al., *supra* note 4, at 533-41.

⁶⁸ INNOCENCE PROJECT, <http://www.innocenceproject.org/Content/205.php> (last visited Nov. 7, 2009).

⁶⁹ Mark R. Chassin & Elise C. Becher, *The Wrong Patient*, 136 ANNALS INTERNAL MED. 829-31 (2002).

⁷⁰ *Id.*

⁷¹ The National Transportation Safety Board dispatches a "Go Team" under the command of an investigator-in-charge to the scene of all accidents. In aviation, the team will include specialists in operations, structures, power plants, systems, air traffic control,

[T]his event shares many characteristics with other well-known and exhaustively researched calamities, such as the Challenger disaster, the Chernobyl nuclear reactor explosion, and the Bhopal chemical factory catastrophe. These events have been termed “organizational accidents” by psychologist and accident expert James Reason because they happen to complex, modern organizations, not to individuals. No single individual error is sufficiently grave to cause an organizational accident. The errors of many individuals (“active errors”) converge and interact with system weaknesses (“latent conditions”), increasing the likelihood that individual errors will do harm.⁷²

Using this perspective, Chassin and Becher reviewed the wrong patient episode, and discovered, reported, and analyzed at least seventeen distinct errors. For example, the patient’s face was draped so that the attending physicians could not see it; a resident left the lab assuming the attending had ordered the invasive procedure without telling him; conflicting charts were overlooked; contradictory patient stickers were ignored. But the crucial point for Chassin and Becher was that *no single one of the seventeen errors they catalogued could have caused the adverse event by itself.*⁷³

Their analysis disclosed not only mistakes by individual doctors and nurses, but abiding latent problems. Communications among the hospital staff members were terrible: “Physicians failed to communicate with nurses, attendings failed to communicate with residents and fellows, staff from one unit failed to communicate with those from others, and no one listened carefully to the patient.”⁷⁴ A patchwork of information mini-systems that did not speak to each other characterized the hospital; teams failed to function, and no one was surprised or bothered when the teams *did* fail to function because of “a culture of low expectations.”⁷⁵ There were gaps on charts, contradictions in conduct, and no legible entry in the charts explaining why the patient was in the hospital in the first place, but “[t]he culture of low expectations led [practitioners] to conclude that these red flags signified not unusual, worrisome harbingers but rather mundane repetitions of the poor communication to which they had become inured.”⁷⁶ Deviations from accepted practice had become “normal,” and, as a result, a disaster occurred.⁷⁷

The wrong patient review captured the fact that while every accident is unique, each depends in part on latent, abiding features that can encourage other accidents in the future. It indicated that the lessons of Chernobyl and

weather, and human performance. Examples of the reports compiled and issued by these teams can be found at: <http://www.nts.gov/ntsb/major.asp> (last visited Nov. 7, 2009).

⁷² Chassin & Becher, *supra* note 69, at 829.

⁷³ *Id.*

⁷⁴ *Id.* at 829-31.

⁷⁵ *Id.* at 829-30.

⁷⁶ *Id.*

⁷⁷ *Id.*

the Challenger could be applied to health care. It also indicated that although there is less complex technology implicated in delivering medicine to the wrong patient than in launching a doomed space shuttle, it does not follow—human beings being as various and extraordinary as they are—that there is less complexity in a human-dominated system. Nor does it follow that the human system’s complexity must be synonymous with permanent incomprehensible chaos. The wrong patient analysis uncovered things that could be changed.

IV. HARMFUL, HARMLESS, AND HELPFUL ERROR

The legal system, habituated to its tradition of dividing errors between harmful errors that threaten the reliability of verdicts and harmless errors that do not, has generally overlooked the potential of helpful errors. The contemporary criminal justice system lacks a routine for identifying and analyzing its unspectacular errors and a template for reporting their lessons.

When an eyewitness misidentification results in a “wrong man” conviction, an examination of the event from a perspective similar to the wrong patient review would reveal an organizational accident, which is constructed out of a constellation of individual errors and latent conditions. We would see that most wrongful convictions are caused, as Diane Vaughan said of the Challenger tragedy, by “a mistake embedded in the banalities of organizational life.”⁷⁸

This sort of examination does not happen often. Boston Police Commissioner Edward F. Davis, who as a Lowell detective investigated the cases against Denis Maher, provides a more characteristic specimen of current reactions to complex malfunctions. When asked recently whether he owed Maher an apology for the years he had spent in prison, Davis replied, “No, because I didn’t do anything wrong.”⁷⁹ Although Davis stressed that he did “feel terrible that this system did not work for [Maher],” in the absence of a bad apple, fatalism had taken over.⁸⁰ In Davis’s formulation, “this system” becomes a synonym for “God’s will.”

Although advocates for the system-variable perspective on eyewitness identifications avoid the bad apple mindset, they substitute a focus on isolated frailties in investigative techniques and do not examine the confluent, cascading failures in any individual case. The double-blind, sequential procedure may be a good thing, but the wrongful conviction of Denis Maher required much more than the use of sub-optimal eyewitness

⁷⁸ VAUGHAN, *supra* note 66, at xiv.

⁷⁹ Chassin & Becher, *supra* note 69, at 829-30.

⁸⁰ Maria Cramer, *Police Head Defends Role After Wrongful Imprisonment*, BOSTON GLOBE, Dec. 25, 2008, at B1.

procedures. Maher's wrongful conviction was an "organizational accident." How did it happen? Was there exculpatory physical evidence on the Denis Maher crime scenes that was not collected? Was that a training, supervision, or resource issue? All three? Did the first responders adequately communicate the full descriptions to the detectives? Were the eyewitnesses' memories protected from contamination? At the scenes? In their interviews? Was any of this documented for later use? Were contaminations dictated by training gaps, or simple facility shortages? Were the witnesses aware of each other's accounts? Was there a protocol for handling multiple eyewitnesses? How were the discrepancies in descriptions overlooked? Was "tunnel vision" an issue?⁸¹ Was "production pressure" (caseload levels and clearance rate evaluations) a contributor?⁸² Was there training in place to prevent tunnel vision? Did the prosecutors adequately challenge the police on alternative suspects? What allowed the actual perpetrator to escape? Did the defense investigation serve its purpose? Why not? Was it a performance issue? A training issue? A funding issue? A discovery issue? Did the trial process provide a clear picture of events? Were the jurors adequately instructed on the nature of memory evidence? Did small failures interact in unexpected and disastrous ways?⁸³

Who was responsible for the Denis Maher wrongful conviction? Everyone was involved, to one degree or another. The Maher conviction was an organizational accident: the police followed the witnesses, who chose the wrong man, but the redundancies of the trial and appellate systems also failed to prevent the initial mistake from taking effect.⁸⁴

It is not that this sort of review has never been suggested or performed in criminal justice systems.⁸⁵ For example, Peter Neufeld and Barry Scheck, the co-founders of the Innocence Project, began to argue in the earliest days of the DNA exonerations that a review function modeled on

⁸¹ See generally Susan Bandes, *Loyalty to One's Convictions: the Prosecutor and Tunnel Vision*, 49 HOW. L.J. 475 (2006); Keith A. Findley & Michael S. Scott, *The Multiple Dimensions of Tunnel Vision in Criminal Cases*, 2006 WIS. L. REV. 291.

⁸² See generally VAUGHAN, *supra* note 66; Diane Vaughan, *Theorizing Disaster: Analogy, Historical Ethnography, and the Challenger Accident*, 5 ETHNOGRAPHY 340-45 (2004) (describing need to study cultural and historical factors affecting organizational accidents).

⁸³ PERROW, *supra* note 2.

⁸⁴ Keith A. Findley, *Innocents at Risk: Adversary Imbalance, Forensic Science and the Search for the Truth*, 38 SETON HALL L. REV. 893 (2008).

⁸⁵ See, e.g., Keith A. Findley, *Learning from Our Mistakes: A Criminal Justice Commission to Study Wrongful Convictions*, 38 CAL. W. L. REV. 333 (2002); Robert Carl Scher & Jamie Sears, *Innocence Commissions: Due Process Remedies and Protection for the Innocent*, 13 CRITICAL CRIMINOLOGY 181 (2005).

the National Transportation Safety Board was needed. As an example, they noted that the government of Ontario investigated the events leading to the wrongful conviction of Guy Paul Morin.⁸⁶ That study fills two volumes, and the executive summary is forty pages long. It dissects the impacts of forensic scientists, jailhouse informers, and tunnel vision. Canada now produces intensive reviews of all known wrongful convictions. Other reviews of comparable depth have been produced from time to time in the aftermath of public scandals growing out of wrongful convictions. There has been high-quality investigative journalism that has attempted to wrestle with systemic failures⁸⁷ and book-length examinations of individual exonerees' experiences.⁸⁸

But these reviews were seen as heroic efforts to face up to extraordinary situations and to set a reform agenda. To some extent, this can be explained by the historical context. The DNA exonerations provided the first inarguable proofs of miscarriages of justice. Activists, appalled by the findings, tried to provoke action as quickly as possible. As a result, although reasonably detailed case studies are sometimes compiled, the first generation exoneration case accounts were distilled to identify one or two "causes" in each case, a prerequisite to prioritizing the "most frequent causes."⁸⁹

⁸⁶ Scheck & Neufeld, *supra* note 11, at 99-101; see FRED KAUFMAN, REPORT OF THE KAUFMAN COMMISSION ON PROCEEDINGS INVOLVING GUY PAUL MORIN (1998), available at <http://www.attorneygeneral.jus.gov.on.ca/english/about/pubs/morin/>.

⁸⁷ See, e.g., Maurice Possley, Steve Mills & Flynn McRoberts, *Scandal Troubles Even Elite Labs: Flawed Work, Resistance to Scrutiny Seen Across U.S.*, CHI. TRIB., Oct. 21., 2004, at C1; see also Terence M. Garrett, *The Waco, Texas, ATF Raid and the Challenger Launch Decision: Management, Judgment and the Knowledge Analytic*, 31 AM. REV. PUB. ADMIN. 66 (2001).

⁸⁸ MARGARET EDDS, AN EXPENDABLE MAN: THE NEAR-EXECUTION OF EARL WASHINGTON, JR., (2003); TIM JUNKIN, BLOODSWORTH: THE TRUE STORY OF THE FIRST DEATH ROW INMATE EXONERATED BY DNA (2004); RACE TO INJUSTICE: LESSONS LEARNED FROM THE DUKE LACROSSE RAPE CASE (Michael L. Siegel ed., 2009); Giovanna Shay, *What We Can Learn About Appeals from Mr. Tillman's Case: More Lessons from Another DNA Exoneration*, 77 U. CIN. L. REV. 1499 (2009).

⁸⁹ This path was essentially the one followed by the eyewitness identification procedural reforms, when mistaken identifications placed the greatest number of cases on the compiled list. Cf. N.Y. STATE BAR ASSOC., PRELIMINARY REPORT OF THE NEW YORK STATE BAR ASSOCIATION'S TASK FORCE ON WRONGFUL CONVICTIONS (2009), available at <http://www.nysba.org>. The Task Force gave individual attention to each case of wrongful conviction, but then published a ranked list of "specific causes linked to wrongful convictions," rather than the detailed case studies themselves, relying instead on synopses. *Id.* at 7. The synopses did frequently note the presence of more than one "cause," and the Preliminary Report did accompany its chart ranked causes with the observation that most wrongful convictions resulted from multiple causes. For a discussion of mechanisms for aggregation of errors in order to enable improvements, see Garrett, *supra* note 46, at 383.

Once law reform is chosen as the goal, it exerts a pull toward problems that law reform might effectively address and narratives that have some utility in the legislative advocacy process.⁹⁰ The end product is a collection of case features that *correlate* with wrongful conviction rather than a close analysis of the causes, active and latent, of specific events. The medical reformers would understand the energy that went into these efforts because similar efforts dot the history of medicine from the time of Ernest Codman. But they would also see this approach as futile, or even worse, as perpetuating a pattern of staying one catastrophe behind. As Lucien Leape wrote:

Efficient, routine identification of errors needs to be part of hospital practice, as does routine investigation of all errors that cause injuries. *The emphasis is on "routine."* Only when errors are accepted as an inevitable, although manageable, part of everyday practice will it be possible for hospital personnel to shift from a punitive to a creative frame of mind that seeks out and identifies the underlying system failures.⁹¹

Contemporary medicine treats errors as "sentinel events": important opportunities to illuminate hidden flaws.

The practice of criminal justice produces sentinel events on a daily basis. Wrongful convictions, after all, are not the only examples of error in criminal justice. To begin with, every wrongful conviction is also a wrongful acquittal because an actual rapist or killer goes free. Other errors can lead to the same result: loss or contamination of evidence, failures to follow legal rules that result in suppression of evidence, and in all of these examples the perpetrator goes free. Why? A misdemeanor defendant was detained pretrial and forgotten until six months beyond the expiration of his maximum sentence.⁹² How did that happen? A dangerous inmate was released after "wrapping up" his sentence, but he should have been held because three new warrants and detainers should have been lodged.⁹³ When he then kills two innocent people,⁹⁴ was the cause a bad apple's negligence or a system failure? And how should we treat a domestic homicide when the troubled family was on the system's screen, but no effective intervention occurred?⁹⁵ Where was the mistake when a cooperating

⁹⁰ See generally Michael J. Saks et al., *Toward a Model Act for the Prevention and Remedy of Erroneous Convictions*, 35 NEW ENG. L. REV. 669, 670 (2001).

⁹¹ Leape, *supra* note 1, at 1856 (emphasis added).

⁹² Thomas Farragher, *Sentence Tracking System Is Tossed Out*, BOSTON GLOBE, Nov. 28, 2007, at B1.

⁹³ Shelley Murphy & Megan Woolhaus, *Worcester DA Knew Killer Went to Wash*, BOSTON GLOBE, Nov. 28, 2007, at B1.

⁹⁴ *Id.*

⁹⁵ Kevin Flynn, *Jump in Brooklyn's Murder Rate Spurs Prosecutor to Form Study Panel*, N.Y. TIMES, July 26, 1999, at B1.

witness is killed after his name and whereabouts are leaked before trial?⁹⁶ All of these errors might teach us something about latent conditions that are “accidents waiting to happen.”

The reservoir of helpful errors is expanded if we recognize that we can learn—and probably learn more and better—by paying careful attention to “near misses” that were caught (this time) by special alertness or good luck.

If Denis Maher’s case were to arise today, it would follow the same course—up to a point. An eyewitness would identify the suspect and biological evidence would be collected. In all likelihood, if Maher were to be arrested today, he would still have been the primary suspect at the beginning of the investigation. But when the DNA lab would release its report (generally a matter of weeks, sometimes months), Maher would be exonerated.

Today’s Maher case, in other words, would be a “near miss” from which we might learn many lessons: lessons useful in future sexual assault cases, but also useful in cases in which there would be no DNA safety net (such as robbery).⁹⁷ Police likely would damage their pursuit of the *right* man by mistakenly focusing on the wrong man while waiting for the lab results. Many lessons could be learned from this damage. Why did they have to wait so long for the lab results? Was the exclusive focus on Maher too early? Too exclusive? Did someone decide to deluge the lab with property crime evidence? Who? Was the lab short-staffed, or were the staff poorly trained? Did the defense lawyer sit on an alibi that could have been shared? Why? Would a quicker, cheaper, preliminary analysis, designed to rule out innocent suspects rather than to generate evidence for trial have eliminated Maher earlier and put investigators back on the right trail sooner?

It is not unreasonable for a front-line detective like Edward Davis to bridle when he is asked to “take the fall” for a wrongful conviction such as Maher’s; front-line practitioners are painfully vulnerable to officious second-guessing. As Charles Perrow points out:

[V]irtually every system we will examine places “operator error” high on its list of causal factors—generally about 60 to 80 percent of accidents are attributed to this

⁹⁶ See generally PETER FINN & KERRY MURPHY HEALEY, PREVENTING GANG- AND DRUG-RELATED INTIMIDATION (1996).

⁹⁷ Routine “near miss” analyses would also address the fear of commentators that, as DNA testing becomes routine, known wrongful convictions will disappear, and the study of error therefore will be starved. See, e.g., Findley, *supra* note 85, at 353. In fact, most contemporary eyewitness, false confession, and jail house snitch cases in which there is DNA recovered would still provide an opportunity—well-after “prime suspect” status is awarded to a defendant—for any DNA results to contradict the first determination, illuminating potential errors in non-DNA cases.

factor. But if, as we shall see time and again, the operator is confronted by unexpected and usually mysterious interactions between failures, saying that he should have zigged instead of zagged is possible only after the fact. Before the accident no one could know what was going on and what should have been done.⁹⁸

But the way out of that trap is not a wistful shrug; it is an assessment of the role played by conditions latent in the system and the mechanisms by which they take effect.⁹⁹

The recognition that wrongful convictions—among other criminal justice errors—are complex organizational accidents can focus attention on the barriers that hamper the front-line troops who are usually singled out for bad apple treatment. Accidents are caused not only by the active errors of people at the sharp end of the system—pilots, operators, and doctors; eyewitnesses, cops, lawyers, and jurors—but also by the mistakes of people far from the scene—managers, designers, accountants, legislators, policy-makers, funders, and appellate courts.¹⁰⁰

V. THE CLEARINGHOUSE FOR HELPFUL ERRORS

The criminal justice system needs a workable facility to collect and disseminate detailed, reliable, factual accounts of helpful errors. Aviation has found regular vehicles for communicating the facts of its disasters and near-misses through NTSB investigations, the internet, and *Flying* magazine. Medicine has done the same through journals such as *Lancet* and *Annals of Internal Medicine*. The Innocence Project has supported an improvised error-tracking function as an appendage to its casework operations for many years. Criminal justice journals could easily provide an analogous clearinghouse for the voluntary reporting of errors and the sharing of lessons distilled from those errors.

But if you build it, will they come? Of course there will be jurisdictions and agencies that will decline an error review opportunity. Some will never participate; others will only participate in scattered instances. With this reality in mind, Scheck and Neufeld have argued that innocence commissions armed with coercive powers (including subpoena power) would be indispensable.¹⁰¹ But there is no reason to forestall voluntary efforts while awaiting the arrival of a mandatory innocence commission with subpoena power. There is no guarantee that the legislative action necessary to create such bodies will be forthcoming any

⁹⁸ PERROW, *supra* note 2, at 9.

⁹⁹ See Garrett, *supra* note 46, at 395-400 (discussing the increasing “systemization” of criminal processes).

¹⁰⁰ See generally Douglas McCarthy & David Blumenthal, *Stories from the Sharp End: Case Studies in Safety Improvement*, 84 MILBANK Q. 165 (2006).

¹⁰¹ Scheck & Neufeld, *supra* note 11.

time soon. Even if versions of such bodies are created, it is not clear that they will be either individually useful or generally compatible since they will be the products of uncoordinated and partisan state legislative processes. Besides, industrial and medical quality experts would argue that the more the process appears to its nervous participants to be a voluntary professional inquiry rather than a blame-oriented inquisition, the better the analyses and educational products are likely to be. Faced with an end-of-process coercive inquisition, the players generally devote energy to trying to game the inquisition that could be better spent on trying to understand the incident's complexities and ambiguities.

Routine error reporting will be very difficult to impose from the outside on the criminal justice system's practitioners.¹⁰² The system's rank and file have been taught throughout their careers that silence on these matters is usually the safest policy. They are a resourceful group; they have seen the frank admission of error used to burn colleagues, and they have developed many tactics and strategies for avoiding comment on imperfections. Any movement in the direction of systematic attention to error will initially provoke calls to "let sleeping dogs lie" or, failing that, to "keep this in house." Even very careful practitioners will be tempted to say that tomorrow will inevitably bring new cases and we will just have to try harder tomorrow.

Despite this reflexive skittishness, collecting evidence to fuel self-criticism is not alien to contemporary law enforcement. Many police leaders value evidence of sub-optimal performance (such as Commissioner William Bratton's well-known advocacy of the COMPSTAT¹⁰³ program) and use error data to amend practice in self-consciously designed "teaching departments" and "learning organizations."¹⁰⁴ Police leaders on the cutting

¹⁰² In medicine, where accreditation bodies wield enormous leverage, this is not the case. Mandatory reporting—at least of harmful errors—is not uncommon, although the consensus is that under-reporting persists. TO ERR IS HUMAN, *supra* note 55, at 8-10. In criminal justice there has been a movement toward accreditation of individual component agencies (police or corrections departments, or crime labs), but no suggestion that accreditation of a jurisdiction's system, *as system*, should or could be attempted. That is not to say that it would be a bad idea, and the steps suggested in this Article might prepare the ground for such an effort. The National Institute of Medicine recommended mandatory reporting of harmful errors and the creation of vehicles guaranteeing confidentiality to encourage the reporting of errors that did not result in harm to patients. *Id.* Something of the kind is not out of the question in criminal justice, but it would seem appropriate to allow a period of experimentation before imposing it on a broad scale.

¹⁰³ WILLIAM BRATTON, TURNAROUND 233-40 (1998). See generally JON M. SHANE, WHAT EVERY CHIEF EXECUTIVE SHOULD KNOW: USING DATA TO MEASURE POLICE PERFORMANCE (2007).

¹⁰⁴ See, e.g., William A. Geller, *Suppose We Were Really Serious About Police Departments Becoming "Learning Organizations"?*, 234 NAT'L INST. JUST. J. 2 (1997).

edge are convinced that robust feedback loops—operating in as close to real time as possible—are essential to progress. Line cops understand and accept that, every time they fire their guns, the decision will be reviewed. Although prosecutors are generally less involved in day-to-day tracking of performance, in the jurisdictions where the Green Book spurred reforms, prosecutors always consented to (and frequently led) the reform efforts. The reasonable expectation is that police, prosecutors, defenders, and judges, although they—like doctors, hospital administrators, risk managers, and health insurance executives—see error through various lenses, all share a desire to stamp it out and will take steps to analyze it when a framework for doing so is presented to them.

The practitioners themselves are the best gatekeepers for the process, and many of their worries can be answered by formally assigning them the gatekeeper's responsibilities. Manageable limits could be placed on the number of errors under study—excluding errors that threaten to waste time and effort—by requiring authorized professional practitioners to certify an event as salient when they nominate it for study and report. The process of collecting and reporting errors has no chance of succeeding if it appears to be an end in itself or—even worse—a suicidal effort to fuel uncomprehending outsider attacks.

The practitioners' first requirement for participating in the process will be: "Don't waste our time." There are incidents—a shooting accident on the police target range, for example—where the "keep it in house" strategy works perfectly well,¹⁰⁵ and those incidents are unlikely candidates for clearinghouse participation. The "all-stakeholders" approach will offer no advantage over the traditional approach in those cases, and those cases can be quickly discarded.

What will be included for examination will be as important to winning practitioners' assent to the process as what will be off-limits. Every stakeholder group should be assured that it is licensed to nominate any error it considers worthy of all-stakeholder analysis and dissemination. The prosecutor who was left holding the bag in the courtroom should have the opportunity to suggest an "organizational accident" review of police and lab procedures;¹⁰⁶ the crime lab director, vulnerable to scapegoating, should have access to the same forum. The defense lawyer who uncovered the alternative suspect who turned out to be the perpetrator, or the alibi that was ignored, could nominate for study the question of how the system was

¹⁰⁵ See Michael T. Charles, *Accidental Shooting: An Analysis*, 8 J. CONTINGENCIES & CRISIS MGMT. 151 (2000).

¹⁰⁶ See Sylvia Moreno, *New Prosecutor Revisits Justice in Dallas*, WASH. POST, Mar. 5, 2007, at A4.

misled until its final phase. The trial judge, presented with a mess at the terminal point in the process, could ask that a study group be convened to explain how that situation was created. Victims' services professionals could call attention to instances of mishandling of victims' safety issues. No one would have a veto, but without a consensus there would be little value in reviewing the nominated error, and the process could be terminated. By definition, a process proceeding without broad participation would be a waste of time in any event. The only price for embarking on the process would be an agreement to see it to its end: to follow the facts of an individual event wherever they lead, while still retaining autonomy regarding any more general recommendations or reforms that might be suggested by whatever facts are found.

Front-line participants will quickly lose patience with any review that applies rarefied national standards without considering local resource problems, legal constraints, and institutional histories. Obviously, most errors are local, and the study of those errors will require a fine-grained knowledge of local systems, practices, and challenges.¹⁰⁷ Practitioners may be happy to share the lessons of their findings with others, but their first priority will be to learn something practical about the frailties in their own environments and to repair them.

If these locally focused analyses are pursued according to a national protocol, reported within a uniform national template, and recorded in a national clearinghouse, the accuracy and the teaching power of local findings can be enhanced by way of national access to these error reports. If the review function supports a modest cadre of national or regional experts (as does the NTSB Go Teams model), the presence of a few outside representatives on a local error review team could provide process consistency across local reviews and supplementary subject-matter expertise which might be unavailable to local practitioners. Practitioners will not feel that they have been wasting their time if they know that they are being saved from reinventing a wheel that has been invented and tested in other jurisdictions. At the same time, a neutral outsider with a national or regional perspective might help insure against another potential waste of time by warding off any drift toward an incestuous insiders' process devoted to either mutually protective logrolling or to finger-pointing aimed at paying off ancient, local inter-agency grudges.

¹⁰⁷ Some events have causes that are so local and idiosyncratic—for example, “Judge Doyle was off his medication”—that they can quickly be consigned to the “waste of time” bin. Other events cannot be understood without examining historical and cultural contexts. See Vaughan, *supra* note 82, at 340-45 (describing the need to study cultural and historical factors affecting organizational accidents).

The second demand of criminal justice leaders will be: “Don’t embarrass my people.” Practitioner leaders will not commit professional suicide or casually sacrifice subordinates to a learning-from-error process. Even so, because they will recognize that the particularly humiliating harmful errors—the apparent scandals that any agency would most like to keep “in house”—are the errors least likely to stay in house, they will accept a measure of discomfort as the price of participation. Since the spectacular disasters are likely to be exposed, an examination by a sober professional panel is an attractive substitute for the pillory of “gotcha” journalism, the accusatorial atmosphere of a grand jury, an internal affairs proceeding, a grandstanding plaintiff’s lawyer in a civil law suit, or a finger-pointing competition among the implicated agencies. At worst, a careful, dry—even boring—professional review of an error would be a useful non-blaming supplement to the accusatory brickbats of journalists and perennial adversaries.

The extra effort that clearinghouse participation will require from practitioners is justified for cases of organizational accidents that have roots in more than one agency or individual performance and cannot be scrutinized within any single “house.”¹⁰⁸ Police commissioners who send mistaken officer-involved shootings *only* to internal affairs know that they will be accused of whitewash. The prosecutors who volunteer an in-house investigation and then discover a problem with the police force know that they will be accused of passing the buck which will simultaneously jeopardize an ongoing working relationship.

In these situations, an organizational accident review, while it will not eliminate the potential for embarrassment, will at least ensure that embarrassment is not gratuitous and that it is appropriately modulated and shared. Regularizing error review within a formal clearinghouse program can actually enhance protections against the personal humiliation of individual staff members. Complete peer review confidentiality for reports—a feature once advocated (at least for errors that did not harm patients) by the National Institute of Medicine¹⁰⁹—may be out of reach. But measures to limit access to professional colleagues and to exclude

¹⁰⁸ Cf. Berwick, *supra* note 42, at 55 (noting the necessity for cross-department interactions).

¹⁰⁹ TO ERR IS HUMAN, *supra* note 55, at 8-10. Embarrassment will remain an inhibiting factor even though most criminal justice actors enjoy a broad immunity from tort liability that medical personnel can only gaze at in wonder. See Margaret Z. Johns, *Reconsidering Absolute Prosecutorial Immunity*, 2005 BYU. L. REV. 53; see also Van de Kamp v. Goldstein, 129 S. Ct. 855 (2009) (holding that prosecutors enjoy absolute immunity from suit over failure to train regarding documentation and constitutionally required disclosure of exculpatory evidence). This grant of the benefits of immunity from tort liability could be conditioned on paying the price of cooperation with post-error analysis investigations.

casual sightseers are still possible because the formality of the process will allow for the identification of cases in which confidentiality does not conflict with other values. Moreover, drafting standards can offer some measure of anonymity—for example, “the ranking official on the scene” (rather than “Captain Shane”).¹¹⁰ When a report’s audience is not restricted, the investigators can supply context that minimizes the subjects’ embarrassment without adversely affecting the utility of the report. Agency heads or elected officials who submit to NTSB-style investigations can fight on favorable grounds by saying, “let’s wait for the report.” Once a known error puts a public reputation for infallibility out of reach, the public perception that a fallible agency honestly faced and rectified its failure can become a crucial bulwark of the agency’s stature and the system’s legitimacy.¹¹¹

Slightly different considerations apply when the learning-from-error inquiry is extended to near-miss events. In aviation, a pilot who promptly reports a near miss is immunized; the event is studied confidentially, and its lessons are disseminated. In medicine, the Joint Commission, which accredits hospitals, maintains a registry of “sentinel events” including not only incidents of injury, but also of high risk of injury. The criminal justice near miss generates grim sighs of relief, but little incentive for instigating an analysis.

Of course, the creation of a “safe harbor” for near-miss criminal justice events will not be a matter of simply turning a switch. Inertia, self-protective bureaucratic reflexes, and the unfamiliarity of the effort will all present challenges. But the aviation and medicine experiences indicate that it will not be an impossible task either.¹¹²

In many ways the near miss presents the criminal justice practitioners with a tempting target. Simple proximity in time to the events to be investigated enriches near-miss investigations. Documents will be easier to find, participants will be easier to interview; a full, reliable record is much easier to construct than in the review of a thirty-year-old wrongful conviction. Because the events examined are more recent, the lessons they teach will be more useful. Time will not be wasted in dissecting procedures and technologies that have been out of use for twenty years.

¹¹⁰ This is the practice followed in NTSB post-crash reports. See Chassin & Becher, *supra* note 69.

¹¹¹ Conley, *supra* note 13 (noting an elected district attorney’s views on the need for remedial measures).

¹¹² There are approximately five thousand aviation near misses reported annually. The FAA Aviation Safety Action Program is described at <http://www.faa.gov/about/initiatives/asap/> (last visited Nov. 27, 2009).

Cases where something went right can shed important light on cases where everything went wrong. Typically, a near miss provokes a lower level of defensiveness among the operators. There is a sense in which a near miss is a success story. Since the near miss caused no lasting harm, there should be fewer worries that an error report will trigger financial liability, harm a reputation, or cheat the victim out of compensation if it affords confidentiality to the operators. The inhibiting effect of activists exploiting the “myth of innocence,”¹¹³ whom law enforcement veterans see as sanctimonious second-guessers, will be muted.

A coherent program of learning from error that includes the evaluation of near misses offers its rewards both within local systems and across scattered systems. The radical fragmentation of the criminal justice system in the United States into a myriad of federal, state, and local systems cannot be eliminated; there will never be a United States version of Britain’s Home Office. But a common national template for error review, enacted locally and informed and challenged by diverse local experiences, can substantially mitigate the problem. Reading of a distant system’s experience of completed accidents can alert currently isolated practitioners to the operation of dangerous latent features that are present in their own local systems. Reading studies of remote near misses can reveal both those dangerous latent features and potential fail-safe devices or procedures that are *not* present locally. It can counteract the endemic tendency of today’s best practice standards, designed to provide a minimum floor for performance, to calcify into a ceiling that blocks further improvements.¹¹⁴ A compact national clearinghouse dedicated to providing technical and process support to the local study of criminal justice error through a small cadre of experts and veteran practitioners could conveniently be housed under the National Institute of Justice¹¹⁵ or funded (at least initially) with non-profit resources.

¹¹³ See generally Marquis, *supra* note 8.

¹¹⁴ Cf. Berwick, *supra* note 42, at 56.

¹¹⁵ The National Academy of Science’s recent suggestion of the formation of a new National Institute of Forensic Science, if adopted, would also provide a highly suitable home for an error clearinghouse. COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCI. CMTY., NAT’L RES. COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 20-22 (2009). The National Institute of Justice, having convened many “technical working groups” to deal with specific areas of “best practice,” might now consider convening another to develop core “best practices” for a consistent error-review process. The course followed by the medical reformers (begin with the development of small “communities of insight,” follow that phase with a broad nationally representative conference, and then fund a series of national “demonstration projects”) would seem readily adapted to criminal justice reform. See generally KENNEY, *supra* note 3.

If the only impact of the circulation of the organizational accident reports is to provide practitioners in all roles with the regular opportunity to ask, “Could this have happened to *us*?” the process will have value. Even so, everyone in the criminal system has plenty of work to do already, and the question remains whether the organizational accident model of error review will generate positive change, beyond the circulation of damage reports, that repays the investment of effort.

The experience of medicine and aviation demonstrates that the broad participation that organizational accident reporting requires will produce its own benefits, distinct from value of the content of any reports, and potentially more important, too. The *practice* of generating organizational error analysis can place local criminal justice systems on the threshold of a fundamental cultural change. Working at organizational accident analysis can steer the criminal justice system toward the ideal of continuous quality improvement and fulfill the preconditions for the inception of the “culture of safety” in criminal justice that medical reformers have made giant strides toward adopting.

VI. TOWARD THE CRIMINAL JUSTICE TEACHING HOSPITALS

When a rash of highly publicized reports of medical errors and patient deaths (analogous to the sudden tide of DNA exonerations let loose in the Green Book) alarmed the health care community in the 1990s, there was an obvious place to “do reform”: the traditional teaching hospital.¹¹⁶ The medical quality reformers had more success in hospitals than in any other arena.¹¹⁷ The structure of a teaching hospital—in which academic divisions, research components, clinical departments, specialties, and administrative support units are forced to interact—could respond fairly quickly to physician-reformers.

These reformers advocated that hospitals mobilize the findings of “human factors” researchers like James Reason. Reason argued that errors are inevitable in human performance, and that the best path toward reliable performance in complex organizations is the creation of a “culture of safety.”¹¹⁸ According to Reason, a culture of safety: (1) is informed about current knowledge in its fields; (2) promotes the reporting of errors and near misses; (3) creates an atmosphere of trust in which people are encouraged to report safety-related information; (4) remains flexible in

¹¹⁶ KENNEY, *supra* note 3, at 55-59.

¹¹⁷ David Blumenthal & Charles M. Kilo, *A Report Card on Continuous Quality Improvement*, 76 MILBANK Q. 640-42 (1998) (describing the mixed record of, and structural barriers to, continuous quality improvement efforts).

¹¹⁸ See generally JAMES REASON, *MANAGING THE RISKS OF ORGANIZATIONAL ACCIDENTS* (1997).

adapting to changing demands (by, for example, shifting from steeply hierarchical modes into “flatter” team-oriented professional structures); and (5) is willing and able to learn about and adjust the functioning of its safety systems.¹¹⁹ Reformers challenged practitioners to establish a culture of safety in medicine by adopting the ideal of “continuous quality improvement,” which was advocated by industrial expert W. Edwards Deming¹²⁰ and was central to the post-World War II revival of Japanese industrial corporations. In a movement that followed Deming’s principles (captured in *To Err Is Human: Building a Quality Health Care System* and *Crossing the Quality Chasm*),¹²¹ the reformers advocated continuous quality improvement rather than end-of-process inspection as an organizing principle.

Donald Berwick challenged the health care system to save one hundred thousand patients’ lives in eighteen months by applying six simple, evidence-based practices.¹²² These practices had been derived from a process employing Deming’s approach to learning from the organizational accident perspective on error, involving every member of the teams charged with a patient’s care. For example, a janitor in a Pennsylvania hospital’s intensive care unit discovered the cause of a mysterious outbreak of central line infections when he reported that “ambu bags” carrying plentiful bacteria were often left lying uncollected in the ICU.¹²³ Berwick’s Institute for Health Improvement enrolled 3,000 hospitals in the effort and saved over 120,000 patients’ lives, surpassing its goal.¹²⁴ The effort to continuously identify errors and work in inclusive teams built a culture of safety that, in turn, had nourished the ideal of continuous quality improvement. As two leading medical commentators noted, “[a]

¹¹⁹ *Id.*; see also McCarthy & Blumenthal, *supra* note 100, at 167.

¹²⁰ See generally W. EDWARDS DEMING, *OUT OF THE CRISIS* (1982). On the medical reformers’ adoption of Deming, see KENNEY, *supra* note 3, at 27-36; Berwick, *supra* note 42; Paul Batalden, *Organization Wide Quality Improvement in Healthcare*, in *THE TEXTBOOK OF TOTAL QUALITY IN HEALTHCARE* 60, 60 (A.F. Al-Assaf & June A. Schmele eds., 1993).

¹²¹ *TO ERR IS HUMAN*, *supra* note 46; INST. OF MED., *CROSSING THE QUALITY CHASM: A NEW HEALTH SYSTEM FOR THE 21ST CENTURY* (2001).

¹²² The six measures were: deploying rapid response teams at the first sign of patient declines, delivering evidence-based care (including aspirin) to heart attack patients, developing accurate lists of patient medications, preventing intravenous line infections (by, for example, hand-washing), preventing surgical site infections with three steps (including pre-surgery administration of antibiotics), and preventing pneumonia in intensive care units with simple measures (such as elevating the head of beds to thirty degrees). Liz Kowalczyk, *Nonprofit Launches Plan to Reduce Medical Errors*, *BOSTON GLOBE*, Dec. 15, 2004, at A1.

¹²³ KENNEY, *supra* note 3, at 121-25.

¹²⁴ *Id.* at 270-72.

paradoxical insight is that the adoption of specific improvements both furthers—and is furthered by—organization-wide cultural change.”¹²⁵

There are no bricks-and-mortar teaching hospitals in criminal justice that would house the system’s full range of practitioners—first responders, police, prosecutors, scientists, policy-makers, corrections officials, defense lawyers, and judges—but bricks and mortar are the simple part of the problem. Operating rooms and patient beds are not needed here, only meeting rooms. The judiciary usually has a wealth of square footage at its disposal. Every jurisdiction has a law school or a school of criminal justice that can be enlisted to provide a neutral meeting ground, as well as substantive faculty expertise and administrative support services.

The criminal justice system’s fractured structure presents a more serious challenge. Responsibility is divided across many agencies, each having a distinct bureaucratic identity, history, and ideology. Several of these agencies are set in explicitly adversarial relationships. A great deal of the criminal practitioners’ work (especially the legal practitioners’) is spent in a zero-sum atmosphere dominated by the question: “Who wins?”¹²⁶

Still, it is possible to identify a potential durable web of relationships and activities in criminal justice that might fill the teaching hospital’s role. In fact, something like a “virtual teaching hospital” may already exist, in embryo, in the criminal justice system. The “wrong man” reform controversies provide a lens through which this latent structure can be seen.

Relative to traditional lineups, the double-blind, sequential identification technique yields fewer “false positives”—identifications of known innocents—but it also yields more “false negatives”—failures to identify the perpetrator when the perpetrator is present in the lineup.¹²⁷ Arguments now rage over whether the traditional or double-blind, sequential method of lineup administration is “better.”

In criminal justice discourse this question is usually answered according to pre-existing feelings about Blackstone’s ratio of ten guilty men escaping being better than one innocent man convicted. But think of a lineup or a photo array as a screening test for guilt, and then compare its treatment to the reception new proposed screening tests for, say, prostate

¹²⁵ McCarthy & Blumenthal, *supra* note 100, at 196.

¹²⁶ See generally John Griffiths, *Ideology in Criminal Procedure or a Third “Model” of the Criminal Process*, 79 YALE L.J. 359 (1972) (describing an endemic “battle model” ideology of criminal process).

¹²⁷ As Gary Wells himself acknowledges, “[t]his pattern of results is a classic trade-off in which reducing the rate of one type of error, mistaken identifications, can increase the rate of another type of error, failure to identify the culprit.” Jennifer Emily, *Photo Lineup Study Is on Deck: Method in which Witness Sees Pictures One at a Time to Be Considered*, DALLAS MORNING NEWS, Jan. 1, 2009, at 1A.

cancer or breast cancer would receive in the teaching hospital. In a modern hospital, the proposed innovation will be evaluated with the same system-consciousness that characterizes organizational accident reviews of medical error. Propose a new screening test in a hospital setting and the inquiry will begin with how *the rest* of the system might absorb this new “system variable.” The discussion will include the researchers who had developed the new test, along with the clinicians who must use it on the patient floors, the administrators in the office wing who are bound to consider the price, and the risk managers who must weigh the costs and benefits of using it. With any luck, there will be someone from the cancer survivors’ community to represent the patients’ viewpoint. Too many false positives might mean one thing for prostate cancer, where surgical treatment results in impotence and incontinence, and where the cancer is often a slow-growing one that many men die with, but do not die of.¹²⁸ Too many false negatives might mean something quite different for breast cancer patients where early discovery, effective treatment plans, and prognosis are closely related. Are there redundant tests that might be used? Can a science-based scheme of watchful waiting be developed? Are there cheaper ways to assess the situation?

In a teaching hospital, no researcher’s novel screening test (that is, no new system variable) would be proclaimed “better” than an existing test without first carefully assessing how the rest of the system would adjust to its adoption or compensate for its rejection. The new test might be seen as an improved part for the system, but everyone would be mindful of Donald Berwick’s observation that “optimizing *parts* is not a good route to system excellence.”¹²⁹ The same should be true in criminal justice. Which identification procedure is better depends to a significant degree on the uses to which the balance of the system puts the results. The higher number of false positives generated by traditional lineups may not be a bad thing if the rest of the system treats the lineups as a preliminary screening-only test that triggers a conscientious campaign to “rule out” the suspect in the medical sense of trying carefully to eliminate the suspect’s guilt as a possibility and dedicates the time and energy that a “rule out” approach would demand. The same higher number of false positives will be a recipe for disaster if the system (because of, for example, resource scarcity and caseload pressures)

¹²⁸ Gina Kolata, *Studies Show Prostate Test Saves Few Lives*, N.Y. TIMES, Mar. 19, 2009, at A1. The eyewitness “false positive” has an additional detrimental consequence: a false positive “burns” the witness and will undercut the credibility of any later testimony by the mistaken witness (and, less directly, any other witness) if and when someone correctly identifies the actual perpetrator.

¹²⁹ ATUL GAWANDE, THE CHECKLIST MANIFESTO: HOW TO GET THINGS RIGHT 185 (2009) (emphasis added).

must treat every lineup as a conclusive diagnostic test that in effect screens *in* the defendant and triggers only a hunt for supplementary corroboration aimed at courtroom persuasiveness rather than objective investigative fact finding.

Two projects have tested the double-blind, sequential reform's impact on the criminal justice system in the field. Both projects aimed to test potential applications of the lessons learned from eyewitness error, and both explored the effect of a new system variable on an existing system. In other words, both attempted to evaluate the impact of a new system variable practice in much the same way that a teaching hospital might evaluate a new cancer screening test. But the two projects took radically different paths toward the goal they shared.

The first path was followed by a field test of eyewitness identification procedures conducted by the general counsel of the Chicago Police at the direction of the Illinois legislature.¹³⁰ It terminated in a dead end. The Illinois Pilot Program (Pilot Program) began in a crisis atmosphere. It was prompted by a cataclysmic series of highly publicized exonerations in Illinois that produced a Governor's Commission on Capital Punishment. The Commission recommended the adoption of double-blind, sequential lineup practices in the state. Responding to disquiet about that recommendation's possible effect on existing practice, the Illinois legislature mandated that the efficacy of the double-blind, sequential method be compared to traditional methods in a field study, "designed to elicit information for comparative evaluation purposes, and . . . consistent with objective scientific research methodology."¹³¹

The general counsel for the Chicago Police Department assumed responsibility for this study. Although social scientists were engaged to analyze the study's results, its design and execution were maintained under exclusive police control.¹³² The general counsel (who was not trained in social science) devised a program of field tests, selected two other municipalities to participate in the program, chose which techniques would be employed in which jurisdictions, instructed the officers involved, compiled the data, and announced the study's findings. The results were

¹³⁰ 725 ILL. COMP. STAT. 5/107A-10 (2006); *see also* SHERI MECKLENBURG, ILL. STATE POLICE, REPORT TO THE LEGISLATURE OF THE STATE OF ILLINOIS: THE ILLINOIS PILOT PROGRAM ON SEQUENTIAL DOUBLE-BLIND IDENTIFICATION PROCEDURES (2006), *available at* www.chicagopolice.org/IL%20Pilot%20on%20Eyewitness%20ID.pdf (describing the process).

¹³¹ *Id.*

¹³² Roy S. Malpass, *Notes on the Illinois Pilot Program on Sequential Double-Blind Identification Procedures*, 11 PUB. INT. L. REP. 5 (2006).

widely publicized.¹³³ According to the initial press accounts, “[t]he study, the first to do a real-life comparison of the old and new methods, found that the new lineups made witnesses less likely to choose anyone. When they did pick a suspect, they were more likely to choose an innocent person.”¹³⁴ This assessment of the Pilot Program fueled a counter-attack against the gathering momentum of the double-blind, sequential reform initiatives advocated by Gary Wells and his allies.¹³⁵

But an evaluation of the study by a “blue ribbon” group of social scientists (with no prior involvement in “wrong man” issues) concluded that the Pilot Program proved nothing at all.¹³⁶ According to the group, the Pilot Program changed two variables—blind/not blind and simultaneous/sequential—at the same time, which created a “confound” with “devastating consequences for assessing the real-world implications . . . [and] guaranteed that most outcomes would be difficult or impossible to interpret.”¹³⁷ Other critics complained that the Pilot Program, because it made no effort to employ random assignment, violated first principles of the scientific method and was doomed to worthlessness before it began.¹³⁸ Interpretation was further hindered by the Chicago Police Department’s determination that much of the data would be withheld as confidential (varying from scientific conventions).¹³⁹ The vigor with which the Chicago Police Department seized on and publicized their conclusory findings, and the eagerness with which other opponents of Gary Wells’s reforms propagated the Illinois results provoked a backlash from Wells’s allies. Some of the Illinois data needed explaining: for example, in two of the three jurisdictions, fillers were identified in none of the traditional lineups.¹⁴⁰ One defense lawyer pointed out that the Illinois study, because it

¹³³ Kate Zernike, *Study Fuels Debate over Police Lineups*, N.Y. TIMES, Apr. 16, 2006, at A1.

¹³⁴ *Id.*

¹³⁵ See, e.g., Sheri H. Mecklenburg et al., *Eyewitness Identification: What Chiefs Need to Know Now*, POLICE CHIEF, Oct. 2008, at 68.

¹³⁶ Daniel L. Schacter, et al., *Policy Forum: Studying Eyewitness Investigations in the Field*, 32 LAW & HUM. BEHAV. 3, 4 (2007).

¹³⁷ *Id.* at 4.

¹³⁸ Timothy O’Toole, *What’s the Matter with Illinois? How an Opportunity Was Squandered to Conduct an Important Study on Eyewitness Identification Procedures*, CHAMPION, Aug. 2006, at 18, 20.

¹³⁹ This issue is, as of the date of publication, the subject of a lawsuit initiated by the National Legal Aid and Defender Association and brought pursuant to the Freedom of Information Act. The defense bar had not previously been involved in developing the Pilot Program. Nat’l Ass’n of Crim. Def. Lawyers v. Chi. Police Dep’t, No. 1-08-2073 (Ill. App. Ct. filed 2008).

¹⁴⁰ “Filler” identifications in earlier field studies had typically been in the 20% range. See, e.g., Bruce W. Behrman & Sherrie L. Davey, *Eyewitness Identification in Actual*

chose the number of *suspect* identifications (not perpetrator identifications) as its benchmark, created “a huge risk that the benchmark inflat[e] the perceived reliability of the most suggestive procedures, rather than the most accurate ones.”¹⁴¹ In this view, the Pilot Program perversely showed that when police were not “blind” (and therefore were free to steer witnesses), fillers were never chosen, but that when the police *were* prevented from steering, a more realistic rate of “filler hits” accrued. In other words, they argued that the Pilot Program proved that bad apple detectives were the problem.

The merits of the Illinois Pilot Program are not the issue here. The point is that, regardless of its merits, the Pilot Program’s *method* of addressing eyewitness exonerations generated considerably more heat than light. The only thing everyone agreed on is that it should be done differently next time.¹⁴² Future programs conducted along the lines of the Pilot Program would be seen as both a waste of time and an embarrassment by its objects.

Hennepin County (Minnesota) Attorney Amy Klobuchar also explored the effect of double-blind, sequential lineups on the criminal justice system. This second path showed that the elements needed for a continuous quality improvement approach to criminal justice already existed and were waiting to be assembled.

Mistaken eyewitness identifications had not been seen as a major problem in Minnesota,¹⁴³ and there was no crisis or media firestorm brewing. However, Amy Klobuchar had been monitoring the national discussions triggered by the Green Book and the National Institute of Justice’s *Eyewitness Evidence: A Guide for Law Enforcement*. In effect, Klobuchar exploited the national-to-local arc of a national learning-from-error feedback loop: she built on the recommendations of Janet Reno’s Technical Working Group on Eyewitness Evidence and tested them on the streets of her diverse jurisdiction. The insights of Reno’s heterogeneous group of scientists, prosecutors, defense lawyers, and investigators were exposed to the front-line operators of Minnesota’s system.¹⁴⁴ At the same

Criminal Cases: An Archival Analysis, 25 LAW & HUM. BEHAV. 475, 482 (2001).

¹⁴¹ O’Toole, *supra* note 138, at 18, 20.

¹⁴² See, e.g., Schacter et al., *supra* note 112; Beth Schuster, *Police Lineups: Making Eyewitness Identification More Reliable*, 258 NAT’L INST. JUST. J. 2 (2007); Zack L. Winzeler, Comment, *Whoa, Whoa, Whoa . . . One at a Time: Examining the Responses to the Illinois Study on Double-Blind Sequential Lineup Procedures*, 2008 UTAH L. REV. 1595.

¹⁴³ Klobuchar et al., *supra* note 40, at 382.

¹⁴⁴ The defense function was represented in the Hennepin process only indirectly, through the adaptation of Technical Working Group’s findings. For the operation of that group, which included defenders, see DOYLE, *supra* note 16, at 165-76. Although the prosecutors involved in the Hennepin pilot study undoubtedly included predictions of

time, Klobuchar decided to report the Hennepin County results (no matter what they showed) to the national criminal justice community—thereby closing the loop on the national learning-from-error process.

In Minnesota, the front-line practitioners were involved in the development of the project from the beginning (unlike in Illinois where the front-line police were presented with the Pilot Program on a take-it-or-leave-it basis). The project required the prosecutor's leadership to get the police involved, but it did not require coercion. As Klobuchar and her coauthors describe the initiative,

[a]t the outset, police chiefs registered apprehension toward the new [double-blind, sequential] protocol primarily because existing lineup procedures were working well. Nevertheless, discussions and training sessions sponsored by the Hennepin County Attorney's office convinced the chiefs that the pilot was a worthwhile project Bloomington Police Chief John Laux explained, "In my time since 1968 in law enforcement, I've always been willing to experiment, to try something new. I try to be open-minded and say just because it's working doesn't mean it can't work better."¹⁴⁵

This was the Deming/Berwick ideal of continuous quality improvement in action.

As the study unfolded, the police acted as full partners in identifying and unraveling the operational challenges caused by the conversion from the traditional lineups to the double-blind, sequential format. They were aided by Dr. Nancy Steblay, a psychology professor at Minnesota's Augsburg College. Klobuchar assigned Steblay an active role at every stage: designing, monitoring, and evaluating the study. When problems arose, the police consulted Steblay and derived fixes to those problems. For example, the Minneapolis Police, intrigued by the challenge, developed a laptop photo array program that met both the prosecutor's and the psychologist's criteria.¹⁴⁶ Specific police concerns about the new procedures (for example, potential witness discomfort with the entry of the new "blind" officer at the photo array stage) were addressed promptly and considered by all of the actors. As the process unfolded, new empirical questions were identified, and new experiments were designed to answer those questions.¹⁴⁷ The impacts of small adjustments in operational routine on witness memory were discussed and taken into account. The

defense lawyer reaction to the new procedures in their calculations, and although the general defense position on eyewitness evidence had been thoroughly ventilated in the post-Buckhout debates, it can be argued that the Hennepin effort would have been aided by soliciting that reaction directly sooner in the process rather than later in court.

¹⁴⁵ Klobuchar et al., *supra* note 40, at 405.

¹⁴⁶ *Id.* at 408.

¹⁴⁷ *Id.* at 411.

consequences of new, science-based procedures on routine operational tasks (for example, officer assignments) were similarly examined. The eyewitness “system” was scrutinized as a system, with science, operational expertise, and legal criteria all playing a role in the analysis. Specialists in the street, the lab, and the courtroom met in a conference room to work things out. They developed solutions to the eyewitness error problem identified by Klobuchar and also gained insights into other operational problems that they originally had not known existed.

The Hennepin County team emerged with, in Klobuchar’s words, “stronger cases, and more justice.”¹⁴⁸ They created a genuine, serious, two-way, ongoing conversation. They confirmed that the double-blind, sequential photo-array was an improved component for the investigative process, but they remembered that “having great components is not enough.”¹⁴⁹ They provided an example of how the generation of a specific improvement instigated by the recognition of error furthers—and is furthered by—organization-wide cultural change.

VII. CONCLUSION

The DNA catalogue of wrongful convictions delivered a shock to the system in the world of criminal justice. The system’s operators, to their credit, have responded to the shock with extensive investments of time and energy to try to make things right.

The return on these investments can be compounded if we analyze wrongful convictions and other criminal justice errors as “organizational accidents”: that is, as complex events in which small mistakes combined with each other and with latent conditions hidden in the system to produce unexpected tragedies. A national commitment to fostering the local practice of routinely developing NTSB-style factual reports on criminal justice organizational accidents will provide a more accurate and more useful understanding of the causes and cures of recurrent disasters.

¹⁴⁸ Jim Adams, *Minneapolis, Suburbs Test Witness ID System*, STAR TRIB., Nov. 4, 2003, at B3. In Ramsey County, Minnesota (St. Paul, and surrounding jurisdictions) a similar effort produced similar findings concerning operability issues (real and imagined) and identification accuracy. Susan Gaertner & John Harrington, *Successful Eyewitness Identification Reform: Ramsey County’s Blind Sequential Lineup Protocol*, POLICE CHIEF, Apr. 2009, at 26 (describing the prosecutor’s and police chief’s reports on experience). A parallel effort in Suffolk County (Boston and surrounding cities), Massachusetts, occurred when the elected district attorney convened an “all-stakeholders” task force on eyewitness errors, and substantially re-vamped investigative procedures. See Press Release, *supra* note 39. Subsequent experience indicated that the improvements in investigations had improved conviction rates. Shelley Murphy, *DA Brings in Wins in Homicide Cases: Conley Credits New Procedures*, BOSTON GLOBE, Apr. 12, 2009, at B1.

¹⁴⁹ GAWANDE, *supra* note 129, at 184.

Besides, this new orientation can pay dividends that will eclipse the impact of the content of any reports it generates. The practice of organizational accident review leads organically, as it did in medicine, to a new focus on continuous quality improvement and meets the preconditions for the formation of a new “teaching hospital” function in criminal justice.

The organizational accident approach encourages local leaders to jettison the every-agency-for-itself tradition and form teams of diverse practitioners representing every role in the system, supplemented by relevant specialists. The existing efforts of practitioners on innocence commissions, working groups, and similar vehicles provide strong evidence that this can be done. Despite traditional frictions among police, prosecutors, judges, and defenders, veteran practitioners grow up together, handle the same cases, deal with the same victims and defendants, and work in the same courts. They have more in common with each other than they have in common with anyone else, and, despite their clashing perspectives, they *all* hate criminal justice error. Disentangling organizational accident fact-finding from law reform will minimize any temptation to “game” the fact-finding to avoid anticipated law “reforms” and direct participants’ energies toward improving their understanding of their daily practice. This initiative does not require participation by blue-ribbon dignitaries (as law reform oriented innocence commissions often have); it can—and should—be carried out at the major and lieutenant colonel level rather than at the major general level.

Working steadily on organizational error analysis creates an increased system-consciousness among the practitioners who staff the criminal justice system’s components. Today’s police lieutenants will make better police captains next year thanks to their participation in the rigorous organizational accident examination of a known error or near miss. Assuming that all of the participants behave themselves, mutual trust between adversaries can be expected to grow by degrees. As the participants gain from their experiences, the systems they operate will gain from their insights. If their error reports are disseminated through a national clearinghouse, distant justice systems can gain too. A disciplined commitment to team analysis of error will lay the foundation in criminal justice for realizing the new ideal of continuous quality improvement that is transforming medicine.

The Hennepin County experience with the “wrong man” problem provides one small concrete demonstration that such a change is feasible, that teams can be readily created, and that these alliances can function as effective “teaching hospitals” by bridging the gaps within criminal practice and connecting criminal practitioners to scientific experts. The Hennepin example also suggests that with committed leadership from law

enforcement, prosecutors, the judiciary, or the bar, these teams can be nourished and maintained. The history of the medical quality movement indicates that by treating the error review teams as standing resources, ready to be catalyzed by each new local error, we can begin to foster a system-wide commitment to an ideal of continuous improvement and pursue a “culture of safety” in criminal justice.

At the beginning of the medical quality movement, Lucien Leape observed that:

Physicians and nurses need to accept the notion that error is an inevitable accompaniment of the human condition, even among conscientious professionals with high standards. Errors must be accepted as evidence of systems flaws not character flaws. Until and unless that happens, it is unlikely that substantial progress will be made in reducing medical errors.¹⁵⁰

The same is true in the world of criminal justice. There is no reason to avert our eyes from episodes of dishonesty or incompetence when they occur—and they do occur—or to eliminate law reform, accreditation, and inspection efforts.

But no inspection at the end of the criminal process, however searching, can be sufficient. Even if we somehow accounted for every episode of perjury, laziness, and racism, and refined the formal legal standards and procedures to a high level of effectiveness, we would still reap a bitter harvest of tragedy that has its roots in everyday human mistakes. Building a culture of safety in criminal justice can begin with a simple commitment to routine, candid, non-adversarial examination of error.

¹⁵⁰ Leape, *supra* note 1, at 1857.

