Trade Secrecy and Innovation in Forensic Technology

ELI SIEMS, KATHERINE J. STRANDBURG & NICHOLAS VINCENT[†]

Trade secrecy is a major barrier to public scrutiny of probabilistic software tools that are increasingly used at all stages of the criminal system, from policing and investigation through trial and sentencing. Such tools allow prosecutors to leverage imperfect forensic evidence, such as DNA mixtures, smudged fingerprints, and grainy video footage. Probabilistic software tools unavoidably rely on potentially contestable assumptions, parameters, and implementation choices. Judicially recognized trade secrecy in criminal cases impedes scrutiny of these tools by defendants and the public. Previous critics have focused on secrecy's potential to undermine the integrity and fairness of the criminal justice system, invoking the constitutional constraints of criminal procedure, as well as the traditional accuracy and fairness grounds of evidentiary rules. This Article takes a complementary perspective, arguing that trade secrecy against court-mandated disclosure is also unlikely to advance the recognized goals of trade secrecy law. There is thus certainly no basis for courts to assume that the social benefits of trade secrecy outweigh the potential for injustice created by withholding information needed for adversarial vetting of the reliability of forensic evidence tools.

[†] Associate Counsel, The Legal Aid Society of Westchester County. Eli received his J.D. from NYU School of Law in 2019, where he was a member of the Juvenile Defender and Criminal Defense and Reentry Clinics, member and Student Fellow Coordinator of the Information Law Institute's Privacy Research Group, and Senior Articles Editor for *The Review of Law and Social Change.* Alfred Engelberg Professor of Law and Faculty Director, Information Law Institute, New York University School of Law. Professor Strandburg acknowledges the generous support of the Filomen D. Agostino and Max E. Greenberg Research Fund. Associate (Patent), Morrison & Foerster, LLP, J.D., New York University School of Law, 2020; Ph.D., Yale University (Microbiology), 2017. While at NYU, Nicholas served as Editor-in-Chief of the *NYU Journal of Intellectual Property & Entertainment Law*.

The authors are also grateful for terrific research assistance from Thomas McBrien, Christine Song, Tanuj Dayal and Danya Amir and for helpful comments from members of the NYU Privacy Research Group and Information Law Institute and from attendees at the Privacy Law Scholars Conference and NYU's Conference on Trade Secrets and Algorithmic Systems.

TABLE OF CONTENTS

INTRODUCTION

The criminal justice system increasingly relies on probabilistic algorithms at all stages from policing and investigation[1] through trial and sentencing.[2] These software-implemented algorithms are used for various purposes including estimating the likelihood that imperfect evidence, such as DNA mixtures, smudged fingerprints and grainy video footage, is associated with a particular suspect, as well as predicting whether a person is likely to commit a crime if released before trial or on parole and directing policing resources.[3] One common thread among these tools is their reliance on computational techniques that embed contestable choices, parameters, and assumptions. These tools, and the efforts by their developers to shield their underlying source code from disclosure, have been the subject of considerable public and scholarly debate and have recently spurred activism from a perhaps surprising group—mathematicians. In October 2020, the Notices of the American Mathematical Society published a letter with over 2,000 signatories calling for mathematicians to sever ties with police departments and to "demand that any algorithm with potential high impact face a public audit."[4]

Trade secrecy is one major obstacle to the scrutiny called for by these mathematicians and many other commentators. This Article focuses on assertions of trade secrecy regarding probabilistic software algorithms used to produce evidence for criminal trials. While questions about the "black box dangers" of non-transparent forensic technologies are hardly new,[5] the issue is of growing practical importance because of the growing reliance on probabilistic

---

1. In essence, predictive policing is an algorithm-based method/analysis that uses large data sets to forecast crime, but there are concerns that the underlying algorithms play a greater role in reinforcing racial bias than in successfully predicting criminal behavior. *See* Tim Lau, *Predictive Policing Explained*, BRENNAN CTR. (Apr. 1, 2020), https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained.

2. Recidivism risk assessment tools are algorithms that purport to calculate the likelihood an offender will commit another crime in the future. These risk assessment tools assign a predictive score to an individual based on a dataset comprised of crime and offender statistics. Commentators have noted that the datasets are often incomplete or inaccurate, and that there is little reason to believe that the data in the datasets are sufficiently accurate to make accurate or meaningful recidivism predictions. Stephanie Lacambra, Jeremy Gillula & Jamie Williams, *Recidivism Risk Assessments Won't Fix the Criminal Justice System*, ELEC. FRONTIER FOUND. (Dec. 21, 2018), https://www.eff.org/deeplinks/2018/12/recidivism-risk-assessments-wont-fix-criminal-justice-system.

3. *See* Rebecca Wexler, *It's Time to End the Trade Secret Evidentiary Privilige Among Forensic Algorithm Vendors*, BROOKINGS (July 13, 2021), https://www.brookings.edu/blog/techtank/2021/07/13/its-time-to-end-the-trade-secret-evidentiary-privilege-among-forensic-algorithm-vendors; Lacambra et al., *supra* note 2.

4. Tarik Aougab, Federico Ardila, Jayadev Athreya, Edray Goins, Christopher Hoffman, Autumn Kent, Lily Khadjavi, Cathy O'Neil, Priyam Patel & Katrin Wehrheim, *Letter to the Editor*, 67 NOTICES AM. MATHEMATICAL SOC'Y 1293 (2020); Lilah Burke, *Mathematicians Urge Ending Work With Police*, INSIDEHIGHERED.COM (June 24, 2020), https://www.insidehighered.com/news/2020/06/24/mathematicians-urge-cutting-ties-police; *see also* Ethan Zell, *Let's Take Responsibility For Our Math*, AMS: BLOGS (June 27, 2020), https://blogs.ams.org/mathgradblog/2020/06/27/lets-take-responsibility-for-our-math.

5. *See* Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245 (2016); Jessica Gabel Cino, *Deploying the Secret Police: The Use of Algorithms in the Criminal Justice System*, 34 GA. ST. U. L. REV. 1073 (2018); Charles Short, *Guilt by Machine: The Problem of Source Code Discovery in Florida DUI Prosecutions*, 61 FLA. L. REV. 177 (2009).

software algorithms in the criminal justice system.[6] Unlike more traditional forensic analyses performed in crime labs, probabilistic software tools are commonly procured from private companies, which assert trade secrecy not only to avoid disclosure in court, but also to keep information secret from the crime labs, law enforcement agencies, and prosecutors who are their primary customers. Judges have been perhaps surprisingly willing to deny disclosure to defendants on trade secrecy grounds. As a result of these technological and judicial trends, trade secret algorithms increasingly underlie evidence offered at the pre-trial, trial, and sentencing phases and undoubtedly cast shadows beyond the courtroom to plea bargaining and charging decisions.[7]

We are not the first to criticize courts' willingness to accept trade secrecy as reason for denying disclosure of the underpinnings of forensic evidence. Previous critiques have centered on secrecy's potential to undermine the integrity and fairness of the criminal justice system,[8] invoking the constitutional constraints of criminal procedure, as well as the traditional accuracy and fairness grounds of evidentiary rules. For example, the Justice in Forensic Algorithms Act of 2019 (reintroduced in April 2021 as the Justice in Forensic Algorithms Act of 2021)[9] aimed "[t]o prohibit the use of trade secrets privileges to prevent

6. To our knowledge, the first judicial opinion analyzing whether trade secrecy could preclude discovery of the software used to produce forensic evidence was issued in 2006 and dealt with breath alcohol testing. *See* Moe v. State, 944 So. 2d 1096 (Fla. 5th DCA, Nov. 17, 2006). Nonetheless, while several other courts dealt with breathalyzer software in the following years, it was not until 2013 that criminal courts began to issue opinions addressing trade secrecy claims for other software-based forensic evidence tools. *See* United States v. Ocasio, 2013 U.S. Dist. LEXIS 79313 (W.D. Tex. June 6, 2013). The first opinion specifically considering an assertion of a trade secret privilege in a criminal case was written even more recently, in 2015. *See* State v. Superior Court (Chubbs), B258569 2015 WL 139069 (Cal. Ct. App. Jan. 9, 2015).

7. *See, e.g.*, Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972 (2017); Meghan J. Ryan, *Secret Conviction Programs*, 77 WASH. & LEE L. REV. 269 (2020).

8. *See, e.g.*, Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018); Natalie Ram, *Innovating Criminal Justice*, 112 NW. U. L. REV. 659 (2017–2018); Jennifer N. Mellon, *Manufacturing Convictions: Why Defendants Are Entitled to the Data Underlying Forensic DNA Kits*, 51 DUKE L.J. 1097, 1119 (2001); Steven M. Bellovin, Matt Blaze, Susan Landau & Brian Owsley, *Seeking the Source: Criminal Defendants' Constitutional Right to Source Code*, 17 OH. ST. TECH L.J. 1 (2021); Vera Eidelman, *The First Amendment Case for Public Access to Secret Algorithms Used in Criminal Trials*, 34 GA. ST. U. L. REV. 915 (2018); Jeanna Matthews, Marzieh Babaeianjelodar, Stephen Lorenz, Abigail Matthews, Mariama Njie, Nathaniel Adams, Dan Krane, Jessica Goldthwaite & Clinton Hughes, *The Right to Confront Your Accusers: Opening the Black Box of Forensic DNA Software*, *in* AIES '19: PROCEEDINGS OF THE 2019 AAAI/ACM CONFERENCE ON AI, ETHICS & SOCIETY 321 (2019), https://doi.org/10.1145/3306618.3314279.

9. Press Release, Rep. Mark Takano, Reps. Takano and Evans Reintroduce the Justice in Forensic Algorithms Act to Protect Defendants' Due Process Rights in the Criminal Justice System (Apr. 8, 2021), https://takano.house.gov/newsroom/press-releases/reps-takano-and-evans-reintroduce-the-justice-in-forensic-algorithms-act-to-protect-defendants-due-process-rights-in-the-criminal-justice-system. At the time of writing, the 2021 Bill had been referred to both the Committee on the Judiciary (Subcommittee on Crime, Terrorism, and Homeland Security) and the Committee on Science, Space, and Technology (Subcommittee on Research and Technology). H.R. 2438, 117th Cong. (2021), https://www.congress.gov/bill/117th-congress/house-bill/2438/all-actions?s=1&r=3. The two versions of the Bill share the same core purpose, but there are several changes worth noting. For example, the stated purpose of the 2019 Bill was to establish "Computational Forensic Algorithm Standards," while the stated purpose of the 2021 Bill is to establish "Computational Forensic Algorithm Testing Standards" as well as a new "Computational Forensic Algorithm Testing Program." *Compare*

defense access to evidence in criminal proceedings, provide for the establishment of Computational Forensic Algorithm Standards, and for other purposes."[10] In introducing the legislation, Representative Mark Takano (D-Cal.) emphasized that "[t]he trade secrets privileges of software developers should never trump the due process rights of defendants in the criminal justice system."[11] Despite the weight of these critiques, judges mostly continue to deny disclosure about forensic technologies in response to trade secrecy assertions.[12]

Moreover, judicial opinions in this area reflect an assumption that standard innovation-based justifications for trade secrecy are persuasive in this context. For example, the Pennsylvania Superior Court's seminal decision denying disclosure of probabilistic genotyping source code in a criminal case noted that "TrueAllele is proprietary software; it would not be possible to market TrueAllele if it were available for free."[13] A fortiori, these disclosure denials cannot be justified if they do not even advance trade secrecy's conventional policy goals. We therefore analyze whether standard justifications for legal trade secrecy protection are applicable to forensic evidence technology. Our analysis is thus complementary to critiques based on the importance of disclosure to the integrity of the justice system. We conclude that there is no basis for assuming that the purported social benefits of trade secrecy outweigh the potential for injustice created by withholding information needed for adversarial vetting of the reliability of forensic evidence tools. Indeed, trade secrecy's justifications are largely inapplicable to court-ordered disclosures about forensic evidence technology.

This Article focuses on probabilistic software-implemented tools, for which trade secrecy is particularly troublesome because of their reliance on embedded assumptions.[14] Our analysis is illustrated and informed by a detailed

---

H.R. 4368 116th Cong. (2019), https://www.congress.gov/bill/116th-congress/house-bill/4368/text *with* H.R. 2438, 117th Cong. (2021), https://www.congress.gov/bill/117th-congress/house-bill/2438/all-actions?s=1&r=3. Although Article V of the 2019 version of the Bill aimed to remove trade secrecy protections in any criminal case when defendants would otherwise be entitled to obtain evidence, Article 2(b) of the 2021 Bill merely states that "There shall be no trade secret evidentiary privilege to withhold **relevant evidence** in criminal proceedings in the United States courts." H.R. 2438 117th Cong. § 2(b)(1) (2021) (emphasis added). Importantly, the 2021 version of the Bill also clarified that the defendant shall have access to the source code of the computational forensic software. H.R. 2438 117th Cong. § 2(f) (2021).

10. H.R. 4368 116th Cong. (2019).

11. Press Release, *supra* note 9; *see, e.g.*, H.B. 118. 65th Leg., 1st Reg. Sess. (Idaho 2019) (requiring pretrial risk assessment algorithms be "transparent" and further specifying that "[n]o builder or user or a pretrial risk assessment algorithm may assert trade secret or other protections in order to quash discovery in a criminal matter by a party to a criminal case").

12. *See infra* Part III.D.4.

13. Commonwealth v. Foley, 38 A.3d 882, 889 (Pa. Super. Ct. 2012). Other opinions reflect similar concerns. *See, e.g.*, People v. Lopez, 23 N.Y.S.3d 820, 829 (Sup. Ct. Bronx Co. 2015) ("To the extent they claim it would be easier to perform the calculations with the actual program software, the computer program itself is proprietary and the court is not ordering its disclosure.").

14. Bellovin et al., *supra* note 8, at 1–2; Katherine Kwong, Note, *The Algorithm Says You Did It: The Use of Black Box Algorithms to Analyze Complex DNA Evidence*, 31 HARV. J.L. & TECH. 275, 290 (2017); Bess Stiffelman, *No Longer the Gold Standard: Probabilistic Genotyping is Changing the Nature of DNA Evidence*

case study of probabilistic genotyping, a technique for analyzing DNA samples that is often used in trials for serious crimes and has been at the center of considerable recent controversy. Probabilistic genotyping ("PG") employs computational optimization techniques to analyze DNA evidence that is not amenable to gold standard direct comparison approaches, often because it contains mixtures of DNA from several unknown individuals.[15] Probabilistic genotyping is a useful case study because it has been around long enough for market trends in its adoption and use to be observed.[16] Moreover, trade secrecy-based refusals to disclose PG source code and other implementation details have been repeatedly challenged by defendants over the past seven or eight years,[17] with little success, producing a rich judicial record.

United States federal and state trade secrecy laws are violated when trade secret information is misappropriated. Obtaining the same information through independent derivation or reverse engineering, however, is perfectly legitimate.[18] Legal trade secrecy protection has two main policy justifications, both grounded in concerns about market failure. First, trade secrecy law aims to help ensure a well-functioning market by punishing misappropriation and deterring wasteful investments in an economic espionage arms race. Second, as a practical matter, secrecy can enhance incentives to invest in innovation by preventing competitors from free riding on those investments. By deterring misappropriation, trade secrecy laws can also bolster those incentives. Of course, this incentive enhancement occurs only for information that can be kept secret in the first place and is not adequately covered by other intellectual property protections. Moreover, the free rider justification involves inevitable tradeoffs, well recognized in intellectual property law, between the social benefits of the incentives provided by market exclusivity and the social benefits of competition.

As this Article explains in detail, these standard trade secrecy justifications are weak for forensic evidence technologies because the feared market failures are simply unlikely to arise. Most obviously, the first set of concerns about economic espionage and misappropriation are simply irrelevant to court-ordered disclosures. As explained in detail below, court-ordered disclosures are also unlikely to raise the second set of concerns about competitor free riding. Markets for forensic evidence technology are quite different from the consumer product markets that have shaped trade secrecy law and policy. Demand in such markets is dominated by concerns about admissibility, and hence shaped by evidence

in Criminal Trials, 24 BERKELEY J. CRIM. L. 110, 113 (2019); Katherine L. Moss, Note, *The Admissibility of TrueAllele: A Computerized DNA Interpretation System*, 72 WASH. & LEE L. REV. 1033, 1072 (2015).

15. Stiffelman, *supra* note 14, at 118.

16. *History*, CYBERGENTIC, https://www.cybgen.com/company/history.shtml (last visited Mar. 21, 2022).

17. *See, e.g.*, People v. Wakefield, 9 N.Y.S.3d 540, 541 (N.Y. Sup. Ct. 2015); People v. Dominguez, 239 Cal. Rptr. 3d 71, 75, 77 (Ct. App. 2018).

18. *See, e.g.*, Andrew A. Schwartz, *The Corporate Preference for Trade Secret*, 74 OHIO ST. L.J. 623, 630 (2013).

doctrine and judicial rulings. Judicial admissibility rulings build upon one another in a classic "rich get richer" fashion, creating a barrier to entry by competing firms. Those first mover advantages create barriers to entry that tend to outweigh the advantages that disclosure gives to potential free riders in this context.

Moreover, secrecy regarding the underpinnings of forensic technology is likely to exacerbate a different sort of market failure. The social benefits of markets depend on their ability to ensure the production of socially valuable goods and services by responding to customer preferences.[19] In forensic technology markets, law enforcement agencies and individual crime labs are the dominant customers, whose preferences guide producers. Unlike ordinary consumers, however, these customers serve as purchase agents for society. They are imperfect agents, however, because they are rewarded directly for solving crimes and convicting perpetrators and only indirectly, at best, for avoiding mistaken arrests and convictions.[20] This misalignment is a classic principal-agent problem, which evidence doctrine, judicial gatekeeping, and the adversarial process are intended to address. These realignment mechanisms cannot work, however, if the technology is not disclosed or adequately validated. Trade secrecy, along with shortcomings in validation requirements,[21] undermines the efficacy of these mechanisms, especially for probabilistic software tools. As a result, market demand will tend to deviate from society's goals and steer innovation in sub-optimal directions.

Part I of this Article motivates and introduces our qualitative case study of probabilistic genotyping, a software tool for analyzing DNA evidence that has been at the center of recent controversy. Part II discusses the standard market-based justifications for trade secret protection. It analyzes how the context of court-ordered disclosure narrows the applicability of these concerns. It then uses the probabilistic genotyping case study to illustrate how the distinctive features of forensic evidence technology markets tend to extend first mover advantages, mitigating free rider concerns. It also explains how secrecy can exacerbate principal-agent problems in these markets, thereby skewing innovation incentives in socially sub-optimal directions. Part III briefly considers whether and how our analysis might change in the context of technologies that are not developed primarily for forensic evidence use. It considers incidental evidence

19. James C. Anderson & James A. Narus, *Business Marketing: Understand What Customers Value*, HARV. BUS. REV., Nov.–Dec. 1998, at 53, 54, https://hbr.org/1998/11/business-marketing-understand-what-customers-value.

20. *See* Roth, *supra* note 5*;* Jeanna Neefe Matthews, Graham Northup, Isabella Grasso, Stephen Lorenz, Marzieh Babaeianjelodar, Hunter Bashaw, Sumona Mondal, Abigail Matthews & Mariama Njie, *When Trusted Black Boxes Don't Agree: Incentivizing Iterative Improvements and Accountability in Critical Software Systems*, *in* AIES '20: PROCEEDINGS OF THE AAAI/ACM CONFERENCE ON AI, ETHICS, AND SOCIETY 103 (2020), https://doi.org/10.1145/3375627.3375807.

21. See Bellovin et al., *supra* note 8, for a discussion of the difficulties of validating complicated software systems. We will discuss the shortcomings of current standards of validation and admissibility in the context of software-based evidentiary tools elsewhere.

technologies, developed and marketed primarily for private commercial use; and dual-purpose technologies, for which both sorts of demand are significant.

I. PROBABILISTIC GENOTYPING: A FORENSIC EVIDENCE TECHNOLOGY CASE STUDY

Currently, controversy swirls around trade secrecy claims related to probabilistic genotyping software for interpreting crime scene samples. The technique is popular because it aims to extract information from samples that cannot be analyzed with more established techniques, including those containing complex mixtures of DNA from more than one individual.[22] It has been employed to produce evidence for criminal trials in the United States since at least 2009 and adopted by crime labs in numerous jurisdictions. Defense requests for disclosure of trade secret PG source code and other implementation details have been the subject of a growing number of judicial opinions. This robust record makes PG an excellent case study to inform and illustrate our theoretical analysis. Subpart A of this Part compares probabilistic genotyping to more traditional DNA analysis and explains why it depends much more heavily on assumptions made by those who develop and use the software. Subpart B explores the value of disclosure to defendants—and highlights the perils of secrecy regarding source code and other implementation details. Subpart C introduces the major players in the market for PG tools based on a review of company websites, writings by the founders of the dominant companies and other relevant materials. Finally, Subpart D maps the network of judicial opinions on admissibility disclosure, demonstrating how the market appears to be driven by a sort of "rich get richer" effect.

A. PROBABILISTIC GENOTYPING AND CRIME SCENE DNA

DNA evidence obtained from robust, single-source samples is widely deemed the gold standard of forensic evidence and is thus highly persuasive to courts and juries.[23] This sort of DNA analysis has been used not only to obtain convictions, but also to exonerate and free the wrongly imprisoned.[24] Because "DNA evidence" has such an excellent reputation and track record, it is important to explain why probabilistic genotyping is different—more complicated, more subjective, and more prone to error.

Probabilistic genotyping is used to analyze crime scene DNA samples that are much less robust and more complex because they are mixtures of DNA from an unknown number of individuals, contain less genetic material and may be

22. Matthews et al., *supra* note 8, at 322; Kwong, *supra* note 14, at 276.

23. *See, e.g.*, Sarah Hammond, *The DNA Factor: Lawmakers are Expanding the Use of Forensic Technology to Battle Crime*, 36 STATE LEGS. 12 (2010), https://www.ncsl.org/research/civil-and-criminal-justice/the-dna-factor.aspx.

24. GERALD LAPORTE, NAT'L INST. JUST., WRONGFUL CONVICTIONS AND DNA EXONERATIONS: UNDERSTANDING THE ROLE OF FORENSIC SCIENCE 2 (2018), https://www.ncjrs.gov/pdffiles1/nij/250705.pdf.

degraded in various ways. Genetic profiles obtained from these samples present immense, likely insurmountable, challenges for traditional comparison by human analysts.[25] Instead, PG tools employ complicated statistical fitting techniques that can only be implemented by computers.[26] While the basic numerical techniques employed by these tools are well-understood, their implementation is not entirely routine. PG software enshrines various assumptions about how to perform these fits, as well as employing parameters that must be supplied by crime lab analysts based on their own assumptions.[27] Because of these complexities, ensuring that a given tool has been properly implemented and used in a given case is not a simple matter of ex ante validation for a few representative samples.[28] While leaving details to a technical appendix, this Subpart provides a sketch of the PG technique and its assumptions to explain why defendants seek disclosure and why disclosure would serve the public interest in fair and just trials.

1. Traditional DNA Evidence

The development of traditional forensic DNA analysis was a huge breakthrough for finding matches in situations where high quality DNA samples with only one unknown contributor were available. The vast majority of DNA samples collected during the early use of forensic DNA analysis were collected in sexual assault cases[29] and contained relatively high-quality DNA from only two contributors—the victim and the perpetrator.[30] In such cases, the question for crime lab analysis was relatively simple—does the DNA profile of the suspect match the perpetrator profile extracted from the sample? While not devoid of contestable judgment calls, this sort of comparison was more scientifically rigorous than many other forensic techniques and quickly became the gold standard for both prosecution and defense.[31]

Forensic DNA analysis begins with using well-established laboratory techniques to sequence the crime scene sample to obtain a DNA profile. Forensic DNA analysis does not focus on the biologically active parts of the genome[32] (that is, "genes") that are known to "code" for proteins and to determine an

25. Kwong, *supra* note 14, at 300.

26. *Id.* at 281.

27. Matthews et al., *supra* note 20, at 103.

28. *See* Hinda Haned, Peter Gill, Kirk Lohmueller, Keith Inman & Norah Rudin, *Validation of Probabilistic Genotyping Software for Use in Forensic DNA Casework: Definitions and Illustrations*, 56 SCI. & JUST. 104, 107–08 (2016).

29. *See* Michael D. Coble & Jo-Anne Bright, *Probabilistic Genotyping Software: An Overview*, 38 FORENSIC SCI. INT'L GENETICS 219, 220 (2019).

30. *Id.*; T.M. Clayton, J.P. Whitaker, R. Sparkes & P. Gill, *Analysis and Interpretation of Mixed Forensic Stains Using DNA STR Profiling*, 91 FORENSIC SCI. INT'L 55, 64; *see also* Yolanda Torres, Inmaculada Flores, Victoria Prieto, Manuel López-Soto, María José Farfán, Angel Carracedo & Pilar Sanz, *DNA Mixtures in Forensic Casework: A 4 Year Retrospective Study*, 134 FORENSIC SCI. INT'L 180, 181 (2003).

31. *See* Coble & Bright, *supra* note 29.

32. The genome is the complete set of genetic information in an organism. *Genome*, SCITABLE BY NATURE EDUC., https://www.nature.com/scitable/definition/genome-43 (last visited Mar. 21, 2022).

individual's genetic traits. Instead, forensic DNA analysis involves a relatively small number of "short tandem repeat" ("STR") sequences, which are repeated several times at predetermined loci in the genome.[33] Forensic analysis in the U.S. ordinarily employs up to twenty to twenty-four such loci.[34]

While the STR sequences used in forensic DNA analysis are not generally of significant biological importance (they are "noncoding"), profiles based on them are valuable because they are distinctive. The exact number of STR repeats at a given locus varies from individual to individual,[35] making it highly statistically unlikely that two individuals' DNA profiles will match at all of the pre-determined locations. The profile produced by sequencing a high quality, single source sample can thus be used as a "DNA fingerprint" for comparison with a profile based on a sample taken from a known suspect.[36]

In a forensic profile output measuring the presence of various STRs, there will be several peaks plotted against each predetermined locus used in the analysis. Different peak patterns correspond to different genetic variants or "alleles."[37] An individual generally has two alleles at a given locus for a coding gene, one contributed by each parent. If those variants are the same, the person is "homozygous" at that locus; if the variants are different, the person is "heterozygous" at that locus.[38] For the non-coding STR loci that are used in forensic DNA analysis, each "allele" peak corresponds to a different number of repeats of that STR. The sequencing readout from the crime scene sample can

33. A locus is a fixed position on a chromosome that contains genetic information encoding a particular gene or genetic marker. While the location of genes and genetic markers is the same from person-to-person, the actual genetic material encoded at each locus (that is, alleles) can be used to identify individuals. *Allele*, SCITABLE BY NATURE EDUC., https://www.nature.com/scitable/definition/allele-48 (last visited Mar. 21, 2022).

34. JOHN M. BUTLER, HARI IYER, RICH PRESS, MELISSA K. TAYLOR, PETER M. VALLONE & SHEILA WILLIS, NAT'L INST. OF STANDS. & TECH, DNA MIXTURE INTERPRETATION 21 (2012), https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8351-draft.pdf; *see also* JOHN BUTLER, ADVANCED TOPICS IN FORENSIC DNA TYPING: METHODOLOGY (2012); JOHN BUTLER, ADVANCED TOPICS IN FORENSIC DNA TYPING: INTERPRETATION (2015).

35. Stephanie Feupe Fotsing, Jonathan Margoliash, Catherine Wang, Shubham Saini, Richard Yanicky, Sharona Shleizer-Burko, Alon Goren & Melissa Gymrek, *The Impact of Short Tandem Repeat Variation on Gene Expression*, 51 NATURE GENETICS 1652, 1652 (2019), https://www.nature.com/articles/s41588-019-0521-9 (stating that STRs represent a large source of genetic variation with mutation rates that are orders of magnitude higher than other portions of the genome and further stating that "each individual is estimated to harbor around 100 de novo mutations in STRs").

36. *See What is a DNA Fingerprint?*, YOURGENOME (July 21, 2021) https://www.yourgenome.org/facts/what-is-a-dna-fingerprint.

37. Karen Norrgard, *Forensics, DNA Fingerprinting, and CODIS*, SCITABLE BY NATURE EDUC., https://www.nature.com/scitable/topicpage/forensics-dna-fingerprinting-and-codis-736 (last visited Mar. 21, 2022) ("For instance, the STR known as D7S820, found on chromosome 7, contains between 5 and 16 repeats of GATA. Therefore, there are 12 different alleles possible for the D7S820 STR. An individual with D7S820 alleles 10 and 15, for example, would have inherited a copy of D7S820 with 10 GATA repeats from one parent, and a copy of D7S820 with 15 GATA repeats from his or her other parent."). In coding regions of DNA, an allele generally refers to a version of the gene, not the number of copies of a repeat present. In the case of a non-coding region like an STR, an allele generally refers to the number of repeats present, since that, technically, represents the possible "versions" of that segment of DNA (or locus) in the population.

38. *Allele*, NAT'L HUM. GENOME RSCH. INST., https://www.genome.gov/genetics-glossary/Allele (last visited Mar. 21, 2022).

then be compared to the sequencing readout of the suspect's DNA, resulting in a final determination about whether the suspect was involved in or present at the potentially criminal act.

2. Complex Crime Scene Samples and Probabilistic Genotyping

In many circumstances, crime scene DNA samples are not amenable to the gold standard analysis described above because they are contaminated, degraded or simply too small to produce highly accurate sequencing results.[39] These problems reduce the certainty of DNA matching even for single-source samples. Real crime scene samples may also contain DNA from more than one, and often an unknown number of, individuals. Laboratory analysis, however, gives only one combined profile, which is an entangled superposition of profiles from all of these DNA contributors. Sequencing alone cannot ascertain who contributed the DNA, how many people contributed to the mixture, or what the possible individual genetic profiles are. These sources of uncertainty often compound one another in real crime scene samples.

Probabilistic genotyping software uses sophisticated computational techniques to try to estimate how likely it is that a particular suspect's DNA is included in such a mixed, potentially degraded crime scene sample. The output of probabilistic genotyping software programs is reported in the form of a likelihood ratio ("LR"), which expresses the relative probability that the observed sequencing results reflect a scenario including the suspect, compared to the likelihood that a randomly chosen individual was at the scene.[40] The computational approach uses a statistical technique to generate many possible origin stories for the observed profile and assign each a relative probability based on how well it fits the available data and how plausible it is in light of what is known or presumed about population genetics.[41] PG algorithms thus are based on biological modeling; human, molecular, and population genetics; statistical analysis techniques; and computer science.[42] PG analysis is controversial because it depends much more heavily than traditional DNA analysis, and less straightforwardly, on various assumptions made by programmers and crime lab analysts.[43]

---

39. In addition, the analysis may have to account for drop-out (when an allele is missing from the sequencing data because it was not present or because it was present in too low levels to be detected by the sequencing apparatus) and stutter (an artifact of DNA sequencing that can result in inaccurate reads). *See* Moss, *supra* note 14, at 1034, as well as the Appendix below.

40. *See* John S. Buckleton, Jo-Anne Bright, Simone Gittelson, Tamyra R. Moretti, Anthony J. Onorato, Frederick R. Bieber, Bruce Budowle & Duncan A. Taylor, *The Probabilistic Genotyping Software STRmix: Utility and Evidence for its Validity*, 64 J. FORENSIC SCI. 393, 394 (2019). It is important to note that a likelihood ratio expresses the ratio of two mutually exclusive events. That is, the POI either is, or is not, a contributor to the mixture. *Id.*

41. *See* THERMOFISHER SCI., PROBABILISTIC GENOTYPING (2018), https://assets.thermofisher.com/TFS-Assets/LSG/Flyers/prob-geno-hps-flyer.pdf; *See* Coble & Bright, *supra* note 29, at 219.

42. *See* THERMOFISHER SCI., *supra* note 41.

43. *See* Coble & Bright, *supra* note 29, at 223.

A simplified example illustrates some of the ways that such assumptions can affect the analysis. Assume we use a simplified DNA profile based on only one locus (instead of the usual twenty to twenty-four) and that there are three possible alleles at that locus, which we will call A, B, and C. The DNA profile from a crime scene sample might then contain two peaks of height *x* at positions A and B and one of height 2*x* at position C. (See Figure 1.) To use one of the available PG tools to interpret this data, the forensic analyst must input an assumption about the number of contributors.[44] Experienced technicians can make educated guesses by eyeballing the DNA sequencing data, but for full DNA profiles of real crime scene samples the determination is anything but certain. For our over-simplified example, suppose our forensic analyst assumes that the sample contains equal amounts of DNA from two contributors. Assume also that all alleles of all contributors are present and detected in the mixture. (In practice, it is possible that not all alleles of all contributors are present in the mixture or that one or more go undetected by sequencing technologies.)[45] Essentially, we assume that the profiles from scenario (AC, BC) and scenario (AB, CC) (and only those profiles) fit the data perfectly. The analysis thus eliminates scenarios involving people with genotypes other than AC, BC, AB and CC. Even with these prosecution-friendly assumptions and our toy one-locus profile, there are two scenarios that could have produced the sequencing graph shown in Figure 1. First, the sample could contain DNA from one contributor with genotype AC and one with genotype BC at this locus. Such a mixture would produce the observed profile with peak of height x at A and B and height 2x at C. Alternatively, the sample could contain DNA from contributors with genotypes AB and CC, which would produce exactly the same readout. Unless we have further information, we must assume that genotypes AC, BC, AB and CC are equally likely to be found in the "reference population" of people who might have visited the crime scene, so that the scenarios (AC, BC) and (AB, CC) are equally likely to explain the DNA data.

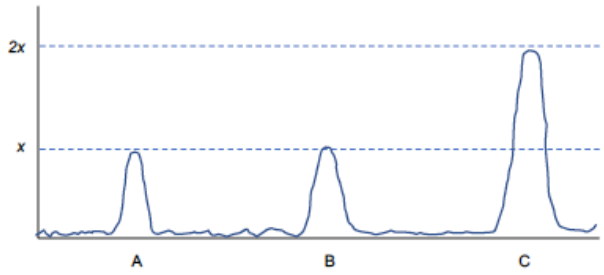
---

44. Catherine McGovern, Kevin Cheng, Hannah Kelly, Anna Cieck, Duncan Taylor, John S. Buckleton & Jo-Anne Bright, *Performance of a Method for Weighting a Range in the Number of Contributors in Probabilistic Genotyping*, 48 FORENSIC SCI. INT'L GENETICS 102352 (2020); *see also* Coble & Bright, *supra* note 29, at 220.

45. This is referred to as "allele dropout." *See What is a DNA Fingerprint?*, *supra* note 36. For example, assume the readout illustrated in the text (i.e., A = height *x*; B = height *x*; C = height *2x*). It is also theoretically possible that there were three (or even more) contributors present, but that, due to sample quality or amount, one or more alleles went undetected during sequencing. Assume three contributors present at the crime scene and the presence of an A allele and B allele that went undetected in the mixture. With perfect knowledge, we would know that the sequencing should have detected heights of *2x* at each position A, B, and C. As a result, the real contributor profile could be AA, BB, CC; AA, BC, BC; AB, AB, CC; AC, AC, BB; AB, AC, BC; etc. Each of these differs meaningfully from the AC, BC or AB, CC predictions based on the readout presented in the text. When the possibility of allele dropout is added to even this simple example, the number of potential contributor profiles and their potential complexity expands rapidly. A realistic probabilistic genotyping algorithm must account for the likelihood of allele dropout in light of the quality of the input data. (*See* Figure 1.)

FIGURE 1: ILLUSTRATION OF HYPOTHETICAL SAMPLE

Figure 1. This figure illustrates the hypothetical example posed in Part II.A.



In our simplified example that assumes two contributors to a complex mixture of DNA, this profile could represent one contributor with the genotype AC and one contributor with the genotype BC, or alternatively, it could represent one contributor with the genotype AB and one contributor with the genotype CC. A, B, and C represent three possible alleles found at a single locus. The values *x* and *2x* represent the relative amount of DNA found in the hypothetical sample assumed in the example.

Now suppose we have a suspect with genotype AC. Because there is a 50:50 chance that the (AC, BC) scenario is correct, there is a 50% chance that someone with genotype AC was at the crime scene. What is the probability that our suspect was that someone? In the best-case scenario for the prosecution, there might be reason to believe that there is only one individual in the reference population with genotype AC. In that case, the probability that our suspect's DNA is in the sample would be 50%. Otherwise, if there are N_{AC} individuals with genotype AC in the reference population, the probability that our suspect's DNA is in the sample is lower (50% divided by N_{AC}).

Rather than reporting this sort of straightforward probability estimate, PG tools generally output a "likelihood ratio."[46] The likelihood ratio divides the PG estimate of the probability that the suspect's DNA was in the crime scene DNA sample by the analyst's estimate of that probability before seeing the crime scene sequence data.[47] Before seeing the crime scene profile in Figure 1, the analyst in our example can only guess that our suspect's DNA is in the crime scene sample with probability 1/N, where N is the size of the reference population. Thus, the likelihood ratio multiplies the probability that the suspect's DNA is in the crime scene sample by the size of the reference population.

46. *See* Appendix.

47. Mark W. Perlin, *Explaining the Likelihood Ratio in DNA Mixture Interpretation*, *in* THE PROCEEDINGS OF PROMEGA'S TWENTY FIRST INTERNATIONAL SYMPOSIUM ON HUMAN IDENTIFICATION 6 (2010), https://www.promega.ee/~/media/files/resources/conference%20proceedings/ishi%2021/oral%20presentations/perlin.pdf.

Reported Likelihood Ratios are often much, much larger than the odds that the suspect's DNA is in the sample because they assume a large reference population. Thus, in a situation similar to our hypothetical, in which the analysis revealed two equally plausible scenarios, a standard calculation might put the odds that the suspect was at the crime scene at 1 out of 300 million before the DNA analysis and somewhere around 50% after the analysis, giving a likelihood ratio of about 150 million.

The likelihood ratio approach to expressing the results of DNA analysis can thus be highly misleading because it measures the wrong thing from an evidentiary perspective.[48] The relevant question is not how much the DNA analysis has reduced the number of plausible scenarios, but how many plausible scenarios are left after the analysis. In our toy hypothetical, the DNA analysis leaves us with two equally plausible scenarios, one of which excludes the suspect. No matter how large the likelihood ratio, the suspect should not be convicted unless there is evidence that rules out the alternative scenario. The example also shows how sensitive the likelihood ratio is to assumptions about the reference population. The larger the reference population, the larger the likelihood ratio, even if the reference population includes scenarios that are highly implausible in the context of the particular crime.

The hypothetical in Figure 1 is illustrative, but unrealistic. In reality, the sequencing result from a crime scene sample will contain several peaks at up to twenty to twenty-four loci,[49] may represent more than two contributors, and may be compromised by sample quality or other artifacts. As a result, rather than two possible scenarios that fit the crime scene sample sequencing data perfectly, there are likely to be many alternative scenarios that fit the data reasonably well. PG tackles this problem using a computational approach called Markov Chain Monte Carlo (MCMC),[50] which essentially imagines an enormous number of possible scenarios and estimates their likelihood in light of how well they fit the data and how prevalent similar profiles are in the population. The likelihood ratios calculated by PG tools are highly dependent on assumptions made by both the analyst in the particular case and the creator of the program about how many individuals are represented in the sample, how the comparison population of unknown persons is composed and the quality of the sample. These calculations

48. *See* Stiffelman, *supra* note 14 (providing an extensive critique of the evidentiary use of the likelihood ratio).

49. BUTLER ET AL., *supra* note 34, at 21.

50. There are two forms of MCMC analysis used in PG software technologies: semi-continuous and fully-continuous. In fully continuous MCMC, the approach used in the most prevalent PG tools, including those discussed later in this paper (STRmix and TrueAllele), the analysis factors the allele peak height and other biological parameters into the calculations, whereas semi-continuous methods do not. Buckleton et al., *supra* note 40, at 394. Both methods are based on estimating the probability of observing the complex DNA profile. Semi-continuous MCMC methods also use a different nuisance parameter (allele dropout). There are perceived benefits to fully continuous MCMC methods for PG because they more effectively use all of the collected data and do not "waste" data that has been collected and reported as part of the sequencing.

may also be affected by the settings chosen for various parameters and thresholds in the software.[51]

B. THE PERILS OF ALLOWING TRADE SECRECY TO IMPEDE DISCLOSURE OF FORENSIC SOFTWARE

Though, as we shall see, courts have largely accepted evidence produced by probabilistic genotyping, there is ongoing disagreement about its reliability, especially for more complex mixtures.[52] In September 2016, the President's Council of Advisors on Science and Technology (PCAST) issued a report about validating forensic methods. The report gave probabilistic genotyping mixed reviews. It concluded that "evidence supports the foundational validity of analysis, with some programs, of DNA mixtures of three individuals in which the minor contributor constitutes at least 20 percent of the intact DNA in the mixture and in which the DNA amount exceeds the minimum required level for the method."[53] The two dominant PG companies, STRmix and TrueAllele, strongly disputed these limitations. STRmix asserted that it had "demonstrate[d] the foundational validity of STRmix™ for complex, mixed DNA profiles to levels well beyond the complexity and contribution levels suggested by PCAST,"[54] while TrueAllele's founder argued that PCAST was attempting to impose "arbitrary limits (e.g., number of contributors) on a scientifically validated solution."[55] In late 2019, the United States Government Accountability Office ("GAO"), Science Technology Assessment, and Analytics, affirmed that PG technology "is not yet fully mature."[56] Some problems the GAO highlighted include the lack of consistency (even when using the same software package) and the lack of outside validation.[57]

Questions highlighting the uncertain validity of probabilistic genotyping software have continued to persist. The National Institutes of Standards and

---

51. A future article from the authors will more fully explore the challenges of validating software-based forensic tools, and, in particular, validating the software used in probabilistic genotyping. That article will focus on two central sets of validation challenges—first, those related to source code validation and second, those related to laboratory-specific implementation and validation. Many probabilistic genotyping validation efforts have primarily focused on lab-specific validation, but we argue that source code validation is also essential to ensuring accurate and valid software performance and results. *See also* Bellovin, *supra* note 8 (discussing the importance and challenges of effective validation of software reliability).

52. Coble & Bright, *supra* note 29, at 222; Buckleton et al., *supra* note 40, at 397.

53. PRESIDENT'S COUNCIL ADVISORS SCI. & TECH., REPORT TO THE PRESIDENT: FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS 82 (2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

54. *Update on STRmix Research in Response to PCAST*, STRMIX (Aug. 18, 2017), https://www.strmix.com/news/update-on-strmix-research-in-response-to-pcast/?acceptCookies=6188d2cdf3a68.

55. Letter from Mark W. Perlin, Cybergenetics Chief Sci. to Dr. John Holdren, PCAST co-chair at 2 (Sept. 16, 2016), https://www.cybgen.com/information/newsroom/2016/sep/files/letter.pdf.

56. U.S. GOV'T ACCOUNTABILITY OFF., GAO-20-306T, SCIENCE AND TECHNOLOGY: OVERVIEW OF GAO'S ENHANCED CAPABILITIES TO PROVIDE OVERSIGHT, INSIGHT, AND FORESIGHT 25 (2019).

57. *Id.*

Technology ("NIST"), a non-regulatory research agency within the United States Department of Commerce, released a draft report in June 2021 re-emphasizing that further information and research are needed to determine the reliability of probabilistic genotyping software.[58] In response, a group of criminal defense attorneys associated with groups including the Legal Society of New York and the Bronx Defenders, called on NIST to impose a moratorium on the use of probabilistic genotyping software until a set of five requirements were met. The requirements focus on having laboratories and developers provide sufficient data for NIST to complete an independent assessment of the reliability of the software and for laboratories to demonstrate that their analysts are proficient dealing with various types of DNA mixtures. The letter also calls for a racial impact assessment to determine how the current use of the software has impacted historically oppressed groups.[59]

PG-based likelihood ratios have the potential to make strong impressions on jurors and judges and are dependent on a number of assumptions.[60] It is thus essential that defendants be given the information they need to probe whether those likelihood ratios are produced in a manner that is trustworthy, accurate, and statistically sound. Indeed, it is clear that assumptions made in implementing and using PG software can affect the results. The PG tools produced by STRmix and TrueAllele have been known to produce significantly different values for the likelihood ratio of the same sample and their founders have sparred publicly about the differences.[61]

Prosecutors and PG companies have routinely opposed defense requests for disclosure of PG source code, asserting that the reliability of these tools can be validated without source code disclosure.[62] In future work, we will analyze the ways in which current evidence doctrine and its judicial implementation fail to demand meaningful validation of probabilistic genotyping in particular, and software-based probabilistic forensic technologies more generally. For present

58. BUTLER ET AL., *supra* note 34, at 89; *see also* Matthews et al., *supra* note 20 (reporting an empirical study of the effects of variations in PG software implementation).

59. Criminal Defense Letter from the Legal Aid Society to Nat'l Inst. Standards & Tech. (Aug. 23, 2021).

60. Coble & Bright, *supra* note 29, at 222.

61. *See* Decision and Order, New York v. Hillary (Hon. Felix Catena, J.) (St. Lawrence Co., Aug. 26, 2016) (excluding STRmix evidence, and explaining that the New York State Police Crime Lab first sent data to Cybergenetics and then, when the results came back inconclusive, sent the data to ESR); *see also* Open Letter from Mark Perlin, Misrepresentation of DNA evidence in *People of New York v. Oral (Nick) Hillary* (July 29, 2016) (on file with authors) (criticizing the methodology of John Buckleton and ESR in interpreting the data in this case); Jesse McKinley, *Judge Rejects DNA Test in Trial Over Garrett Phillips's Murder*, N.Y. TIMES (Aug. 27, 2016), https://www.nytimes.com/2016/08/27/nyregion/judge-rejects-dna-test-in-trial-over-garrett-phillipss-murder.html; Stephanie M. Lee, *People are Going to Prison Over DNA Software—But How It Works is Secret*, BUZZFEED NEWS (Mar. 18, 2016), https://www.buzzfeednews.com/article/stephaniemlee/dna-software-code ("Differing ratios may not always change jurors' minds, like when one method claims a 1 in 5 million chance of being wrong and another claims 1 in 81 billion (as was the case with a rapist in Pennsylvania). But errors in how these ratios are calculated can really matter when two methods end up with wildly different results, like 1 in 420 versus 1 in 18 billion (as was the case in a fatal 2008 shooting.")).

62. *See generally infra* Part III.D.4.

purposes, we demonstrate the value of source code disclosure with a cautionary tale involving the Forensic Statistical Tool, or "FST," a probabilistic genotyping tool developed in house by the New York Office of the Chief Medical Examiner ("OCME").[63] Two early cases diverged as to the admissibility of FST evidence: it was admitted in *People v. Rodriguez* and rejected in *People v. Collins*. The *Collins* decision quickly proved to be an outlier, however, as FST evidence was admitted in other cases throughout New York City,[64] with courts relying explicitly on "the testimony and findings in Rodriguez as settling the questions posed by the defendant"[65] and finding it "not necessary [] to duplicate those efforts."[66]

As things turned out, a duplication of those efforts might have proved worthwhile. In late 2016, Judge Valerie Caproni of the Southern District of New York ordered FST's source code disclosed to a defendant's experts under a protective order.[67] The OCME moved to quash, arguing that its "property rights should be respected."[68] Judge Caproni was not persuaded, however, and the FST source code was released to the defense. After defense expert Nate Adams reviewed the source code, journalists from ProPublica intervened, successfully moving to vacate the protective order.[69] The ProPublica investigation that ensued revealed that, after completing its full validation of FST and bringing it online in New York City labs, the OCME "recoded" portions of the tool to deal with problems encountered in real-world applications and "did not inform the state oversight commission about the change, nor did they run another full validation study on the program."[70] The investigation also uncovered substantial substantive weaknesses in the tool, including that "FST's inventors had acknowledged a margin of error of 30 percent for one key input of the program, and that the program could not take into consideration that family members might share DNA."[71]

The story of FST illustrates two important points. The first is that fieldwork invariably presents cases outside the scope of a tool's initial validation. The ill-

63. CRAIG O'CONNOR, N.Y. OFF. OF THE CHIEF EXAM'R, PROBABILISTIC GENOTYPING: THE USE OF THE FORENSIC STATISTICAL TOOL (FST) (2014), https://www.nist.gov/system/files/documents/forensics/CraigOConnor_DNA-2.pdf (stating that "OCME developed and validated FST.").

64. *See, e.g.*, People v. Carter, No. 2573/14, 2016 WL 239708 (N.Y. Sup. Ct. Jan. 12, 2016).

65. People v. Lopez, 23 N.Y.S.3d 820, 825 (N.Y. Sup. Ct. 2015).

66. *Carter*, 2016 WL 239708, at *3.

67. Order at 1, United States v. Kevin Johnson, No. 15-CR-565 (VEC) (S.D.N.Y. June 7, 2016) (Notably, Judge Caproni added that "[t]he Court is prepared to enter a protective order if OCME wishes, although the Court questions why a public laboratory would need a protective order in this context").

68. Letter from Florence Hunter, General Counsel, Office of the Chief Medical Examiner to Judge Valerie E. Caproni (S.D.N.Y. June 15, 2016) (on file with authors).

69. Lauren Kirchner, *Federal Judge Unseals New York Crime Lab's Software for Analyzing DNA Evidence*, PROPUBLICA (Oct. 20, 2017), https://www.propublica.org/article/federal-judge-unseals-new-york-crime-labs-software-for-analyzing-dna-evidence.

70. *Id.*

71. Lauren Kirchner, *New York City Moves to Create Accountability for Algorithms*, PROPUBLICA (Dec. 18, 2017), https://www.propublica.org/article/new-york-city-moves-to-create-accountability-for-algorithms.

fated, and untested, modification to the FST software was made to paper over the tool's inability to correctly handle just such a case. Because courts were willing to admit FST evidence and deny disclosure based on the original validation studies, while relying on its admission in earlier cases, there was no path for uncovering *later-arising problems* with the validity of the tool. Unfortunately, just such later-arising problems are endemic to complicated software.[72]

Second, the FST story illustrates how secrecy can mask problematic behavior, limitations, and mistakes that are unlikely to be detected by validation studies or black box testing. Validation can be gamed, or simply incomplete, and once a flawed product is on the market and protected by trade secrecy, its developers have incentives to turn a blind eye to, or even cover up, its flaws. Disclosing source code disincentivizes intentional cover-ups and allows defense experts not only to contest explicit and implicit assumptions, but also to expose unwitting mistakes and careless errors.[73] Anticipating disclosure, developers would have much stronger incentives to ferret out, investigate, and correct limitations and errors revealed through applications of their tools in the field.

C. THE PROBABILISTIC GENOTYPING MARKET

Software employing some form of probabilistic genotyping methods has been around for approximately two decades. Today, there are two dominant probabilistic genotyping tools used to produce forensic evidence. TrueAllele, the first of these two to appear on the market, was developed by Dr. Mark Perlin, who founded Cybergenetics in 1994. Cybergenetics initially focused on medical applications of DNA studies, but switched to forensic DNA work in 1999[74] and developed its continuous MCMC PG tool, TrueAllele, in the early 2000s. According to the Cybergenetics website, TrueAllele technology was first adopted for analysis of crime scene evidence in 2004 by the British Forensic Science Service.[75] In 2006, Cybergenetics was awarded a contract to use TrueAllele software to help identify victims of the World Trade Center attack.[76] Three years later, in 2009, evidence obtained using TrueAllele was first admitted

72. Bellovin, *supra* note 8, at 39.

73. *See id.* at 31–35 (describing the importance of adversarial testing for software reliability).

74. *History*, CYBERGENETICS, https://www.cybgen.com/company/history.shtml (last visited Mar. 21, 2022).

75. Press Release, Cybergenetics, Cybergenetics Accelerates the UK National DNA Database (July 11, 2001), https://www.cybgen.com/information/press-release/2001/Cybergenetics-Accelerates-the-UK-National-DNA-Database/page.shtml.

76. *Cybergenetics Awarded Contract to Identify World Trade Center Victim Remains Using TrueAllele Technology*, CYBERGENETICS (Sept. 8, 2006), https://www.cybgen.com/information/press-release/2006/Cybergenetics-Awarded-Contract-to-Identify-World-Trade-Center-Victim-Remains-Using-TrueAllele-Technology/page.shtml.

in the U.S. in a Pennsylvania case.[77] The Cybergenetics website reports uses of TrueAllele in about fifteen cases per year since 2015.[78]

Development of STRmix began in 2010, as a joint project of two government-funded laboratories: Forensic Science South Australia (FSSA) and the forensic arm of New Zealand's Institute of Environmental Science and Research (ESR). The program was first used for casework in Australia and New Zealand in 2012. Dr. John Buckleton, the senior member of the STRmix team, is a consummate forensic science insider, whose "caseworking experience covers 33 years in the United States, Australia, the Netherlands, the United Kingdom and New Zealand."[79] STRmix evidence was first accepted by a U.S. court in 2015 and, according to Buckleton's blog, has been the subject of at least 17 admissibility hearings in North America.[80]

Crime labs in a given state (when there is more than one) all tend to adopt the same tool. Currently, the PG market in the United States is divided almost exclusively between states that primarily use TrueAllele and states that primarily use STRmix. States that began using TrueAllele before STRmix entered the market tend to have stuck with it (at least so far), while STRmix has become the favorite of later adopters, which are now in the majority. As discussed in Subpart B, New York used the state-developed "FST" for several years. When that tool was discredited after its source code was disclosed, New York laboratories adopted STRmix. Strikingly, while several open-source versions of PG software are available, only one state (Colorado) has employed an open source PG tool (Lab Retriever) to any considerable degree, and it has more recently adopted STRmix as well.

The developers of TrueAllele, STRmix, and FST have all argued on numerous occasions, in response to defense requests, that their source code and other implementation details constitute trade secrets.[81] Courts have routinely endorsed these arguments, denying defense requests for disclosure on that basis. While STRmix has repeatedly resisted court-ordered production of source code, it does allow defense experts to gain sharply limited access to the JavaScript code, governed by a confidentiality agreement and carried out under restrictive conditions.[82]

Because crime labs purchase or license these tools for the purpose of producing admissible evidence, one way to follow the growth of the market is

77. Commonwealth v. Foley, 38 A.3d 882 (Pa. Super. Ct. 2012).

78. *Cases Where Cybergenetics Testified About TrueAllele® Evidence*, CYBERGENETICS, https://www.cybgen.com/news/trials.shtml (last visited Mar. 21, 2022).

79. *John Buckleton*, GOVERNING, http://www.governing.com/authors/John-Buckleton.html (last visited Mar. 21, 2022).

80. John Buckleton, *STRmix*, WORDPRESS, https://johnbuckleton.wordpress.com/strmix (last visited Mar. 21, 2022).

81. *See*, *e.g.*, State v. Superior Court (Chubbs), No. B258569, 2015 WL 139069, at *2 (Cal. Ct. App. Jan. 9, 2015); People v. Wakefield, 9 N.Y.S.3d 540, 543 (N.Y. Sup. Ct. 2015).

82. STRMIX, ACCESS TO STRMIX™ SOFTWARE BY DEFENCE LEGAL TEAMS (2016), https://www.strmix.com/assets/Uploads/Defence-Access-to-STRmix-April-2016.pdf.

by tracing references in judicial opinions and orders. These opinions usually address defense requests for disclosure of the source code and other implementation details in the context of admissibility determinations and/or assertions of a trade secret privilege.[83] We have collected representative opinions and orders through searching Lexis, Westlaw, and other online sources based on suggestions from defense attorneys active in this area. We include opinions addressing both admissibility and trade secret privilege because they address similar disclosure-related questions and also because courts often cite rulings of one sort in their decisions about the other.[84] In the vast majority of such cases, with a few recent exceptions, courts have upheld trade secrecy and denied disclosure to defense experts.[85] As we discuss in further detail in Part III, judges considering these issues frequently rely heavily (or even exclusively) on previous decisions, often from outside jurisdictions.[86] These references create a network that tracks adoption and is also generally representative of the evolution of the market. Figures 2 and 3 show that network for cases up through 2021.

83. *See generally infra* Part III.D.4.

84. We believe we have been reasonably thorough, but cannot guarantee that our collection is comprehensive. While many of the court orders discussed in this Article are unreported and accessible only through tools such as PACER or by request to local courts, PDF copies of the orders cited here are on file with the authors unless otherwise noted.

85. *See infra* Part III.

86. *See infra* Part III.

FIGURE 2: GEOGRAPHIC DISTRIBUTION OF OBSERVED NETWORK EFFECT

Figure 2. This figure shows the geographic distribution of the observed network effect. Cases are superimposed on the originating jurisdiction. Connecting edges illustrate the citing relationship between the cases. Blash is plotted off of the continental United States, as it is a United States Virgin Islands Superior Court decision. Foley refers to the 2012 decision, and Foley (2) refers to the 2009 decision. The case names are color-coded relative to the PG algorithm. Black: TrueAllele; Maroon: STRMix; Blue: FST. The network was created using Cytoscape version 3.9.1. [87]

87. In addition to the cases cited in this article, Figure 2 also includes *United States v. Gissantaner*, 990 F.3d 457 (6th Cir. 2021); *State v. Bah*, No.17CR00938 (Ga. Super. Ct. Oct. 23, 2019); *State v. Battle*, No. A17A1753 (Ga. App. Ct. May 31, 2017); and Order, *State v. Sewell*, No. 17CR01675 (Ga. Super. Ct. Aug. 7, 2019).

FIGURE 3: INTERACTION NETWORK OF CITING AND CITED CASES

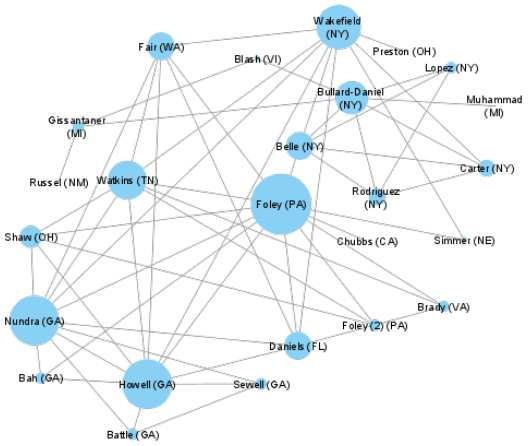


Figure 3. This figure illustrates the interaction network of citing and cited cases and illustrates the observed network effect. The nodes (corresponding to cases) are scaled according to the number of times the case was cited relative to the other cases in the interaction network. Connecting edges illustrate the citing relationship between the cases. Foley refers to the 2012 decision, and Foley (2) refers to the 2009 decision. The network was created using Cytoscape version 3.9.1.[88]

II. TRADE SECRECY, INNOVATION AND MARKETS FOR FORENSIC EVIDENCE TECHNOLOGY

There are two primary justifications for legal trade secrecy protection, each premised on avoiding a different sort of market failure.[89] First, like other forms of intellectual property protection, trade secrecy law is justified partly as a mechanism for promoting innovation.[90] Under this rationale, trade secrecy preserves innovation incentives by providing a period of market exclusivity during which an innovator can recoup R&D investments without fear of

88. In addition to the cases cited in this article, Figure 3 also includes *Gissantaner*, 990 F.3d 457; *Bah*, No. 17CR00938; *Battle*, No. A17A1753; Order, *Sewell*, No. 17CR01675.

89. Government secrecy regarding certain investigatory techniques and procedures is also sometimes justified by fears of "gaming the system." *See, e.g.*, 5 U.S.C. § 552(b)(7)(E) (2016). One of us has critiqued the scope of such assertions elsewhere. *See* Ignacio Cofone & Katherine J. Strandburg, *Strategic Games and Algorithmic Secrecy*, 64 MCGILL L.J. 623, 627 (2019). In any event, this argument is distinct from, and of entirely different legal and analytical scope than, the trade secrecy privilege assertions at issue here.

90. *See* Robert G. Bone, *Trade Secrecy, Innovation and the Requirement of Reasonable Secrecy Precautions*, *in* THE LAW AND THEORY OF TRADE SECRECY: A HANDBOOK OF CONTEMPORARY RESEARCH 46, 64 (Rochelle C. Dreyfuss & Katherine J. Strandburg eds., 2011); Serge Pajak, *Do Innovative Firms Rely on Big Secrets? An Analysis of IP Protection Strategies with the CIS 4 Survey*, 25 ECON. INNOVATION & NEW TECH. 516, 528 (2016); Michael Risch, *Why Do We Have Trade Secrets?*, 11 MARQ. INTELL. PROP. L. REV. 1, 26 (2007) (explaining that promoting innovation is a minor, but extant, justification for trade secrecy).

competition from free riding copyists. In fact, the period of market exclusivity, and the corresponding market advantage, extends as long as the trade secret is kept secret.[91]

Second, and unlike other forms of intellectual property, trade secrecy has another primary goal of regulating market behavior by punishing and deterring unethical mechanisms for obtaining information from competitors, thereby also avoiding (or at least diminishing) wasteful investments in an economic espionage arms race.[92] This distinctive mission emerges because trade secrets are largely a matter of self-help.[93] Commercial actors can, and do, use a whole range of practical mechanisms, as well as contractual non-disclosure agreements, to control access to and disclosure of economically valuable secrets. Trade secret law merely provides back-up protection to these other measures. By contrast, the scope of copyright and patent protections is under legislative control and, though doctrinal controversies abound, there is widespread agreement that the goal is to balance between the innovation incentives created by market exclusivity and the offsetting social costs borne by consumers and follow-on innovators.[94] Because of its dual mission and back-up stance, trade secrecy doctrine cannot even attempt such a nuanced balancing act.

This Part argues that trade secret privileges for forensic evidence technology are largely unjustified by the traditional rationales for trade secrecy law because court-ordered disclosure departs from the ordinary context of trade secrecy law in two highly significant ways. First, trade secrecy law's role in regulating undesirable market behavior is irrelevant to court-ordered disclosure. Second, the distinctive characteristics of forensic evidence technology markets largely undercut the need for trade secrecy to preserve incentives for innovation against free-riding competitors.

---

91. *See* Bone, *supra* note 90, at 73; David S. Levine, *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 FLA. L. REV. 135, 145 (2007); U.S. DEP'T OF COMMERCE, *How Long Does Patent, Trademark or Copyright Protection Last?*, STOPFAKES.GOV (Feb. 25, 2021), https://www.stopfakes.gov/article?id=How-Long-Does-Patent-Trademark-or-Copyright-Protection-Last.

92. *See* RESTATEMENT (FIRST) OF TORTS § 757 cmt. a (AM. L. INST. 1939) ("[T]he liability rests upon . . . breach of contract, and abuse of confidence or impropriety in the method of ascertaining the secret.").

93. *See* Bone, *supra* note 90, at 46 (explaining how the Uniform Trade Secrets Act makes taking reasonable precautions to maintain secrecy an essential element in an enforceable trade secret.).

94. *See, e.g.*, Bonito Boats v. Thunder Craft Boats, 489 U.S. 141, 146 (1989) ("The Patent Clause itself reflects a balance between the need to encourage innovation and the avoidance of monopolies which stifle competition without any concomitant advance in the 'Progress of Science and useful Arts.' As we have noted in the past, the Clause contains both a grant of power and certain limitations upon the exercise of that power."); Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975) ("The limited scope of the copyright holder's statutory monopoly, like the limited copyright duration required by the Constitution, reflects a balance of competing claims upon the public interest . . . The immediate effect of our copyright law is to secure a fair return for an 'author's' creative labor. But the ultimate aim is, by this incentive, to stimulate artistic creativity for the general public good.").

A. LEGAL TRADE SECRECY PROTECTION'S SCOPE AND PURPOSES

Trade secret protection is a longstanding feature of state (and colonial) law and, since Congress enacted the Defend Trade Secrets Act in 2016, has also found a place in the U.S. Code.[95] While definitions vary somewhat, trade secrecy laws generally protect a wide variety of commercial information as long as it (1) is actually secret, in that it is not generally known or readily ascertainable to others who can obtain economic value from it, (2) is more valuable because it is secret, and (3) is protected by reasonable secrecy preservation measures.[96]

Significantly for present purposes, trade secrecy laws proscribe only "misappropriation," defined to include acquisition by "improper means" and culpable downstream uses and disclosures. The misappropriation requirement, along with the requirement of actual secrecy, distinguish trade secrecy from copyright, which proscribes all unauthorized copying, regardless of means or intent, and patent, which penalizes even independent invention.[97] "Improper means" generally include typical forms of economic espionage and theft, as well as violations of employment policies and non-disclosure agreements. Reverse engineering and independent invention, however, are legitimate means for obtaining previously secret information.[98] These mechanisms are consistent with trade secrecy law's focus on misappropriation and are viewed as crucial limitations on trade secrecy's downstream social costs. They are, however, crude mechanisms for tailoring the scope of protection when compared with the detailed scope tailoring and defenses embodied in copyright and patent law.

B. MISAPPROPRIATION AND COURT-ORDERED DISCLOSURE

Trade secret law's misappropriation element sets it apart from other intellectual property liability regimes, which generally do not turn on defendants' wrongful behavior. The misappropriation requirement recognizes that secrecy is first and foremost a practical tool for market actors, which does not depend on a legal entitlement. The rationale for legal protection against misappropriation was expressed at length by the Supreme Court in its 1974 opinion *Kewanee Oil Co. v. Bicron Corp.*, which held that state trade secrecy law was not preempted by federal patent law:

> [Abolishing trade secret protection] would [lead to] an increase in the amount of self-help that innovative companies would employ. Knowledge would be widely dispersed among the employees of those still active in research. Security precautions necessarily would be increased . . . . Smaller companies

95. Defend Trade Secrets Act, 18 U.S.C. §§ 1836–1839 (2016).

96. *See* UNIF. TRADE SECRETS ACT § 1.4 (UNIF. L. COMM'N 1985).

97. *See, e.g.*, Oskar Liivak, *Rethinking the Concept of Exclusion in Patent Law*, 98 GEO. L.J. 1643, 1657–74 (2010) (critiquing patent on this ground). Copyright and, to a lesser extent patent, laws do have important scope limitations and exceptions, but both are essentially strict liability and do not invoke conceptions of improper means.

98. *See* Jonathan R. Chally, *The Law of Trade Secrets: Toward A More Efficient Approach*, 57 VAND. L. REV. 1269, 1284–86 (2004).

> would be placed at a distinct economic disadvantage, since the costs of this kind of self-help could be great, and the cost to the public of the use of this invention would be increased.
>
> . . .
>
> Nothing in the patent law requires that States refrain from action to prevent industrial espionage. In addition to the increased costs for protection from burglary, wiretapping, bribery, and the other means used to misappropriate trade secrets, there is the inevitable cost to the basic decency of society when one firm steals from another.[99]

The misappropriation justification is, however, irrelevant to court-ordered disclosure, which obviously does not involve "misappropriation"[100] and will not provoke wasteful investments in an economic espionage arms race. If anything, concerns about wasteful investment cut the other way in this context; the *availability* of trade secrecy defenses to disclosure leads to litigation over their applicability, with its attendant transaction costs.

Moreover, trade secrecy protection is routinely limited by disclosure mandates in a wide variety of regulatory contexts. Some such regulations mandate public disclosure, often for the purpose of informing consumer purchasing decisions.[101] Many other regulations mandate disclosure to government regulatory bodies. For the most part, these disclosure mandates are motivated by concerns about market failure due to information asymmetries between suppliers and consumers. Court-ordered disclosures of the workings of forensic evidence technology may have loftier goals related to the legitimacy of the justice system, but they also can prevent similar forms of demand-side market failure, as discussed in Subpart E.[102]

Because the misappropriation justification is inapposite, any justification for trade secret privileges for forensic evidence technology must turn on whether they promote socially valuable innovation. Indeed, judges denying defense requests for disclosures about probabilistic genotyping tools clearly rely on this assumption. Thus, the Pennsylvania Superior Court's seminal decision in *Foley* relies on the assumption that "TrueAllele is proprietary software; it would not be possible to market TrueAllele if it were available for free."[103] Similar

---

99. Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 485–87 (1974).

100. Trade secret law's definitions do, however, govern the scope of information for which trade secret privileges are potentially available. *See, e.g.*, CAL. EVID. CODE § 1061(a) (setting the definition of "trade secret" for purposes of evidentiary privilege equal to the standard definition in Cal. Civil Code § 3426.1(d)). Nonetheless, in practice, cases involving trade secret privileges for black box evidence devote surprisingly little attention to determining which information about the technology actually qualifies as trade secret.

101. Laws requiring disclosure of ingredients in food, drugs and cosmetics are of this nature. *See, e.g.*, 21 U.S.C. §§ 321–92 (requiring labeling of cosmetic ingredients); 21 U.S.C. § 343 (explaining definition of misbranded food); *see also* N.Y. ENV'T CONSERV. LAW § 35-0107 (McKinney 1972) (requiring disclosure of ingredients in household cleaning products).

102. *See infra* Part III.E.

103. Commonwealth v. Foley, 38 A.3d 882, 889 (Pa. Super. Ct. 2012).

sentiments are echoed in many other opinions dealing with TrueAllele, STRmix, and even the state-owned FST.[104]

C. COPYRIGHTS, PATENTS AND TRADE SECRETS, OH MY!

Copyrights and patents are expected to serve their constitutional purpose of "promot[ing] the progress of science and useful arts,"[105] by providing a limited period of market exclusivity. Their rationale posits that, in the absence of intellectual property protection, competitors can cheaply copy technological inventions or creative works.[106] Because free riding competitors do not have to make their own R&D or creative investments, they can charge low prices that undercut innovators' ability to recoup such investments. Anticipating competition from free riders, potential authors and inventors will presumably be deterred from investing, and creating, in the first place.[107] This situation produces a market failure if consumers would have been willing to pay the higher prices necessary to cover the R&D costs. Patent law also aims to promote progress by requiring disclosure of technological advances as a "quid pro quo" of patent exclusivity.[108]

Despite trade secrecy's long legacy, its role in intellectual property law has always been a bit puzzling and, as a result, controversial. While legal trade secrecy protection presumably extends the market exclusivity afforded by practical secrecy, secrecy is in tension with patent law's strong emphasis on promoting disclosure of new inventions. Moreover, because trade secret law encompasses secret information that would not qualify for patent (or copyright) protection, it is also in some tension with patent doctrine's "notion that [unpatentable] concepts within the public grasp, or those so obvious that they readily could be, are the tools of creation available to all."[109] Above and beyond these foundational questions about the social costs of secrecy, there has also been considerable scholarly debate about whether, even in ordinary commercial

104. *See, e.g.*, People v. Lopez, 23 N.Y.S.3d 820, 829 (N.Y. Sup. Ct. 2015) ("To the extent they claim it would be easier to perform the calculations with the actual program software, the computer program itself is proprietary and the Court is not ordering its disclosure.").

105. U.S. CONST. art. I, § 8, cl. 8.

106. *See, e.g.*, JAMES BOYLE, THE PUBLIC DOMAIN: ENCLOSING THE COMMONS OF THE MIND 47–48 (2008) (explaining the idea of copyright as a response to the "tragedy of the commons").

107. *See, e.g.*, Yafit Lev-Aretz & Katherine J. Strandburg, *Regulation and Innovation: Approaching Market Failure from Both Sides*, 38 YALE J. REGUL. ONLINE BULL. 1, 3 (2020) https://digitalcommons.law.yale.edu/jregonline/2.

108. *See* NAT'L RSCH. COUNCIL, A PATENT SYSTEM FOR THE 21ST CENTURY 36 (Stephen A. Merrill, Richard C. Levin & Mark B. Myers eds., 2004) ("The quid pro quo for giving the patent holder the right to exclude others is to compel disclosure of the invention in terms that enable others to replicate, modify, and circumvent it."); Timothy R. Holbrook, *Possession in Patent Law*, 59 S.M.U. L. REV. 123, 131–32 (2006).

109. Bonito Boats v. Thunder Craft Boats, 489 U.S. 141, 156–57 (1989).

contexts, trade secrecy protection is overbroad in light of tradeoffs between the social costs and benefits of market exclusivity.[110]

Hovering over this debate about trade secrecy doctrine is the question of whether innovators should be forced to rely on more carefully tailored patent and copyright protections, rather than trade secrecy law, to deter free riders. Though the Supreme Court confronted this puzzle to some extent in *Kewanee Oil Co. v. Bicron Corp.*,[111] its holding rested heavily on the misappropriation-related concerns expressed in the quote above, while its analysis of trade secrecy as an innovation promoter was arguably both empirically dubious[112] and analytically questionable in light of the Court's own later pronouncements.[113]

In the ordinary commercial context, the misappropriation and free-rider-based justifications for trade secrecy are unavoidably intertwined because those who produce trade secret information would have the option to rely on practical secrecy even if legal trade secrecy protection were not available. The policy question is thus not whether trade secrecy itself is a good idea but whether legal trade secrecy protection is a socially beneficial supplement to practical secrecy. Practical secrecy is not an option, however, for many innovations and essentially all expressive works, which are self-disclosing once they are put on the market.[114] In these run-of-the-mill situations, we assume that the policy balances enshrined in patent and copyright doctrines, which cover some aspects of some products and deny protection to others, are sufficient.

It is thus worth bearing in mind that court-ordered disclosure of the workings of forensic evidence tools would leave their creators no worse off than the many creators of self-disclosing innovations. Innovative forensic evidence tools are protectable by copyrights and patents to the same extent as other technologies. Software source code, for example, is covered by copyright to the

110. *See generally* Bone, *supra* note 90; Robert G. Bone, *The (Still) Shaky Foundations of Trade Secret Law*, 92 TEX. L. R. 1803 (2014). THE LAW AND THEORY OF TRADE SECRECY: A HANDBOOK OF CONTEMPORARY RESEARCH 490 (Rochelle C. Dreyfuss & Katherine J. Strandburg, eds., 2012).

111. Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470 (1974).

112. *See, e.g.*, *id*. at 487–90, where the Court asserts that inventors of patentable or potentially patentable technologies will not opt to rely on trade secrecy because "trade secret law provides far weaker protection."

113. *Compare, e.g.*, *Kewanee*, 416 U.S. at 484 ("Certainly the patent policy of encouraging invention is not disturbed by the existence of another form of incentive to invention"), *with Bonito Boats*, 489 U.S. at 156–57 ("Both the novelty and the nonobviousness requirements of federal patent law are grounded in the notion that concepts within the public grasp, or those so obvious that they readily could be, are the tools of creation available to all . . . . Moreover, through the creation of patent-like rights, the States could essentially redirect inventive efforts away from the careful criteria of patentability developed by Congress over the last 200 years.").

114. We note that the core of some innovations remain secret even when the item is put on the market. A classic example of this is a recipe—like the secret formula for Coke. That is, the introduction of the product to the market does not provide information on the hidden contents of the product, which can remain protected by trade secret. In contrast, other innovations are self-disclosing when put on the market, such as writings or products that can be easily reverse engineered. For this second group of innovations, secrecy is necessarily destroyed by putting the innovative item on the market. Katherine J. Strandburg, *What Does the Public Get: Experimental Use and the Patent Bargain*, 2004 WIS. L. REV. 81, 105–07 (2004).

extent it contains protectable expression,[115] meaning that potential competitors could not simply copy and adopt it wholesale, but would probably have to engage in significant recoding. Patent protection, while recently reined in by the Supreme Court, also remains available for certain sorts of software-related inventions.[116] In fact, since court-ordered disclosure would presumably apply to all players in forensic technology markets, it would be particularly easy for holders of copyright and patent rights to enforce those protections by monitoring their competitors' mandated disclosures. There is longstanding debate about the desirability and adequacy of copyright and patent protections for software.[117] Whatever one's perspective on this general debate, there seems no reason to expect that patent and copyright protections are *distinctively* inadequate to incentivize forensic evidence tool innovation.

D. TRADE SECRECY, FREE RIDERS AND FIRST MOVER ADVANTAGES IN MARKETS FOR FORENSIC EVIDENCE TECHNOLOGY

This Subpart explores how the characteristics of forensic evidence technology markets tend to further undermine the force of the free rider justification for trade secrecy by prolonging first mover exclusivity. The distinctive first mover advantages in these markets arise from the fact that they are driven by customer demand for judicial admissibility. This Subpart uses our case study of PG as a springboard for analyzing these distinctive characteristics and their implications for the free rider justification for trade secrecy.

From an intellectual property perspective, market exclusivity for innovators should be sufficient to allow them to recoup their free-rideable investments. Beyond that, however, the goal is to cap exclusivity so as to promote—not avoid—healthy market competition and follow-on innovation.[118] Free-rideable investments are those, such as R&D expenses, that competitors can avoid by simply copying the innovator. Other sorts of investments—in raw materials, building a manufacturing plant and so forth—are not free-rideable. Moreover, not all free-rideable investments require legal protection. Even in the absence of intellectual property protections, first movers ordinarily enjoy some period of market exclusivity as a result of factors such as the time it takes competitors to ramp up production and convince consumers to purchase the new product. In some markets, the first mover exclusivity period is further prolonged by various sorts of barriers to entry.[119] First mover advantages alone are often sufficient to allow innovators to recoup their free-rideable investments without

115. 17 U.S.C. § 117; *see also* Apple Computer, Inc. v. Franklin Computer Corp., 714 F.2d 1240, 1247 (3d Cir. 1983) (declaring computer software code to have the same copyright protection as literary works).

116. *See, e.g.*, U.S. Patent No. 8,898,021 (filed Feb. 2, 2001) (TrueAllele patent).

117. *See, e.g.*, Meghan J. Ryan, *Secret Algorithms, IP Rights, and the Public Interest*, 21 NEV. L.J. 61 (2020); JAMES BESSEN AND MICHAEL J. MEURER, PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK (2009).

118. *See* Bonito Boats v. Thunder Craft Boats, 489 U.S. 141, 146 (1989).

119. *See* Lev-Aretz & Strandburg, *supra* note 107.

the need for further protection. Indeed, this point is a primary justification for patent law's refusal to award patents to "obvious" inventions.[120]

To understand whether a trade secrecy privilege is important for promoting innovation in forensic evidence technology, we thus need to understand (1) the extent to which the trade secret privilege covers costly free-rideable investments that are not protectable by copyrights and patents and (2) the sources of first mover exclusivity in those markets. Court-ordered disclosure will not threaten innovation incentives if first mover exclusivity from other sources, combined with copyright and patent protection, is sufficient to allow innovators to recoup their free-rideable investments.

Forensic evidence technology markets are driven by admissibility doctrines and judicial practices. These driving forces create distinctively robust first mover exclusivity mechanisms, over and above the conventional sources. Admissibility doctrines also cabin the sorts of trade secrets that suppliers of these tools can maintain. Taken together, these features lessen the chance that court-ordered disclosure will be the last straw that deters innovation.

1. Admissibility Drives Markets for Forensic Evidence Technology

Anyone seeking to enter any market must assess current or potential customer demand and determine how to meet that demand. The direct customers for forensic evidence technology are forensic laboratories. Their demand piggybacks primarily on the demand of prosecutors and law enforcement for tools that will produce admissible evidence that will lead to convictions. Admissibility in court is thus critical to a forensic evidence technology's market viability. With the possible exception of certain trend-setting federal laboratories, those who procure forensic evidence tools will strongly prefer to purchase tools that they can be confident will produce admissible results.

Purveyors of probabilistic genotyping tools clearly recognize the importance of admissibility to marketability. For example, an entire page on the website of Cybergenetics, the corporate home of TrueAllele, is devoted to documenting admissibility opinions and orders signed by trial judges from around the country,[121] providing potential customers with up-to-date admissibility precedent. Similarly, the STRmix website and the blog of STRmix

120. *See, e.g.*, Graham v. John Deere, 383 U.S. 1, 11 (1966) ("The inherent problem [underlying the nonobviousness requirement] was to develop some means of weeding out those inventions which would not be disclosed or devised but for the inducement of a patent."); Robert P. Merges, *Uncertainty and the Standard of Patentability*, 7 HIGH TECH. L.J. 1, 31 (1992) ("In a recent study of a large number of companies, a team of economists found that in most industries advantages associated with a head start, including establishment of production and distribution facilities, and moving rapidly down a learning curve, were judged significantly more effective than patents in enabling a firm to reap returns from innovation.").

121. *TrueAllele Admissibility*, CYBERGENETICS, https://www.cybgen.com/information/admissibility/page.shtml (last visited Mar. 21, 2022).

creator John Buckleton are regularly updated with posts reporting favorable admissibility decisions.[122]

The admissibility of scientific evidence is governed by two different standards:[123] some state courts continue to employ the *Frye* standard, based on a 1923 D.C. Circuit decision,[124] while federal courts and a majority of state courts now apply the more multi-faceted *Daubert* standard, first enunciated by the Supreme Court in 1993[125] as an interpretation of Rule 702 of the Federal Rules of Evidence.[126] The *Frye* standard looks to whether a technique is generally accepted in the "relevant scientific community,"[127] while the *Daubert* rule purports to position judges as the "gatekeepers" of scientific evidence,[128] who are to employ factors that include whether the technique has been "subjected to peer review and publication" and whether it is "generally accepted in the scientific community." Importantly, under both *Daubert* and *Frye*, validation studies, especially those published in peer-reviewed journals, play a central role in the admissibility determination. Such studies appear to address both *Frye*'s over-riding concern with community acceptance and *Daubert*'s over-arching concern with accuracy and reliability. (Of course, it is possible that validation studies also play a more direct role in signaling the quality of the tool to potential adopters.)

The importance of admissibility to marketability tends to enhance first mover exclusivity and suppress the extent to which disclosure facilitates free riding in three ways: First, because general principles must be disclosed to meet admissibility standards and software is protected by copyright, disclosure of

122. *STRmix*, NICHEVISION, https://nichevision.com/strmix (last visited Mar. 21, 2022) ("Additionally, there have been at least 22 successful admissibility hearings for STRmix™ in the U.S."); *News*, STRMIX, https://www.strmix.com/#news (last visited Mar. 21, 2022) (including posts detailing state court admissibility decisions).

123. Bruce Kaufman, *States Slow to Adopt Daubert Evidence Rule*, BLOOMBERG L. (Apr. 26, 2016) https://news.bloomberglaw.com/environment-and-energy/states-slow-to-adopt-daubert-scientific-evidence-rule ("More than two decades after the U.S. Supreme Court adopted the *Daubert* test for evaluating the reliability of scientific evidence in federal courtrooms, nearly a quarter of the states have retained their own standards. In many of the holdout jurisdictions—including California, New York, New Jersey, Illinois, Maryland, Washington and the District of Columbia—the standard for admitting expert evidence in courtrooms closely follows the century-old *Frye* test, which was developed for evaluating then-novel polygraph testimony.").

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

125. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993).

126. FED. R. EVID. 702. Rule 702 was amended in 2000, following the Supreme Court's ruling in *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999), clarifying that the standard applies to all expert testimony. The 2000 amendment "affirms the trial court's role as gatekeeper and provides some general standards that the trial court must use to assess the reliability and helpfulness of proffered expert testimony." FED. R. EVID. 702 advisory committee's notes to 2000 amendment.

127. *See* Kaufman, *supra* note 123; *Frye*, 293 F. at 1014.

128. *See Daubert*, 509 U.S. at 597. The opinion set out what it characterized as a nonexclusive set of factors that "bear on the inquiry[:]" (1) whether the technique "can be (and has been) tested," (2) whether the technique has been "subjected to peer review and publication," (3) the technique's "known or potential rate of error," (4) "the existence and maintenance of standards controlling the techniques operation" and, folding in the old *Frye* test, (5) the degree of acceptance in the relevant expert community. *See id.* 593–94. These factors now dominate the admissibility analysis in *Daubert* jurisdictions.

source code and implementation details provides only limited opportunities for competitor free riding. Second, once one tool is deemed admissible, the needs to implement and validate a new tool and demonstrate its admissibility in court creates switching costs for customers. Third, and perhaps most significantly, judges routinely treat previous decisions about admissibility or trade secret privilege for a given product as highly persuasive precedent, even when those decisions are from other jurisdictions and are sparsely reasoned.[129]

2. Admissibility and Limits on Free-Rideable Secrets

Court-ordered disclosure of trade secrets can only facilitate problematic free riding to the extent that it reveals secrets that reflect costly, free-rideable investments in innovation. Scientific evidence doctrine cabins this possibility from both directions.

On the one hand, the emphasis on published validation studies and the general scope of the admissibility inquiry limits the extent to which evidence technology innovators can rely on trade secrecy regarding truly innovative tools. Both courts and validation study peer reviewers will demand disclosure of underlying principles and methods and general implementation descriptions.[130] Because these basics must be disclosed for purposes of admissibility, they are not protectable by trade secrecy.[131]

That leaves a window of free-rideable investment corresponding to the cost of "development," that is, determining how to implement the underlying principles in a useable tool. The size of that cost, both in absolute terms and relative to the overall investment required to enter the market, depends on specifics. For PG, as discussed above, the general principles and methods of probabilistic genotyping are publicly available,[132] while trade secrecy is asserted in source code, parameters and the like.[133] Court-ordered disclosure of these sorts of secrets would, at most, save competitors the cost of doing their own coding and parameter selection. Of course, companies will argue that these remaining secrets embody the "secret sauce" they depend upon for competitive advantage, but it is unclear whether they truly involve substantial innovation (or investment).

For source code, the potential for free riding on court-ordered disclosure is narrowed still further by copyright protection, which accrues automatically.

129. *See, e.g.*, State v. Simmer, 935 N.W.2d 167, 181 (Neb. 2019); People v. Lang, No. F075921, 2019 WL 5205997, at *6 (Cal. Ct. App. Oct. 16, 2019); People v. Belle, 16 N.Y.S.3d 793 at *4 (N.Y. Sup. Ct. 2015).

130. *See supra* Part II.D.1.

131. In any event, it seems as though such basic innovations often are taken from academic research or other public sources.

132. *See supra* Part I.A.

133. *See* State v. Superior Court (Chubbs), No. B258569, 2015 WL 139069, at *2 (Cal. Ct. App. Jan. 9, 2015) ("As pertinent here, the People explained that they requested the source code from Cybergenetics, but Cybergenetics did not turn it over because it is a trade secret. The People argued that disclosure of the source code would be 'financially devastating' to Cybergenetics.").

Though copyright does not cover software's functionality, and it is hard to predict how courts will interpret its scope in particular programs, it certainly precludes rote copying of source code and requires competitors to engage in significant re-coding.[134]

On the flip side, copying an innovator's technology does not exempt competitors from the need to validate their own tools for admissibility purposes. While courts do put some weight on prior validation of general methods, and sometimes even of related tools, validation costs are at best partially free-rideable. Thus, even if a later competitor claims to employ a previously validated method, the new implementation of the method ordinarily will still have to be validated by separate, preferably published, studies. Validation studies will ordinarily be both time-consuming and costly. The time required for peer review and publication alone can be substantial and thus extend the first mover exclusivity period.[135] In fact, the validation process may even be more costly and slower for latecomers than for first movers. Validation studies often require specialized equipment, samples or data that are most cheaply accessible from forensic laboratories or law enforcement sources. Those players may have few incentives to cooperate with second comers, especially if they are not forensic insiders.

Aspects of the market for probabilistic genotyping tools support this analysis. Despite trade secrecy protection of TrueAllele and STRmix source code, there are several open-source tools available, suggesting that implementing basic PG methods in code is not terribly costly.[136] Nonetheless, it is private, closed-source technology that is dominating the marketplace, and no open-source tools appeared in our case study into lower-court admissibility decisions. This situation at least suggests that barriers to entry associated with validation costs are significant. Notably, John Buckleton, the force behind STRmix's successful competitive entry into the PG market, is a consummate forensic insider, who undoubtedly had an inside track to validation resources.

In sum, admissibility standards for forensic evidence limit the scope of potentially free-rideable information that could be covered by the trade secret

134. The extent of re-coding required would depend on the amount of protectable "expression" contained with the code. The amount of re-coding required depends on specifics and involves a fairly complicated copyright analysis. It is, however, probably safe to say that re-writing a competitor's code to avoid copyright infringement is a substantial undertaking, probably requiring legal consultation, particularly when it is virtually certain that the originator will have easy access to the new version for purposes of enforcement.

135. *See* Vivian M. Nguyen, Neal R. Haddaway, Lee F. G. Gutowsky, Alexander D. M. Wilson, Austin J. Gallagher, Michael R. Donaldson, Neil Hammerschlag & Steven J. Cooke, *How Long Is Too Long in Contemporary Peer Review? Perspectives from Authors Publishing in Conservation Biology Journals*, 10 PLOS ONE 1, 2 (Aug. 12, 2015), https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0132557 (explaining that peer review "may still stretch into months or even years").

136. *See, e.g.*, Øyvind Bleka, Mayra Eduardoff, Carla Santos, Chris Phillips, Walther Parson & Peter Gill, *Open Source Software EuroForMix Can Be Used to Analyse Complex SNP Mixtures*, 31 FORENSIC SCI. INT'L GENETICS 105 (2017); Sho Manabe, Chie Morimoto, Yuya Hamano, Shuntaro Fujimoto & Keiji Tamaki, *Development and Validation of Open-Source Software for DNA Mixture Interpretation Based on a Quantitative Continuous Model*, 12 PLOS ONE 1 (Nov. 17, 2017), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5693437.

privilege and, concomitantly, the impact that disclosure of trade secrets can have on innovation incentives. Meanwhile, all market entrants confront substantial non-free-rideable validation costs, which may even be heightened for later entrants. Court-mandated disclosure of the source code, parameters and similarly specific implementation details that are currently protected by trade secret privileges are unlikely to provide much advantage to free riding competitors. Notably, however, such disclosure can have significant public payoffs in uncovering problematic implementations, as illustrated by the FST debacle described above.[137] These social benefits can be large even when the potential for free riding is small because they are unrelated to the size of the R&D investments that produced the secret information.

3. Switching Costs

The market significance of admissibility creates distinctive consumer switching costs in these markets that also can extend first mover exclusivity periods. Switching to a competing tool is likely to require costly re-training of laboratory personnel and internal validation of a particular lab's implementation. Switching costs will be amplified by the need to establish the admissibility of the new tool in the local courts. The importance of admissibility to purchasers is captured clearly in the Cuyahoga County Board of Control's agenda for a meeting that resulted in a decision to purchase the forensic tool TrueAllele:

> "The True Allele Casework system has been extensively validated and used by forensic laboratories in the United States. In addition, the True Allele casework system has been through admissibility hearings in 6 US states including Ohio which is a great advantage for the laboratory. This means that the laboratory will not have to go through the admissibility hearing to get the True Allele results accepted in court."[138]

These sorts of cost considerations will surely affect any agency's willingness to switch to a competitor's tool once a first mover's tool has been adopted. The pattern of PG tool adoptions supports this observation. States that initially adopted TrueAllele have stayed with it, despite the current popularity of STRmix among new adopters of PG technology.

The switching costs associated with establishing admissibility in one jurisdiction will diminish once the admissibility of a new entrant's tool becomes established elsewhere. But adoption is apt to be delayed nonetheless while laboratories sort out the collective action problem associated with shouldering the costs of the first few admissibility hearings.

---

137. *See supra* Part I.B.

138. *Meeting Agenda of the Board of Control of Cuyahoga County*, CUYAHOGA CTY. BDS & COMM'N (June 15, 2015), http://bc.cuyahogacounty.us/en-US/Board-of-Control.aspx?Year=2015.

4. Judicial Precedent and the Network Effects of Admissibility Decisions

In this Subpart, we use an in-depth study of judicial opinions regarding trade secrecy and admissibility for probabilistic genotyping software to illustrate how judicial practices regarding admissibility produce a rich-get-richer network effect that is likely to substantially enhance first mover market exclusivity, even if disclosure is required. These network effects arise from the way in which courts rely on adoption by forensic labs as evidence of a tool's scientific acceptability and then piggyback on previous admissibility determinations, even from outside of their own jurisdictions.

One of the first U.S. courts to directly address the admissibility of PG in a written opinion was the Pennsylvania Superior Court in 2012. During the murder trial of former Pennsylvania state trooper Kevin James Foley, prosecutors sought to admit a likelihood ratio generated by the TrueAllele casework system as evidence that Foley's DNA was present at the crime scene.[139] The defense objected under Pennsylvania's particular variant of the *Frye* test on grounds that the tool was both "novel" and not "generally accepted" by the relevant community of scientists.[140] The trial court rejected the defense challenge at the first stage of analysis, finding that probabilistic genotyping was not "novel" because, or so the judge believed, it was little more than a "refined application of the 'product rule,' a method for calculating probabilities that is used in forensic DNA analysis."[141] Foley appealed.

While not directly endorsing the largely erroneous[142] reasoning that PG is merely a refinement of the product rule, the Superior Court affirmed the finding that PG was not novel on the grounds that there was "no legitimate dispute regarding the reliability of the expert's conclusions."[143] The court based this conclusion in part on then-current uses "by New York State for all of their data banking and bringing their casework system on board" and by the UK's Forensic Science Service and on the fact that "Allegheny County Crime Lab has been using our system as a service and recently purchased the system for looking at mixtures in complex cases and DNA evidence" and that the World Trade Center had engaged the company to conduct analysis relating to the identification of victim remains.[144] The court also relied heavily on the existence of two peer-reviewed validation studies, both authored by TrueAllele's creator, Mark Perlin.[145]

For many reasons, *Pennsylvania v. Foley* is an important early case in this area. The Superior Court's conclusions regarding the admissibility of TrueAllele evidence in spite of the tool's secret source code have been repeatedly echoed,

139. *See* Commonwealth v. Foley, 38 A.3d 882, 888 (Pa. Super. Ct. 2012).
140. *Id.*
141. *Id.*
142. *See* Appendix.
143. *Foley*, 38 A.3d at 888.
144. *Id.* at 889.
145. *Id.*

or directly cited, by trial courts all across the country.[146] In this way, the acceptance by the Pennsylvania Superior Court of TrueAllele evidence in this case establishes the first node in a network of favorable admissibility decisions, which are referenced again and again by courts, creating a rich get richer effect regarding the admissibility of these tools.

The effect of judicial reliance on earlier admissibility decisions is exacerbated by a distinct tendency to cite judicial admissibility decisions as evidence of "general acceptance." In doing so, judges appear to conflate the familiar concept of persuasive legal precedent with the more relevant question of acceptance by the scientific community. While a legal interpretation may become more persuasive if more judges have adopted it, a forensic tool does not become more generally accepted *by the scientific community* simply because more courts have agreed that it is generally accepted by that community.

For example, in January 2019, a trial judge in Georgia issued a ruling which included, under the heading "TrueAllele's Widespread Acceptance," the statement: "Courts accepting TrueAllele evidence include California, Florida, Indiana, Louisiana, Maryland, Massachusetts, Michigan, Nebraska, New Hampshire, New York, Ohio, Pennsylvania, South Carolina, Tennessee, Texas, Virginia, Washington, the United States Federal Courts (Eastern District of Virginia), United States Marine Corps, Northern Ireland, and Australia."[147] The same order included a list of "[e]ighteen admissibility decisions in the United States"[148] under the heading "TrueAllele is Reliable."

The compounding of judicial admissibility rulings is evident when one looks further into this list. The earliest of the cited cases was the Pennsylvania Superior Court's opinion in *Foley*. Notably, a striking number of the other cited opinions themselves cite to *Foley*, including orders from *Virginia v. Brady* (2013),[149] *Ohio v. Shaw* (2014),[150] *New York v. Wakefield* (2015),[151] *Washington v. Fair* (2017),[152] and *Florida v. Lajayvian Daniels* (2018).[153] Many of those opinions are themselves cited in other decisions on the list. For example, the 2015 decision in *New York v. Wakefield* was cited by an Ohio appellate court in 2021 in a discussion of the tool's general acceptance, and is referenced for its "compilation of cases accepting True Allele."[154] Together, all of the TrueAllele

146. *See, e.g.*, State v. Simmer, 935 N.W.2d 167, 181 (Neb. 2019); People v. Lang, No. F075921, 2019 WL 5205997, at *6 (Cal. Ct. App. Oct. 16, 2019); People v. Belle, 16 N.Y.S.3d 793 at *4 (N.Y. Sup. Ct. 2015).

147. Order, State v. Nundra, No. 18-CR-134, at *2–3 (Ga. Super. Ct. Jan. 21, 2019).

148. *Id.* at *5–6.

149. Order, Commonwealth v. Brady, Nos. CR11-465-01,-02,-03, & -04 and CR11-494-01, -02, -03, & -04, (Va. Cir. Ct. Dec. 17, 2013).

150. Order, State v. Shaw, No. CR-13-575691 (Ohio Ct. Com. Oct. 10, 2014).

151. Order, People v. Wakefield, 9 N.Y.S.3d 540 (N.Y. Sup. Ct. 2015).

152. Order, State v. Fair, No. 10-1-09274-5 SEA (Wash. Super. Ct. Apr. 4, 2016).

153. Order, State v. Daniels, No. 2015CF009320AMB, at 3 (Fla. Cir. Ct. Oct. 31, 2018) (Under the heading "TrueAllele's Widespread Acceptance" we find: "TrueAllele's reliability has been confirmed in appellate precedent in Pennsylvania. *See* Commonwealth v. Foley, 38 A.3d 882 (Pa. Super. 2012).").

154. State v. Preston, No. CR-18-634913-A (Ohio Ct. App. July 1, 2021).

admissibility decisions we were able to gather create the striking cross-jurisdictional network of citation and reliance illustrated in Figures 2 and 3.

"Network effects" traditionally arise when a product's value to a potential consumer grows with the number of existing consumers.[155] Network effects make it more difficult for new entrants to compete or gain a foothold in the market, even if they introduce qualitatively better products. Social networking services provide a classic example of products that benefit from a network effect: the more people there are who use the service, the more appealing the network becomes to potential future users. Here, our research suggests an analogous effect. A forensic evidence tool's attractiveness to potential purchasers depends on the likelihood that its results will be deemed admissible by local courts. Because of the way that courts rely on previous admissibility decisions, the marketability of a forensic evidence tool grows as it accumulates favorable admissibility decisions, regardless of jurisdiction. A favorable admissibility decision for a PG company, therefore, confers a market advantage upon a product that extends beyond the product's acceptance in the jurisdiction holding the hearing. This cumulative effect makes it more difficult for a newcomer to enter the market and thus extends first mover advantages.

Prosecutorial submissions of court orders from other jurisdictions as exhibits to admissibility hearings suggest that they are well aware of these effects. For example, at a 2018 *Daubert* hearing in Davidson County, Tennessee, prosecutors submitted a "binder containing 13 decisions from other jurisdictions,"[156] while a court in Chemung County, New York listed among the People's evidence "People's Exhibit #12 ('Admissibility Rulings'), comprising 20 court decisions ruling on TrueAllele's admissibility"[157]

Judicial orders suggest that this prosecutorial tactic is persuasive. For example, a number of recent orders admitting TrueAllele results incorporate nearly identical paragraphs including the sentence: "Courts accepting TrueAllele evidence include California, Florida, Indiana, Louisiana, Maryland, Massachusetts, Michigan, Nebraska, New Hampshire, New York, Ohio, Pennsylvania, South Carolina, Tennessee, Texas, Virginia, Washington, the United States Federal Courts (Eastern District of Virginia), United States Marine Corps, Northern Ireland, and Australia."[158] In effect, admission by other courts seems to be playing the role of the *Frye* standard's "general acceptance in the relevant scientific community."[159]

155. Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CALIF. L. REV. 479, 483 (1998).

156. State v. Watkins, No. 2017-C-1811, at 12–14 (Tenn. Crim. Ct. Dec. 17, 2018).

157. State v. Wilson, No. 2013-331, at 3 (Chemung, N.Y. Cnty. Ct. May 1, 2019).

158. State v. Daniels, No. 2015CF009320AMB, at 3 (Fla. Cir. Ct. Oct. 31, 2018); State v. Nundra, No. 18-CR-134, at 2–3 (Ga. Super. Ct. Jan. 21, 2019); State v. Baugh, No. 2017-CR-618, at 8 (Ga. Super. Ct. Mar. 22, 2019).

159. Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

Cases dealing with TrueAllele's key competitor, STRmix, exhibit the same sort of network pattern. Like the TrueAllele cases, the STRmix cases begin with a cornerstone written admissibility decision. In a 2015 order in the case of *Michigan v. Elamin Muhammad*,[160] a Muskegon County trial court admitted evidence generated by the STRmix tool in what appears to be the first recorded decision on the tool. In support of its decision to admit STRmix evidence, the court emphasized that "at least two cases in New York utilized opinions based on STRmix evaluations."[161] This first STRmix decision also illustrates the limitations on trade secrecy's role in this market, given the doctrinal requirement that basic principles must be disclosed. Its decision not only relied on previous STRmix cases, but also noted that "courts in Pennsylvania, Virginia, New York and Ohio have admitted results from a program based upon similar principles [TrueAllele]."[162]

The next court to take up the issue of STRmix admissibility in a written order was back in New York.[163] Interestingly, that case makes no mention of the two New York opinions cited by the Michigan court in *Muhammad*, but instead cites heavily to the Michigan court.[164] Thus, the New York court in *Bullard-Daniel* states that "this case concerns the first judicial review, as far as this court is aware, of STRmix in New York," and that "[t]here is only one reported decision involving STRmix, from Michigan, where the court applied *Daubert* and upheld the admissibility of DNA test results."[165] The *Bullard-Daniel* court devotes substantial attention to the Michigan decision, stating the name of the case eight times and summarizing that court's reasoning in substantial detail.[166]

---

160. State v. Muhammad, No. 14-65263-FC (Muskegon Co. Dec. 15, 2015) (Hon. William C. Marietti) (on file with authors).

161. *Id.* The two New York opinions noted by the Muskegon County court appear to be unreported, and the authors have been unable to locate them.

162. *Id.*

163. *See* People v. Bullard-Daniel, 42 N.Y.S.3d 714, 715 (Niagara Cnty. Ct. 2016).

164. *Id.* at 721.

165. *Id.* at 720.

166. The *Bullard-Daniel* court explains:

> Of course, this Court would expect that the statement referred to above would add another arrow in the quiver of defense counsel that would be used to undermine the STRmix results when the issue is presented to the trial jury, but it does not affect the issue of the general acceptance of STRmix within the relevant scientific community. *State v. Muhammad*, No. 14-65263-FC (Muskegon Co. Dec. 15, 2015) (Hon. William C. Marietti) (on file with authors), is the only other reported case in the country regarding the admissibility of STRmix. There, the court concluded as a preliminary matter that statistical evaluation of the DNA analysis's results is a matter of evidentiary weight, not admissibility.' Thus, the court's determination of admissibility falls into the category of dicta. Nonetheless, the court reached several conclusions, which are persuasive insofar as this Court is faced with identical issues. First, the *Muhammad* court found that STRmix 'received adequate validity testing.' Indeed, Dr. Buckleton testified in *Muhammad,* and it was anticipated, based on preliminary representations made to this Court by the People, that he would testify here. His testimony could have resolved several questions raised by the cross-examination testimony of Dr. Simich and the direct testimony of Dr. Skuse. Notwithstanding Dr. Buckleton's failure to testify here, Dr. Simich's testimony was sufficient to meet the People's burden of establishing, by a preponderance of the evidence, that STRmix was generally accepted in the relevant scientific

Despite noting that the out-of-state precedent was only "dicta," the New York court found the Michigan decision "persuasive insofar as this Court is faced with identical issues."[167]

From there, the STRmix cases begin to form a citation network that resembles the one we see in the TrueAllele cases, as shown in Figures 2 and 3. For instance, in 2018, the *Bullard-Daniel* decision was cited in published opinions from California, New Mexico, and the Virgin Islands[168] in support of the acceptance of TrueAllele. Before long, trial courts began issuing decisions expressly relying on a network of inter-jurisdiction admissibility precedent. For example, in a 2020 decision addressing the admissibility of STRmix evidence under Colorado's expert evidence admissibility standard (called *Shreck*), Judge Marcelo Kopcow of the Weld County Court wrote that "[c]onsidering factors similar to those outlined in Shreck, courts in at least Colorado, Illinois, Wyoming, New York, New Mexico, Minnesota, Michigan, Connecticut, Florida, California, and the Virgin Islands have found probabilistic genotyping and STRmix sufficiently reliable to be admitted and submitted to the Jury."[169]

These citation patterns show that courts confronted with challenges to the admissibility of a particular PG tool have relied heavily on previous admissibility opinions concerning that tool, regardless of jurisdiction. This network of admissibility strengthens first mover advantages, making it more difficult for competitors to enter the market, while also sometimes seeming to conflate judicial agreement about admissibility with general acceptance in the scientific community. STRmix entered the market at a time when PG technology was relatively new and many states had yet to adopt it. At that early stage, when the admissibility of the basic method was being established, STRmix was able to benefit to some extent from decisions admitting TrueAllele in other jurisdictions. Once a jurisdiction has adopted a particular tool, however, courts' reliance on precedent in making admissibility decisions compounds the barriers to entry that a new entrant would face.

Recently, a state appellate court in New Jersey bucked this trend, and in doing so, became the first (to our knowledge) to look critically at the network of precedent laid before it by state prosecutors and TrueAllele's Mark Perlin. *New Jersey v. Pickett* became "the first appeal in New Jersey addressing the science underlying the proffered testimony by the State's expert, who designed, utilized,

---

community. Significantly, Dr. Simich testified in *Muhammad* and that court found his testimony relevant and significant. Dr. Simich reported the results of the Commission and the DNA subcommittee and the *Muhammad* court discussed those results in a positive light.

Id. at 725–26.

167. *Id.* at 725.

168. United States v. Russell, No. CR-14-2563 MCA, 2018 WL 7286831, at *8 (D.N.M. Jan. 10, 2018); People v. Dominguez, 239 Cal. Rptr. 3d 71, 71 (2018); People v. Blash, No. ST-2015-CR-0000156, 2018 WL 4062322, at *8 (V.I. Super. Ct. Aug. 24, 2018).

169. Order, People v. Hendrix, No. 2018CR1767 (Weld, Colo. Cnty. Ct. May 4, 2020).

and relied upon TrueAllele."[170] The Pickett court noted that Perlin "[s]ubmitted a seventy-eight paragraph declaration documenting," among other things, "TrueAllele's purported widespread acceptance."[171]

The *Pickett* court proceeded to both recognize and reject the strange network effect of judicial precedent placed before it. This portion of the court's decision, which includes a list of several of the major cases on our network map, tracks so closely with our own observations that it bears quoting at length:

> The first court to address the question of admissibility was *Commonwealth v. Foley*, 38 A.3d 882, 889-90 (Pa. Super. Ct. 2012), where the court accepted Dr. Perlin's assertion that validation studies are the best tests of the reliability of source codes. The court reasoned that "scientists can validate the reliability of a computerized process even if the 'source code' underlying that process is not available to the public," emphasizing that making the source code available would have market consequences. [. . .] Subsequent courts have placed great emphasis on the observation made in *Foley*, without further scrutiny, creating an authority "house of cards." *See, e.g.*, *People v. Superior Court (Chubbs)*, No. B258569, 2015 WL 139069, at *8 (Cal. App. Ct. Jan. 9, 2015); *State v. Daniels*, No. 2015CF009320AMB (Fla. Cir. Ct. Oct. 31, 2018) (slip op. at 3); **307 *State v. Wakefield*, 47 Misc.3d 850, 9 N.Y.S.3d 540, 541 (Sup. Ct. 2015); *State v. Shaw*, No. CR-13-575691 (Ohio C.P. Ct. Cuyahoga Cnty. Oct. 10, 2014) (slip op. at 23); *Commonwealth v. Knight*, No. 379 WDA 2017, 2017 WL 5951725, at *6 (Pa. Super. Ct. Nov. 29, 2017); *State v. Watkins*, No. 2017-C-1811 (Tenn. Crim. Ct. Davidson Cnty. Dec. 17, 2018) (slip op. at 13-14). Published out-of-state judicial decisions, although persuasive rather than binding, carry great weight, especially after they are cited by other courts. A long line of decisions uniformly in favor of a legal proposition suggests that a legal proposition is generally accepted. We are mindful, however, that in science, the repetition of authority does not automatically establish reliability for purposes of a Frye hearing.[172]

The *Pickett* court ultimately ordered the TrueAllele source code released to the defense pursuant to a protective order, holding that "[t]he cases identified by the State include a laundry list of admissibility rulings, but to reiterate, none consider whether the TrueAllele source code itself correctly implements its methods, which can only be tested in the manner defendant and amici advocate for here."[173]

At least one other court has already partially adopted the laudable approach expressed in the *Pickett* decision. The United States District Court for the Western District of Pennsylvania recently cited *Pickett* in declining to quash a subpoena for TrueAllele's source code, holding that "some level of access to the source code, with proper protections, represents a reasonable outcome."[174]

---

170. State v. Pickett, 246 A.3d 279, 283 (N.J. App. Div. 2021).
171. *Id.* at 286.
172. *Id.* at 306–07 (footnote omitted).
173. *Id.*
174. United States v. Ellis, No. 19-369, 2021 WL 1600711, at *1 (W.D. Pa. Apr. 23, 2021).

It remains to be seen how many other courts will adopt the reasoning laid out in *Pickett*, particularly state and local courts in states where appellate authority already exists approving the admissibility of TrueAllele or STRmix without ordering any source code disclosure. Indeed, important state court orders and decisions issued more recently than the *Pickett* decision continue to find general acceptance of TrueAllele technology, without disclosure or review of the source code, based on the same arguments regarding prior inter-jurisdictional admissibility decisions. For example, the New York State Appellate Division, Third Department, deciding on appeal that TrueAllele evidence had been properly admitted, recently found persuasive the fact that "at the time in question, courts in at least three other states had found the TrueAllele Casework system to be reliable under the Frye standard."[175] Another example comes from an appellate decision in Florida challenging the admission of TrueAllele evidence, where, in response to the defendant-appellant's concern that no internal validation had been done on the lab in question, the District Court of Appeals for Florida's Fourth district found it notable that "the Cybergenetics DNA analyst testified that in eight of the admissibility challenges against TrueAllele in prior cases where the TrueAllele evidence was ruled admissible, there was never any internal validation done on the lab from which the data came nor was the lack of internal validation on a specific lab's data an issue for the reliability of the evidence."[176]

These recent decisions, complemented by numerous recent court orders listed on TrueAllele's website,[177] for example, demonstrate the continuing influence that the network effect we have observed exerts on the universe of PG admissibility decisions and on the PG marketplace as a whole. At the same time, the novel analysis of the *Pickett* decision offers a glimpse into what it might look like if, to quote that court, this "house of cards" of judicial authority were to fall. If, let's imagine, under the *Pickett* ruling, state and local courts were to begin ordering source code disclosure under protective order, and if independent review by defense experts were to become more common as a result, perhaps the value of admissibility precedent to these market players would be comparably diminished. For now, though, and for the nearly ten years that have elapsed since the *Foley* decision, the persuasiveness, impact and value of this network effect has been, and remains, measurable and significant.

E. THE DUBIOUS VALUE OF TRADE SECRET PRIVILEGES FOR PROMOTING INNOVATIVE FORENSIC EVIDENCE TECHNOLOGY

Pulling the above observations together, we conclude that it is unreasonable to assume that trade secret privileges are important for preserving incentives for forensic evidence technology innovation. The trade secret

175. People v. Wilson, 143 N.Y.S.3d 466, 468 (2021).
176. Daniels v. State, 312 So. 3d 926, 929 (Fla. Dist. Ct. App. 2021).
177. *TrueAllele Admissibility*, *supra* note 121.

information at issue, such as source code and implementation parameters, is of narrow scope and may not encompass the most innovative aspects of the technology. Even if the trade secret information at issue reflects substantial R&D investment, competitors ordinarily cannot simply copy the court-ordered disclosure and grab a share of the market without falling afoul of copyright or patent protections that survive disclosure. Moreover, potential market entrants must conduct costly and time-consuming validation studies and will confront significant switching costs and network effects created by admissibility doctrine and judicial practice. It thus seems quite likely that, even with court-mandated disclosure of source code, parameters and so forth, first mover exclusivity will be more than adequate to recoup R&D investments in free-rideable trade secret information. The questionable social benefits of trade secrecy are highly unlikely to outweigh the significant social benefits of public disclosure. For this reason, we do not think protective orders covering disclosures about forensic evidence technology should be issued in most cases. If, however, it can be demonstrated that disclosures about a particular tool are especially likely to facilitate problematic free riding, courts are free to bestow further exclusivity by covering the mandated disclosure with a protective order.

F. Secrecy and the Distortion of Demand for Forensic Evidence Technology

Though the innovation benefits of trade secret privileges for forensic evidence technology are likely to be minimal at best, secrecy has the potential to skew market demand for such innovation in socially undesirable directions. Forensic technology is not a private good. It should be designed to serve public purposes. Society's goals and values, as enshrined in the Constitution and the traditions of the criminal justice system, include preferences for more accurate law enforcement, for avoiding false convictions, as well as for practicalities such as lower cost. As already discussed, the "customers" for forensic evidence technology are forensic laboratories and, ultimately, prosecutors and law enforcement agencies. Market demand for innovation in this market, like others, is driven by customer demand. The customers in this market are agents for the public, but imperfect agents, who have various personal and professional motivations, including a desire for "success" in their cases, a preference for lower costs, a concern with accuracy and, probably most immediately, a desire for tools that produce persuasive, admissible evidence.

Because "customer" preferences in this market are only partially aligned with society's goals and values, there is likely to be a mismatch between the technology that would best serve society and the technology that these customers demand. On top of these principal-agent issues, forensic evidence tools purchased from private companies are likely to be "credence goods," meaning that it will be difficult for purchasers to assess their quality through use. The

failures of market demand associated with credence goods are commonly addressed by regulation, often involving mandated explanation or disclosure.[178]

In the criminal justice system, admissibility standards, judicial gatekeeping and the adversarial process are designed to address essentially these problems, though they are not usually described in market failure terms. Judicial gatekeeping and adversarial testing of evidence are foundations of the U.S. criminal justice system and a primary means for closing the gap between social values and prosecutor preferences. If admissibility doctrine, judicial practice and trade secrecy privilege combine to undermine the efficacy of these mechanisms, demand in the market for forensic evidence tool innovation will be misaligned with public values. This sort of market failure cannot be remedied by competition because competitors all respond to the same, misaligned demand signals. Rather than merely slowing the pace of innovation in forensic evidence technology, demand misalignment produces a portfolio of innovative activity that is mis-directed and fails to serve society's goals and values.[179] Admissibility and trade secret privilege doctrines thus play crucial roles in regulating the market for forensic evidence tools.

In a follow-on article, we will argue, these demand-side problems are exacerbated for software-based technologies because current approaches to admissibility and validation fail to account adequately for software's distinctive nature. In light of the growing importance of software-based forensic tools, the inadequacy of current approaches is a matter of major concern. The FST debacle illustrates the way that secret source code can hide post-validation modifications and questionable "fixes." Aside from failing to uncover this sort of misconduct, judges have been willing to allow developers of proprietary code to rely solely on lab-based input-output testing that is not properly designed to uncover coding errors. These inadequate doctrinal and judicial standards, combined with the conflation of precedent with scientific acceptance, strip these markets of incentives for the sorts of innovations that would improve code quality, generalizability and dependability.

III. A FEW WORDS ABOUT INCIDENTAL AND DUAL-PURPOSE FORENSIC EVIDENCE TOOLS

So far, we have implicitly assumed that forensic evidence technology is developed and marketed solely for use by crime labs for the purpose of analyzing forensic evidence, such that innovation is driven mostly by the preferences of law enforcement agencies and prosecutors. While many, if not most, forensic evidence tools fit this pattern, undoubtedly some are also used or marketed for

178. For discussions of the credence goods problem in other markets, see Ariel Katz, *Pharmaceutical Lemons: Innovation and Regulation in the Drug Industry*, 14 MICH. TELECOMM. & TECH. L. REV. 1 (2017); Gillian K. Hadfield, *The Price of Law: How the Market for Lawyers Distorts the Justice System*, 98 MICH. L. REV. 953 (2000).

179. *See* Lev-Aretz & Strandburg, *supra* note 107, at 4.

other commercial purposes. Some technologies may be developed primarily for standard commercial markets and then used as forensic evidence tools at a later time. From an incentive perspective, such incidental forensic uses are unimportant, however, because unanticipated disclosure cannot depress a priori incentives for innovation. Some technologies, however, presumably are developed with both forensic and other applications in mind. It is possible that the expectation of court-ordered disclosure could affect incentives for innovation of such dual-purpose technologies. To get a basic handle on whether and how our analysis might differ for such dual-purpose technologies, it is helpful to distinguish two possibilities.

One possibility is that first mover advantages, along with trade secrecy and other intellectual property protections might be sufficiently robust in ordinary commercial markets to incentivize the development of a dual-purpose technology, but the potential for court-mandated disclosure might deter innovators from marketing the technology for forensic evidence applications. While this might be a perfectly sensible business strategy, it could be unfortunate from a public perspective to deprive courts of the evidence that could be produced by such tools. Another possibility is that some dual-purpose technologies require such large investments that they can only be recouped by marketing to both conventional and forensic evidence markets. In such situations, the fear that court-mandated disclosure could be used to free ride in the conventional market might be enough to deter innovation completely.

The trade secrecy privilege debate may or may not have much bearing on innovators' business decisions for either sort of dual-purpose technology. Recall that we argued earlier that only a limited amount of free-rideable information is truly at stake in the trade secret privilege decision, because so much about principles, methods and validation must be disclosed even under current admissibility doctrine. That argument carries over to dual-purpose technologies. Copyright protection is also still available for dual-purpose software, though it is admittedly easier for free riders to hide infringement in commercial markets, where they can keep their code secret. It is of course possible that, for some dual-purpose technologies, the marginal free riding in conventional markets facilitated by court-mandated disclosure of source code and implementation details could tip the balance. Even for those technologies, a trade secret privilege in criminal cases is not likely to be justified. Instead, courts could simply employ protective orders. That is, after all, the approach used in high-stakes commercial trade secret litigation between competitors.[180]

In sum, even for dual-purpose technologies, public disclosure may not have much impact on incentives for innovation. Where significant impact is likely, disclosure under a protective order might be appropriate. Of course, disclosure

180. Rebecca Wexler, *It's Time to End the Trade Secret Evidentiary Privilege Among Forensic Algorithm Vendors*, BROOKINGS (July 13, 2021), https://www.brookings.edu/blog/techtank/2021/07/13/its-time-to-end-the-trade-secret-evidentiary-privilege-among-forensic-algorithm-vendors.

under a protective order would also be one way to proceed for technologies employed primarily for forensic purposes, such as probabilistic genotyping. As mentioned earlier, while this approach would be an improvement over current practice, we do not endorse it because the potential benefits of trade secrecy are very unlikely to outweigh the social benefits of public disclosure.

CONCLUSION

The evidentiary privilege for trade secrets is premised on a policy of incentivizing innovations by ensuring that advancements are not immediately replicable by "free riding" competitors. This Article analyzes the flaws in this premise in some detail. Rather than reiterate the analysis here, we close with a hypothetical narrative, based on our probabilistic genotyping case study, that encapsulates our argument.

Imagine that a court has ordered the disclosure of TrueAllele's source code and relevant input parameters and that you have obtained a copy. You decide to free ride on this disclosure to start a company to market a competing probabilistic genotyping tool at a discounted rate. But what does this marketplace look like? As it turns out, even armed with TrueAllele's source code, the landscape that greets the free rider is bleak. Two companies have already secured major contracts with crime labs all across the country. These companies' PG tools have been battle-tested in successful admissibility litigation in dozens of states. They carry binders full of validation studies: implementations by state crime labs, even the FBI. When you approach your local crime lab with your new tool, they want to see *your* validation studies. If you decide to invest in such studies, it will take you some time to access the necessary laboratory equipment and samples, conduct the studies and shepherd them through the process of peer review and publication. When you return, considerably poorer, to your prospective customer, you find that the tool the lab currently uses has continued to rack up positive admissibility decisions. Even when you brandish your published validation studies, the laboratory is reluctant to take the risk of relying on your untested tool. To make the sale, you are forced to offer a much deeper discount than you had originally envisioned, leaving you further in the hole. But you remain optimistic that you will eventually be able to make inroads on the market. Now it is time for your tool's first admissibility hearing. Of course, your tool is also subject to court-mandated disclosure. Mark Perlin now has access to your source code. Noticing that it looks suspiciously like his original code, he sues you for copyright infringement. Having no defense, you are compelled to pay damages and are enjoined from making further use of the code. Thus ends your foray into the probabilistic genotyping market. Of course, maybe you were smart enough to consult a copyright attorney and attempt to modify the code to avoid using any of Perlin's protected expression. Now you are even further in the hole (and still not completely certain that you have avoided copyright infringement). And that is not even to mention

Perlin's patents on a "Method and System for DNA Mixture Analysis." Apparently free riding is not all it's cracked up to be.

Now imagine that you are an expert in receipt of the court-mandated TrueAllele disclosure. You may find the sort of misfeasance illustrated by the FST debacle. But even if you do not, you may now be able to probe the limits of validity of TrueAllele's implementation of probabilistic genotyping. Using the information you uncover, you might serve as a defense expert in a case that pushes those limits, avoiding an unjustified guilty verdict. Or, you might see how to devise a better, more accurate PG tool and decide to try to enter the market. Even with your improved technology, entering the market will be a challenge. You will, of course, consult an IP attorney and avoid copying Perlin's code. Still, depending on how your technology builds on Perlin's, you might have to pay patent royalties or even sell him your improvement. You will still have to validate your tool and overcome potential customers' qualms about admissibility. And you will still have to satisfy market demand that is skewed toward prosecutorial preferences. As these two hypotheticals illustrate, markets for forensic evidence technology are far from easy playing fields for free riders. Even a follow-on innovator faces an uphill market entry battle.

Against the arguments that a trade secrecy privilege is needed to promote innovation in forensic evidence technology, stand countervailing concerns about the Constitutional rights and fair treatment of people accused of crimes. These fundamental rights have been addressed by others and thus are not our focus here, but they provide an additional yardstick against which the anemic free rider arguments for trade secrecy must be measured. Our analysis strongly suggests that the economic arguments for a trade secret privilege for forensic evidence technology come up short by any measure.

APPENDIX

A. PROBABILISTIC GENOTYPING TECHNOLOGY

1. Terms

Alleles: Gene variants that are present at predetermined *loci* in the genome. In general, each person has two alleles of each gene. If the person's two alleles are the same, they are said to be *homozygous* for that trait. If the person's two alleles are different, they are said to be *heterozygous* for that trait.[181]

Genotype: The DNA profile (i.e., the composition of alleles) of an organism.[182]

Genome: A complete set of genetic information in an organism.[183]

Heterozygous: When a person's two *alleles* at a particular *locus* are different.[184]

Homozygous: When a person's two *alleles* at a particular *locus* are the same.[185]

Locus (pl. loci): Fixed positions on a chromosome that contains genetic information encoding a particular gene or genetic marker.[186]

Short tandem repeats (STRs): Consecutively repeated units of DNA, typically in noncoding regions of the *genome*.[187]

B. A BRIEF PRIMER ON LIKELIHOOD RATIOS

1. Likelihood Ratios

The LR can be expressed as follows, in the form of a Bayesian conditional probability:

$$LR = \frac{Pr(E|S)}{Pr(E|U)}$$

where Pr(E|S) is the probability the evidence in the DNA mixtures comes from the suspect, and where Pr(E|U) is the probability the evidence in the DNA

181. *Allele*, *supra* note 38.

182. *Genotype*, SCITABLE BY NATURE EDUC., https://www.nature.com/scitable/definition/genotype-234 (last visited Mar. 21, 2022).

183. *Genome*, SCITABLE BY NATURE EDUC., https://www.nature.com/scitable/definition/genome-43 (last visited Mar. 21, 2022).

184. *See Allele*, *supra* note 38.

185. *Id.*

186. BUTLER ET AL., *supra* note 34, at x.

187. Stephanie Feupe Fotsing, Jonathan Margoliash, Catherine Wang, Shubham Saini, Richard Yanicky, Sharona Shleizer-Burko, Alon Goren & Melissa Gymrek, *The Impact of Short Tandem Repeat Variation on Gene Expression*, 51 NATURE GENETICS 1652, 1652 (2019), https://www.nature.com/articles/s41588-019-0521-9.

mixture comes from an unrelated individual.[188] The mathematics underlying PG software and LRs are somewhat simplified when one considers that the term PR(E|S) is often expressed as H_p, or the prosecution hypothesis. This number is often set to 1 since the prosecutor(s) involved in the case would not bring charges against the POI if they did not fully suspect s/he was the perpetrator.[189] PR(E|U) is often expressed as H_d, or the defense hypothesis.[190] This number is often the random match probability (RMP), a mathematical term that represents the likelihood of finding the DNA variant from the sample in the random population.[191] In essence, this represents the defense alleging that the particular DNA in this sample could have been contributed by a random individual, with a probability equal to the H_d.

2. Quantifying Likelihood Ratios

Likelihood ratios can be astronomically high, with some indicating that it is tens of millions of times more likely that the suspected individual is a contributor than an unknown person. This determination would result in cases where H_d, i.e., the probability of finding that particular allele in the general population (i.e., $2pq$), is extremely small. A simple mathematical analysis illustrates this concept: if the LR = H_p/H_d, and H_p=1 and H_d (i.e., the RMP, or $2pq$) is relatively extremely small, the LR will be a relatively large number. Even in the simplest genetic situation, where the only possible alleles of the gene of interest are p and q, $p + q = 1$ (but in nearly all real-world situations pertinent to PG technologies, neither $p = 1$ nor $q = 1$, and with regards to STR repeats used in PG, it is possible that both $p<<1$ and/or $q<<1$). Even in this simplified example, the product $pq < 1$ for all values of p and q, and the product $2pq<1$. This results in LR = $1/n$, where $n<1$. Thus, LR>1.

188. Coble & Bright, *supra* note 29, at 219.
189. *Id.*
190. *Id.*
191. *Id.*

***