VERFICATION DILEMMAS IN LAW AND THE PROMISE OF ZERO-KNOWLEDGE PROOFS

Kenneth A. Bamberger,[†] *Ran Canetti,*[††] *Shafi Goldwasser,*[†††] *Rebecca Wexler*[‡] *& Evan J. Zimmerman*[‡‡]

ABSTRACT

Individuals who wish to access a website or qualify for a loan are expected to expose personally identifying information, undermining their privacy and security. Firms share proprietary information in dealmaking negotiations which, if the deal fails, may be used by the negotiating partner for a competitive advantage. Regulators are expected to disclose their algorithmic tools to comply with public transparency and oversight requirements, a practice that risks rendering these tools circumventable and ineffective. Litigants might have to reveal trade secrets in court proceedings to prove a claim or defense. Such "verification dilemmas"—costly choices between opportunities that require the verification of some fact and risks of exposing sensitive information in order to perform that verification—appear across the legal landscape. Yet existing legal responses to them are imperfect. Legal responses often depend

---

DOI: https://doi.org/10.15779/Z38P55DH8N

 This Article received an Honorable Mention from the Privacy Papers for Policymakers Award as one of the top eight privacy papers in 2022. The award is for "leading privacy scholarship that is relevant to policymakers in the U.S. Congress, at U.S. federal agencies, and among international data protection authorities." *12th Annual Privacy Papers For Policymakers Awardees Explore The Nature Of Privacy Rights & Harms*, Future of Privacy F. (Jan. 13, 2022), https://fpf.org/blog/12th-annual-privacy-papers-for-policymakers-awardees-explore-the-nature-of-privacy-rights-harms. The authors would like to thank: the editors of the *Berkeley Technology Law Journal* for their exceptional editorial assistance; David Ameling, Robert Bartlett, Abraham Cable, Victoria A. Cundiff, Jim Dempsey, Deven Desai, Arthur R. Miller, Jules Polonetsky, Yuval Shany, and participants in the Hebrew University Federmann Cyber Security Center Symposium, and the Berkeley Center for Law and Technology R2P Seminar series, for helpful input and comments;, and I-Wei Wang for reference assistance. Ran Canetti and Shafi Goldwasser were supported by the DARPA SIEVE project, contract no. HR00112020021.

† The Rosalinde and Arthur Gilbert Foundation Professor of Law, University of California, Berkeley; Faculty Director, Berkeley Center for Law and Technology.

†† Professor of Computer Science, Boston University; Director, Center for Reliable Information System and Cyber Security.

††† Director of the Simons Institute for the Theory of Computing, and C. Lester Hogan Professor in Electrical Engineering and Computer Sciences, University of California, Berkeley; RSA Professor of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

‡ Assistant Professor of Law, University of California, Berkeley; Faculty Director, Berkeley Center for Law and Technology.

‡‡ Founder, Jovono; J.D., University of California, Berkeley, School of Law.

on ex post litigation procedures that can be prohibitively expensive for those most in need or are otherwise ineffective.

Zero-knowledge proofs (ZKPs)—a class of cryptographic protocols that enables verification of a fact or characteristic of secret information without learning the actual secret—can help to avoid these verification dilemmas. ZKPs can provide a feasible means for a party who holds secret information to demonstrate desirable properties of this information while keeping the information otherwise hidden. Yet ZKPs have received scant notice in the legal literature. This Article fills that gap by providing the first deep dive into ZKPs' broad relevance for law. It explains ZKPs' conceptual power and technical operation to a legal audience. It then demonstrates how ZKPs can be applied as a governance tool to transform verification dilemmas in multiple legal contexts. Finally, the Article surfaces and provides a framework to address the policy issues implicated by introducing of ZKP governance tools into existing law and practice.

TABLE OF CONTENTS

I. INTRODUCTION

According to a recent legal filing by TargetSmart, a data-based Democratic campaign consulting firm, representatives of GHP, an investment firm, approached it to express an interest in funding potential partnerships or

mergers.[1] As part of the due diligence involved in the subsequent negotiations, the investment firm sought the disclosure of information that "TargetSmart believed . . . went beyond what was required to appraise TargetSmart's business . . . ."[2] After failing to identify other methods to resolve uncertainty about the TargetSmart firm's value—a common challenge for innovative firms needing to verify to potential acquirers, investors, or partners, the value or novelty of their source code, their margins, or their customer base—TargetSmart disclosed confidential business information. This included the proprietary digital code that powered their "VoterBase" and "VoterFile 2.0" products, client information, the terms of third-party relationships, and vendor agreements. Shortly thereafter, it was revealed in a complaint that GHP represented TargetSmart's chief competitor and allegedly had violated the terms of its contractual nondisclosure agreement (NDA) by leaking the confidential information to that rival.[3]

This episode illustrates a problem pervasive in the information age: proving specific facts, knowledge, or capacity to others frequently involves a costly choice. The first option is that parties can disclose information beyond the elements sought to be proven; yet considerations of privacy or security, circumvention or gaming concerns, and other confidentiality needs often make such overdisclosure undesirable. Alternatively, parties can choose not to disclose, which can exclude individuals from participatory opportunities, preclude firms from attracting capital or engaging in valuable partnerships, and thwart important public and private oversight. This Article calls these problems "verification dilemmas."

Verification dilemmas appear across the legal landscape. This Article examines four such contexts. Individuals seeking to prove specific attributes that are required to access certain opportunities (such as legal age, financial capacity, or other attributes required to access digital spaces or databases) must frequently disclose not merely the attribute but also their identity. This system, in turn, raises distinct privacy and security risks that the individuals must endure or otherwise forgo access to the opportunity. Similarly, potential deal partners, like TargetSmart, must reveal sensitive information about customers, models or data sets or otherwise relinquish valuable economic opportunities. Public policymakers face a choice between revealing the algorithms used in regulation, thereby potentially rendering those algorithms circumventable and ineffective or otherwise sacrificing opportunities for public transparency and

1. Complaint, TargetSmart Holdings, LLC v. GHP Advisors, LLC, at *1, No. 18-cv-11365 (D. Mass., Jun. 6, 2018).

2. *Id.* at *8.

3. *Id.* at *2.

oversight. Litigants in trade secret disputes can divulge their proprietary information to their adversaries, exposing the information to further leaks, misappropriation, or even the loss of legal protection or otherwise decide to keep it private and fail to offer evidence in support of their claims.

Legal solutions that seek to address verification dilemmas by limiting the costs of overdisclosure have proven only partially effective. Privacy law's restrictions on the use of personally-identifiable information are uneven at best and have failed to prevent the large-scale aggregation, unauthorized use, and recurring leaks of personal data. Trade secret law poses limited restraints on the use or disclosure of proprietary business information once it is shared with other parties. But trade secret misappropriation claims are not available for all confidential business information and are often challenging to track and enforce. Contractual NDAs can also be difficult to enforce. And even when they are enforced, NDAs merely provide an opportunity to receive (hard to prove) damages after a party violates a contract by disseminating confidential information without authorization. NDAs do not prevent the harm of that unauthorized use or dissemination from happening in the first place.

Ideally, then, there would be a way to limit disclosures to the particular fact, knowledge, or capacity being verified. Individuals could prove attributes about themselves without disclosing their identity. Public and private entities could verify compliance with certain requirements without revealing underlying details about data or algorithms that policy considerations deem sensitive. Firms could prove to potential acquirers, investors, or partners—including competitors—their financial margins, or that their source code operates as claimed, or that their innovative methods differ from those of their rivals, without disclosing the proprietary information itself. Litigants could establish similarities or dissimilarities between their alleged trade secrets without either party having to reveal their proprietary information to the other.

Zero-knowledge proofs (ZKPs) are mathematical objects designed to facilitate just that type of limited disclosure. First conceived by a group of researchers in 1985, including one of this Article's authors,[4] ZKPs provide a class of cryptographic protocols that take place between a "prover" and a "verifier." These protocols allow for the validation of a claim, fact, capacity, or identity, without requiring the "prover" to reveal to the "verifier" any underlying information beyond the validity of the assertion in itself. ZKPs thus allow for verification without overdisclosure: a characteristic of secret information can be verified as true, without revealing the actual secret. In a

4. *See generally* Shafi Goldwasser, Silvio Micali & Charles Rackoff, *The Knowledge Complexity of Interactive Proof-Systems*, 17 PROC. ANNUAL ACM SYMPOSIUM ON THEORY OF COMPUTING 291 (1985).

straightforward application, for example, an individual could use ZKP protocols to prove for legal purposes that they were of the age of majority without revealing their name or any other personal information linking that age to their identity, including even the birthday itself.

Until very recently, ZKPs have been considered as more a theoretical construct than a protocol efficient enough to use in complex applications.[5] But ZKPs' practical power has now been demonstrated by their application to the blockchain, where they have provided a privacy and anonymity backbone to the technology. ZKPs currently provide a means to keep specifics about private blockchain transactions from outside observers, even while the transactions occur on the public network.[6] More broadly, leading technology firms are increasingly using ZKPs to develop products that can protect individuals' detailed financial information when proving capacity for a loan,[7] promote compliance with privacy regulation by replacing data with proofs about that data,[8] or enable enterprises to use public blockchain technology to protect data confidentially in everything from banking to supply chain tracking.[9] Computer scientists have explored the capacity of ZKP protocols to

5. Mark C. Suchman, *Invention and Ritual: Notes on the Interrelation of Magic and Intellectual Property in Preliterate Societies*, 89 COLUM. L. REV. 1264, 1268 n.21 (1989) (opining that although contemporary work on "zero-knowledge proofs" suggests that the phenomenon of the disclosure paradox "may be theoretically surmountable," the "relevance of such abstract findings to real-world knowledge markets, however, is limited at best"); *see also* Mike Jenkins, *3 Real World Applications of Zero Knowledge Proofs*, COINBUREAU.COM (Oct. 25, 2018), https://www.coinbureau.com/adoption/applications-zero-knowledge-proofs/ ("Only now in 2018, more than 30 years after its inception, is [ZKP technology] being widely used in a practical way.").

6. The Zcash token model is the most well-known. *See* https://z.cash/the-basics (last visited Apr. 4, 2022) ("Like Bitcoin, Zcash transaction data is posted to a public blockchain; but unlike Bitcoin, Zcash gives you the option of confidential transactions and financial privacy through shielded addresses.").

7. *See ING launches Zero-Knowledge Range Proof solution, a major addition to blockchain technology*, ING (Nov. 16, 2017), https://www.ingwb.com/en/insights/distributed-ledger-technology/ing-launches-major-addition-to-blockchain-technology (promoting a product that allows verification of financial capacity within a required "range" without disclosing underlying data).

8. Ian Allison, *Deloitte Adds Privacy Tech to Its Education-Credentials Blockchain*, COINDESK (Oct. 29, 2019), https://www.coindesk.com/markets/2019/10/29/deloitte-adds-privacy-tech-to-its-education-credentials-blockchain/ (discussing Deloitte's use of technology allowing verification of educational qualifications without revealing personal details, aiming to comply with the requirements of the European Union's General Data Protection Regulation).

9. Jonathan Rouach, Data Privacy is Forever Changed. Zero-knowledge Proofs are Enterprise's Solution, FORKAST.NEWS (Jan. 22, 2020), https://forkast.news/data-privacy-is-forever-changed-zero-knowledge-proofs-are-enterprises-solution-opinion/ (quoting Jonathan Rouach, QEDIT's CEO) ("[A] supply chain consortium can deploy blockchain technology to track assets along a supply route, without displaying sensitive transactional

verify voting systems while preserving voter confidentiality[10] and to monitor compliance with nuclear agreements.[11] The Defense Department is developing ZKPs as an accountability tool for circumstances in which "the highest levels of privacy and security are required to protect a piece of information, but there is still a need to prove the information's existence and accuracy."[12]

Despite these developing applications and their implications for a swath of thorny legal issues involving the balance between confidentiality and disclosure, the legal literature has paid scant attention to ZKP cryptography. To date, the scholarship reveals only a few dozen mentions of the term[13] and only two examples of substantive engagement.[14] This Article seeks to remedy that gap in the literature in three ways. First, it provides an explanation, aimed at a legal audience, of ZKPs' conceptual power and operational characteristics. Second, it explores verification dilemmas across a range of concrete legal contexts and demonstrates how ZKPs can work in these contexts. Third, it surfaces and examines the policy issues raised by introducing ZKPs as a governance tool to augment (and potentially replace) existing law and practice.

Part II frames the idea of verification dilemmas that law and policy have sought to address. Such problems—in which a party is faced with an all-or-nothing choice between costly or undesirable overdisclosure of information, or no disclosure at all—pervade legal inquiry. This Part explores four disparate

---

details that reveal confidential information pertaining to trading partners, sales volume, or pricing.").

10. Somnath Panja & Bimal Kumar Roy, *A Secure End-to-End Verifiable E-Voting System Using Zero Knowledge Based Blockchain*, 2018 IACR CRYPTOLOGY (2018), https://eprint.iacr.org/2018/466.pdf.

11. *See* Alexander Glaser, Boaz Barak & Robert J. Goldston, *A Zero-Knowledge Protocol for Nuclear Warhead Verification*, 510 NATURE 497 (2014).

12. *Generating Zero-Knowledge Proofs for Defense Capabilities: Program Aims to Advance Method for Making Public Statements Without Compromising Sensitive Underlying Information*, DEFENSE ADVANCED RSCH. PROJECTS AGENCY (July 18, 2019), https://www.darpa.mil/news-events/2019-07-18.

13. WESTLAW, https://1.next.westlaw.com/Search/Home.html (last visited Oct. 1, 2020).

14. The first, Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson & Harlan Yu, *Accountable Algorithms*, 165 U. PA. L. REV. 633, 668–69 (2017), includes zero-knowledge proofs in a suite of computer-science tools useful for ex post algorithmic accountability, in particular as a means for ensuring procedural regularity in algorithmic decision-making and for testing properties of the underlying algorithmic policy, *id.* at 673 ("[P]ublished commitments and zero-knowledge proofs allow overseers and the public at large to verify that the decisions of some authority actually correspond to a specific predetermined policy rather than the arbitrary whim of a decisionmaker."). The second, Yuqing Cui, Note, *Application of Zero-Knowledge Proof in Resolving Disputes of Privileged Documents in E-Discovery*, 32 HARV. J. L. & TECH. 633 (2019), provides a description of the technology and proposes the use of ZKPs to help resolve claims of privilege in e-discovery.

contexts in which these dilemmas recur: (1) *information privacy and security* and the need to disclose one's identity in order to verify certain personal attributes and permissions; (2) *dealmaking* and the necessity of verifying deal-worthiness by revealing sensitive business information; (3) *government oversight* and the tension with the requirement, in some circumstances, that the secrecy of sensitive government algorithms and data be preserved; and (4) litigation that depends on *trade secrets* yet risks compromising them. In each of these contexts, existing legal remedies are often imperfect and can lead to either costly nondisclosure or harmful overdisclosure.

Part III explains zero-knowledge proofs and the ways in which they can change the all-or-nothing disclosure calculus by permitting verification through the partial revelation of information. This Part provides both a conceptual framework for understanding the power (and limitations) of ZKPs and a technical primer for a legal audience on how ZKPs work. By upending assumptions about the scope of information that needs to be disclosed for verification, ZKPs offer an example of a technology that might allow the unshackling of existing legal approaches from those assumptions.

To work through the ZKPs' promise and the issues their implementation would raise, Part IV explores a series of case studies posing verification problems from the four contexts discussed in Part II. For each context, cases are divided into "easier use" cases in which ZKPs have garnered notice to date, and "harder use" novel cases. As a practical matter, these cases suggest the ways that ZKPs might enhance the protection of personal privacy and data security, enable an entire class of beneficial transactions that currently do not occur, promote government accountability by changing the calculus in contexts in which circumvention or gaming concerns undergird public secrecy, and permit the litigation of meritorious legal claims while precluding gaming and curtailing unscrupulous actors. Importantly, applying ZKPs in each of these contexts would reduce but not entirely eliminate the need for trust. As discussed throughout the Article, a ZKP must be accompanied by an additional guarantee that the data considered in the mathematical proof are the same as the objects considered in the legal verification dilemma. The mechanisms to provide that additional guarantee would need to be context-specific.

Finally, Part V considers the policy implications of these case studies for the use of ZKPs as an information governance tool and develops a framework for designing and evaluating appropriate ZKP tools for specific legal, technological, or institutional solutions to verification dilemmas. It also discusses the range of policy choices around disclosure which, in turn, challenges the existing policy balance between disclosure and secrecy. This

balance is rooted in assumptions about the technical capacity for verification that ZKPs may supersede. Surfacing these altered presumptions is necessary to enable critical policy discussions about whether ZKPs should be used in a particular context, what information should be disclosed or kept secret, and how these decisions should be made.

II. VERIFICATION DILEMMAS AND THE LAW

Law is full of verification and trust problems. Verifying facts about information to an untrusting party, such as a potential contracting partner or a litigation adversary, often requires disclosing a substantial amount of sensitive information. If undertaken, the disclosure in turn creates risks that the disclosed information will be misused. Different legal doctrines attempt to solve verification dilemmas in different ways.

This Part presents examples of verification dilemmas drawn from four doctrinal contexts: information privacy and security; dealmaking; government oversight; and trade secret litigation. Although these examples implicate a wide range of policy concerns and legal rules, they all revolve around the core issue of how to verify facts about information while protecting the information from misuse. Importantly, the legal rules on which each of these doctrines relies to solve the verification problem are imperfect. On the one hand, where current legal practice requires or encourages disclosure, the law often seeks to deter post-disclosure misuse of information by offering ex post remedies. However, those remedies are available solely when misuse is detected, and they can be prohibitively costly to pursue. On the other hand, where current legal practice restricts or discourages disclosure, the law sacrifices the trust and verifiability that transparency could provide.

Examining each of the examples below through the lens of the verification dilemma offers a new way of thinking about and analyzing these legal doctrines. More specifically, it exposes certain shared organizing assumptions behind these seemingly-disparate legal rules—a revelation which in turn raises the possibility of devising novel, transdoctrinal solutions to the verification problems that current legal rules are designed to address.

A. INFORMATION PRIVACY AND SECURITY: LINKING VERIFICATION WITH IDENTIFICATION

1. Verification Dilemmas

A key set of verification dilemmas involve the linkage between verification and identification. In current practice, persons seeking to verify particular facts—for instance, that they are over eighteen years of age, are citizens of the United States, have rights to access digital spaces, or possess certain financial

capacity—must frequently reveal other information about who they are.[15] These disclosures are costly in terms of privacy and data security.

As to privacy, revealing who one is exposes unnecessary additional private information to the recipient. More significantly, the amalgamation of large quantities of data online compounds the threat from even apparently small but unnecessary additional disclosures. A mere name disclosure can link a real-world individual and their "digital person"—the "extensive aggregations of data about a person in many databases."[16] For instance, the link could expose private medical information, location histories, or educational records in inappropriate contexts.[17] Because one's personal identity is generally fixed throughout life, disclosing one's personal identity to verify limited personal attributes has particular costs in the big data era. As Dan Solove explains, if aggregation permits the creation of the "digital person," identification "goes a step further—it links the digital person directly to a person in real space."[18] The pernicious nature of this linkage can extend even further, as identification systems "piggyback" on one another, reproducing and reinforcing judgments made about, and now linked to, a person's identity.[19]

Excessive data aggregation also has repercussions beyond potentially harming the individuals whose data are aggregated. It allows the aggregator to discern trends in communities as well as individuals, and use the acquired knowledge to potentially harm third parties or the population at large.[20] If personal identity information were less frequently revealed, then aggregating data at a community level would be harder and thus potentially more limited.

Furthermore, the overdisclosure of personal information, typical of traditional means of user authentication, expands the data exposed to

15. Maria Dubovitskaya, *Take Back Control of Your Personal Data*, TED (Oct. 2015), https://www.ted.com/talks/maria_dubovitskaya_take_back_control_of_your_personal_data#t-1760 ("Whether buying a bottle of wine, making an online purchase or going to a movie, most of us share far more information than is necessary: birthdates, credit card numbers, addresses.").

16. Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 511, 513 (2006).

17. HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE 129 (2010) (discussing how privacy harms occur when "context-relative informational norms" are violated).

18. Solove, *supra* note 16, at 513.

19. *See* JIM HARPER, IDENTITY CRISIS: HOW IDENTIFICATION IS OVERUSED AND MISUNDERSTOOD 77–80 (2006) (discussing decisions by one institution determining the "suitability" of whether to adopt identification decisions by other institutions).

20. *See, e.g.*, Matthew Hindman, *How Cambridge Analytica's Facebook Targeting Model Really Worked—According to the Person Who Built It*, THE CONVERSATION (Mar. 30, 2018), https://theconversation.com/how-cambridge-analyticas-facebook-targeting-model-really-worked-according-to-the-person-who-built-it-94078.

vulnerabilities in security and management. Whether through malicious data breaches, error-based data breaches, or weaknesses in integrated third-party systems that store or transmit passwords, the costs of data theft or misuse are daunting.[21] Over 33% of adults in the United States have experienced identity theft, the act of impersonating others' identities by presenting stolen identifiers or proofs of identity, leading to significant financial and reputational harm.[22] Moreover, one in five companies (19%) that suffered a malicious data breach was infiltrated because of stolen or compromised personal credentials.[23]

2. Imperfect Legal Responses

Despite these costs, current law and policy often come down on the side of permitting or mandating the overdisclosure involved when identification is required for verification. Privacy threats cede to the benefits of accuracy, security, administrability, and efficiency understood to inhere in such overinclusive means of verification. U.S. privacy law recognizes the sensitivity of "personally identifiable information" (PII).[24] Yet it does not generally address the precedent issue of whether identification should occur at all. PII generally acts as a regulatory trigger: privacy laws largely apply only once PII is collected. These laws do not control whether PII is initially collected.[25]

To be sure, policymakers have recognized the costs engendered by the overdisclosures currently required for verification. Policymakers have also specifically recognized the dangers of linking verification and identification. Privacy law explicitly seeks legal solutions to mitigate the risks, but to date those solutions are imperfect. For instance, legal requirements to protect PII

21. *See, e.g.*, *US Expected to Break Data Breach Record in 2021*, SEC. MAG. (Oct. 15, 2021), https://www.securitymagazine.com/articles/96318-us-expected-to-break-data-breach-record-in-2021.

22. *Global Cybersecurity Awareness Survey Reveals 33 Percent of U.S. Respondents Have Experienced Identity Theft, More than Twice the Global Average*, PROOFPOINT (Oct. 11, 2018), https://www.proofpoint.com/us/newsroom/press-releases/global-cybersecurity-awareness-survey-reveals-33-percent-us-respondents-have.

23. IBM SEC., COST OF A DATA BREACH REPORT 2020 9 (2020) https://www.ibm.com/security/digital-assets/cost-data-breach-report/#/ (calculating the average cost of a data breach at almost $4,000,000).

24. Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814, 1816 (2011). The European General Data Protection Regulation (GDPR), focuses on the category of "personal data"—defined as "any information relating to an identified or identifiable natural person," Commission Regulation 2016/679, 2016 O.J. (L 119) (EU), at art. 4(1) [hereinafter GDPR].

25. *See, e.g.*, California Consumer Privacy Act of 2018, CAL. CIV. CODE §§ 1798.100–199.100 (West 2021) [hereinafter CCPA] (providing consumers with a right to know what information is collected about them and a right to opt-out of resale of that information but imposing no other limits on the initial scope of collection).

by mandating anonymization prior to data redistribution are infamously unreliable.[26] They are unreliable because existing technical strategies for performing anonymization are weak and can often be defeated by counter-technical measures.[27] Similarly, laws requiring "notice and consent" before PII may be collected or used have been shown to be ineffective both in communicating the risks to those whose information is at issue and in providing an alternative that would make consent meaningful.[28] Although ex post data breach notification requirements and penalties for failure to maintain "reasonable" security practices[29] seek to provide financial and reputational incentives for better security, such requirements do not address information collection in the first place.

Furthermore, in some cases legal measures intended to protect consumer privacy require verification by means of identification. For example, the California Consumer Privacy Act[30] requires businesses to provide customer access to information about the data kept about them—a privacy-preserving provision intended to foster individual control of information.[31] However, the law provides that the access right is triggered only by consumers' "verifiable"

---

26. *See generally* Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* (Carnegie Mellon Univ., Data Privacy Working Paper No. 3, 2000), https://dataprivacylab.org/projects/identifiability/paper1.pdf (conducting groundbreaking research on the large percentage of the population that can be uniquely identified by ZIP code, birth date, and gender); *see* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1705 (2010) (discussing the literature on reidentification and the ways in which it will amplify privacy harms because "[r]eidentification combines datasets that were meant to be kept apart, and in doing so, gains power through accretion" that "makes all of our secrets fundamentally easier to discover and reveal").

27. Ohm, *supra* note 26, at 1717–22 (describing examples of deanonymization).

28. Lindsey Barrett, *Model(ing) Privacy: Empirical Approaches to Privacy Law and Governance*, 35 SANTA CLARA COMPUT. & HIGH TECH. L.J. 1, 17–20 (2018) (summarizing studies that find, among other things, that most consumers do not understand basic facts about the use of their data); Ehimare Okoyomon, Nikita Samarin, Primal Wijesekera, Amit Elazari Bar On, Narseo Vallina-Rodriguez, Irwin Reyes, Álvaro Feal & Serge Egelman, *On The Ridiculousness of Notice and Consent: Contradictions in App Privacy Policies*, IEEE WORKSHOP ON TECH. & CONSUMER PROTECTION (2019), https://www.ieee-security.org/TC/SPW2019/ConPro/papers/okoyomon-conpro19.pdf (discussing the gaps between disclosed data collection practices as articulated in privacy policies and de facto data collection practices as observed using dynamic analysis tools); *see* Jonathan A. Obar & Anne Oeldorf-Hirsch, *The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services*, 23 INFO., COMM. & SOC. 1, 25 (2018) (noting that of more than 500 surveyed users, 93% accepted a first-born child assignment term and 98% ignored or missed it).

29. *See, e.g.*, CAL. CIV. CODE §§ 1798.81.5(a)–(b) (obligating a company that processes personal information about a California resident "to implement and maintain reasonable security procedures and practices" appropriate to the nature of the information it processes).

30. CAL. CIV. CODE §§ 1798.100–199.100 (West 2021) .

31. *Id.* § 1798.100(a).

access requests.[32] This seems to require that both the data and request are associated with or "linkable" to the requestor's legal identity.[33]

In sum, a recurring challenge for privacy and security is to enable persons to verify facts about themselves or authenticate their rights of access while limiting the risks of revealing one's identity or other sensitive information. As Maria Dubovitskaya, a cryptographer at IBM's Research Lab, puts it: "If your personal data is never collected, it cannot be stolen."[34] To date, legal rules have had little success in promoting this goal.

B. DEALMAKING: VERIFICATION AND ARROW'S PARADOX

1. Verification Dilemmas

The *TargetSmart* case described at the beginning of this Article[35] underscores a type of recurring verification dilemma involving confidential or sensitive business information. Specifically, a target firm or potential partner is faced with a quandary during the due diligence process: The party must disclose enough information to convince the other party—often a rival—to enter deeper negotiations or engage in the transaction. At the same time, this information—including source code, proprietary processes, customer data and contracts, and pricing information—is often confidential and sensitive yet sometimes unprotected by traditional forms of intellectual property protection. This dilemma occurs in a variety of contexts, such as seeking venture funding or negotiating a merger, acquisition, partnership, or joint venture.

This bind exemplifies what economist Kenneth Arrow termed the "fundamental paradox" of information disclosure.[36] Professors Gill and Parchmovosky summarized Arrow's point in the following manner: "information that is not afforded legal protection," such as sensitive business information that is not a trade secret, "cannot be bought or sold on the market" because "in order to sell the information, [a seller] must disclose it to the potential buyer; but once she does, she has nothing left to sell."[37] Because the disclosure of information vitiates its control, revealing information when

32. *Id.* § 1798.100(c).

33. *See* Rebecca Iafrati, *Can the CCPA Access Right Be Saved? Realigning Incentives in Access Request Verification*, 20 J. TECH. L. & POL. 25, 25–27 (2020).

34. Dubovitskaya, *supra* note 15.

35. *See supra* in Part I.

36. Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources for Invention*, *in* THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS 609, 615 (Nat'l Bureau of Econ. Rsch. ed., 1962).

37. Oren Bar-Gill & Gideon Parchomovsky, *Law and the Boundaries of Technology-Intensive Firms*, 157 U. PA. L. REV. 1649, 1653–54 (2009).

disclosure is necessary for verification creates risks of appropriation of proprietary information. When the risk of disclosure is judged too great, the verification dilemma will impede value-creating transactions.[38]

Although this "fundamental paradox" threatens information exchange generally, the possibility for appropriation of sensitive information is particularly menacing in the dealmaking context, where acquiring firms, potential funders, and promising partners are often rivals, or at least operate in a similar business niche. To make matters worse, sometimes, as in *TargetSmart*, "competitors can initiate M&A discussions as a strategic tactic to gain access to sensitive proprietary information and to exploit that information at the expense of the disclosing party."[39]

2. Imperfect Legal Responses

With few exceptions—notably when sharing proprietary information with rivals can raise antitrust concerns[40] or constitute contractual breach of confidentiality clauses[41]—the law reflects a policy choice to leave decisions to disclose many categories of proprietary information to private ordering. Most concretely, significant swaths of sensitive information are left unprotected by intellectual property law once it is disclosed.

The law reflects the costs of disclosure by recognizing the enforceability of contractual NDAs into which parties may enter before negotiations begin.[42]

38. Michael J. Burstein, *Exchanging Information Without Intellectual Property*, 91 TEX. L. REV. 227, 242 (2012) ("An inventor seeking funds or development expertise may be reluctant to disclose information about her invention for fear that the recipients of the information can take it for themselves. On the other side of the transaction, the funders or developers will be unwilling to commit money or resources to the project unless or until they can assess its value.").

39. Jason Bullen & Xi Chen, *Managing Disclosure Risk in M&A Transactions*, MANDAQ (Apr. 26, 2017), https://www.mondaq.com/canada/maprivate-equity/588906/managing-disclosure-risk-in-ma-transactions.

40. *See, e.g.*, Holly Vedova, Keitha Clopper & Clarke Edwards, *Avoiding Antitrust Pitfalls During Pre-Merger Negotiations and Due Diligence*, COMPETITION MATTERS (Mar. 20, 2018, 4:57 PM), https://www.ftc.gov/news-events/blogs/competition-matters/2018/03/avoiding-antitrust-pitfalls-during-pre-merger (noting that the Commission "looks carefully at pre-merger information sharing to make sure that there has been no inappropriate dissemination of or misuse of competitively sensitive information for anticompetitive purposes").

41. Aaron Binstock, *10 Considerations to Protect Confidential Information When Selling Your Company*, LEXOLOGY (Mar. 5, 2015), https://www.lexology.com/library/detail.aspx?g=6960d8dc-61f0-4659-9a0c-7e4d4d40bd69 (noting that, in due diligence, "it is possible to breach a contract just by disclosing its existence").

42. *See* Henry Lesser, Ann Lederer & Charles Steinberg, *Increasing Pressures for Confidentiality Agreements that Work*, MERGERS & ACQUISITIONS, Mar.–Apr. 1992, at 23, 23 (describing the basic function of the agreement as "protecting sellers against misuse of confidential information").

As a remedy for the dilemma of verification, however, such ex post legal remedies are imperfect at best.[43] NDAs are difficult to enforce for reasons ranging from the cost and complexities of proof—that is, was an idea or method stolen or derived independently?—to the establishment of the secrecy of the underlying material.[44]

Even when NDAs are enforceable, they merely provide an opportunity to receive damages after a violation has occurred through unauthorized dissemination of confidential information. They do not prevent the harm of such use or dissemination from happening in the first place. And they are not enforceable against third parties with whom the negotiating partner might have shared sensitive information. Moreover, NDAs cannot prevent the transfer of "knowledge spillovers" inherent in sharing information, which occur even if the recipients never leak or knowingly misuse that information. Information cannot be unseen, and an NDA cannot prevent the proliferation of knowledge gains that mere viewing may precipitate. Finally, in a variety of contexts, such as in venture capital, imbalances in negotiating power have allowed the development of customs by which parties refuse to sign NDAs.[45]

According to Michael Burstein, the challenge for firms seeking to protect sensitive information while preserving their ability to attract transaction partners lies in finding "some kind of optimum level of appropriability that allows for (a) sufficient information to be transferred to link ideas with capital

43. Practical "best practices" in due diligence are often similarly incomplete. The use of "clean rooms," by which companies in negotiations can post sensitive information viewable to a limited number of the opposing party's representatives, may offer technical protections against direct copying of documents or other materials. But clean rooms do not eliminate the reality that the material is being viewed by outsiders who, even if they are acting in good faith, are specifically charged with relaying pertinent information to decisionmakers in their firms. The cognitive reality is that once innovative information is understood, it cannot be "unlearned."

44. *See, e.g.*, nClosures Inc. v. Block and Co., Inc., 770 F.3d 598, 602 (7th Cir. 2014) (refusing to enforce a nondisclosure agreement entered into before plaintiff shared propriety designs and manufacturing knowledge with defendant during negotiations to form a partnership because, inter alia, plaintiff had previously provided its design files to a third-party company that initially manufactured its products without requiring confidentiality agreements).

45. Sergio Marrero, *Why Venture Capitalists Don't Sign NDAs*, MEDIUM.COM (Aug. 19, 2019), https://medium.com/rbl1/why-venture-capitalists-dont-sign-ndas-c87e331a5505; Guy Kawasaki, *The Venture Capitalist Wishlist*, (Jan. 16, 2006), https://guykawasaki.com/the_venture_cap-2 ("Before you even start addressing the hard stuff, never ask a venture capitalist to sign a non-disclosure agreement (NDA). They never do. This is because at any given moment, they are looking at three or four similar deals. They're not about to create legal issues because they sign a [sic] NDA and then fund another, similar company—thereby making the paranoid entrepreneur believe the venture capitalist stole his idea. If you even ask them to sign one, you might as well tattoo 'I'm clueless!' on your forehead.").

and development partners while (b) ensuring that enough value remains in the original information holder so that she still has an incentive to disclose."[46] Burstein suggests two options to balance these interests. First, especially in contexts in which confidential knowledge is tacit, transacting firms can choose to "codify"—and therefore make transferable—only certain aspects of the confidential knowledge. Second, firms can selectively disclose information by "partition[ing] their information so as to reveal some but not all of the relevant information to counterparties"; for example, firms can use modularity in software design to shield certain portions.[47]

As we discuss below in Part III, ZKPs promise to become a powerful enabler and enhancer of Burstein's approach: They allow the secret-holder to disclose partial information about the secret without disclosing the rest, all while reassuring the recipient of the disclosed information that what was disclosed is a valid portion of the secret.

C. GOVERNMENT OVERSIGHT: VERIFICATION AND DATA/ALGORITHMIC ACCOUNTABILITY

1. Verification Dilemmas

Another set of verification dilemmas arise from the fact that the public interest in government accountability through transparency can clash with perceived needs to keep details about data and algorithms secret. Regulator-regulate data exchanges illustrate this tension. Consider Federal Reserve stress tests as a particularly poignant and high-impact example. The Dodd-Frank Act requires the Federal Reserve to annually conduct stress tests of bank balance sheets, ensuring that they can survive a financial crisis without collapsing.[48] These annual stress tests are of significant consequence for financial institutions because the Federal Reserve is empowered under law to take early remediation measures, including cutting or halting dividends or preventing acquisitions.[49] The Federal Reserve guidelines are extremely complex.[50] The formulation itself is complicated, but the details make it even more so because banks trade in securities of uncertain value that require parsing to evaluate risk;

46. Burstein, *supra* note 38, at 254.

47. *Id.*

48. Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111–203, § 165(h)(4)(i) (2010).

49. *Id.* § 166(c)(2).

50. *See Supervisory Stress Test Framework and Model Methodology, Dodd-Frank Act Stress Test 2019: Supervisory Stress Test Results June-2019*, BD. OF GOVERNORS OF THE FED. RSRV. SYS. (Jul. 16, 2019), https://www.federalreserve.gov/publications/june-2019-supervisory-stress-test-framework-and-model-methodology.htm.

since these securities are often illiquid and opaque, the essence of the stress test is ensuring the validity of these calculations.

Despite the importance of these exchanges, this process is surrounded with an enormous amount of secrecy. To start, the tests and scoring criteria are secret.[51] James McAndrews, the chief of research at the Federal Reserve Bank of New York, noted that this secrecy is necessary because publicly disclosing the test's details would make it easier for banks to circumvent the test.[52] Furthermore, banks are not required to disclose much information to the public about their results. Scholars have argued that this too serves a purpose, because too much disclosure would incentivize poor behavior by individual managers, such as holding on to suboptimal loans to game a test.[53] Yet, though the secrecy serves a purpose, it comes with costs. The banks have found that their numbers diverge with the Federal Reserve due to the secrecy of the tests themselves, which results in significant annual controversy.[54] A Government Accounting Office (GAO) report also noted that a lack of transparency risked undermining the stress test's efficacy by inhibiting compliance.[55] At the same time, the lack of transparency may reduce public trust in the regulatory agency, especially when the same secrecy necessary to avoid circumvention introduces risks of regulatory capture because there is no one to watch the watchmen.[56]

Government accountability and transparency can also stand in direct conflict with the need to preserve individuals' privacy. Quintessential examples are government policy decisions based on direct polling of private and identifying information about individuals (such as in the decennial census) or, alternatively, on scientific studies that are in turn based on protected (e.g., medical) information. For instance, this conflict is at the base of controversy

51. Victoria McGrane, *Fed Stands Firm Against Revealing Bank Stress-Test Model*, WALL ST. J. (June 24, 2015), https://blogs.wsj.com/economics/2015/06/24/fed-stands-firm-against-revealing-bank-stress-test-model/.

52. Francine McKenna, *Fed Says Stress Test Models Will Stay a Secret*, MARKETWATCH (June 25, 2015), https://www.marketwatch.com/story/fed-says-stress-test-models-will-stay-a-secret-2015-06-25.

53. Itay Goldstein & Haresh Sapra, *Should Banks' Stress Test Results be Disclosed? An Analysis of the Costs and Benefits*, 8 FOUND. & TRENDS FINANCE 1, 44–47 (2014).

54. *Id.*

55. U.S. GOV'T ACCOUNTABILITY OFF., GAO-17-48, ADDITIONAL ACTIONS COULD HELP ENSURE THE ACHIEVEMENT OF STRESS TEST GOALS 90 (2016).

56. For example, in 2014 Carmen Segarra, a former Federal Reserve employee, blew the whistle on Federal Reserve support granted to Goldman Sachs to circumvent their shortcomings for several years. *See* Nathaniel Popper & Peter Eavis, *Secret Goldman Sachs Tapes Put Pressure on New York Fed*, DEALBOOK (Oct. 2, 2014), https://dealbook.nytimes.com/2014/10/02/secret-goldman-sachs-tapes-put-pressure-on-new-york-fed/.

around a recent Environmental Protection Agency rule requiring researchers to disclose the raw data involved in their public health studies before the agency could rely upon their research conclusions.[57]

Beyond data, government reliance on algorithms can also present verification dilemmas. For example, in 2013, a GAO report indicated that the IRS was using "inappropriate" means to audit certain 501(c)(4) organizations that it believed might be abusing their tax-exempt status to engage in prohibited political activities.[58] Specifically, the IRS was using certain keywords in the names of the nonprofit filings as a factor in whether the applications merited further review, such as "tea party."[59] Although progressive applicants were also subject to review, conservative groups appeared to receive greater scrutiny as measured by the different rates by which the groups were flagged for further review, leading to accusations of bias.[60] Put another way, the IRS had an audit algorithm[61] that had a need for public accountability to ensure it did not have a particular deleterious characteristic. The simple solution is to have a public audit or a trusted entity like the GAO publish a report. However, a public audit would not be feasible because exposing any IRS audit mechanism would make it easier to circumvent, risking an increase in tax fraud. And while the GAO may be highly competent and nonpartisan, at least some of the public may not trust the government to audit itself.

Another particularly high stakes instance of verification dilemmas concerns algorithms used in the criminal legal system to perform surveillance, conduct investigations, analyze forensic evidence, or predict risk of future offenses to guide pretrial incarceration and sentencing decisions. Life, liberty, police accountability, constitutional privacy, racialized mass incarceration, and public safety are all on the line.[62] Criminally accused persons have powerful

57. Strengthening Transparency in Pivotal Science Underlying Significant Regulatory Actions and Influential Scientific Information, 86 FED. REG. 469 (Jan. 6, 2021) (to be codified at 40 C.F.R. pt. 30).

58. TREASURY INSPECTOR GEN. FOR TAX ADMIN., INAPPROPRIATE CRITERIA WERE USED TO IDENTIFY TAX-EXEMPT APPLICATIONS FOR REVIEW 5 (2013), https://www.treasury.gov/tigta/auditreports/2013reports/201310053fr.pdf.

59. *Id.* at 6.

60. John D. McKinnon, *IRS Inspector Firm on One-Sided Targeting*, WALL ST. J. (June 27, 2013), https://www.wsj.com/articles/SB10001424127887323873904578571363311816922 ("Internal Revenue Service employees flagged for extra scrutiny fewer than a third of progressive groups applying for tax exemptions from mid-2010 through mid-2012, compared with 100% of conservative applicants.").

61. *See* PEDRO DOMINGOS, THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD 1 (2015) ("An algorithm is a sequence of instructions telling a computer what to do.").

62. For a compelling recent critique of risk assessment instruments that rely on carceral data and discredit community knowledge sources, see Ngozi Okidegbe, *Discredited Data*, 107

constitutional, statutory, and common law rights to scrutinize and test the evidence against them, including outputs from investigative[63] and forensic software tools.[64] Moreover, the public has an interest in democratic oversight of police and forensic technologies and the criminal court proceedings that rely on them, in part to ensure accurate outcomes that properly balance protections for individual rights and public safety.[65]

Meanwhile, there can be compelling security reasons to maintain some secrecy concerning some algorithms used in the criminal legal system. For instance, if algorithms that flag suspicious activity associated with illegal trading were publicly disclosed, then it would be easier for insider traders to avoid getting caught.[66] Likewise, if algorithms used to identify child sexual abuse materials (CSAM) on the internet were publicly disclosed, then it would be easier for CSAM possessors and distributors to evade detection.[67] Yet even legitimate secrecy interests can conflict with criminal defendants' rights to scrutinize the evidence against them and with the public's interest in oversight of criminal proceedings. There is also a risk that law enforcement or algorithm developers might overclaim their secrecy interests, whether due to mistake, exaggeration, or an attempt to evade scrutiny of potential flaws or biases in the algorithms themselves.[68] Courts generally defer to these types of security secrecy claims,[69] so there are few examples of courts ordering disclosure that resulted in the exposure of flaws or biases in the tools.

CORNELL L. REV. (forthcoming 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3835414.

63. *See* Elizabeth E. Joh, *The Corporate Shadow in Democratic Policing: Technology Companies can Elude Accountability*, 374 SCI. 274, 275 (2021).

64. *See generally* Andrea Roth, *Machine Testimony*, 126 Yale L.J. 1972, 1980 (2017) (outlining "testimonial safeguards for machine sources of information").

65. *See generally* Hannah Bloch-Wehba, *Democratic Algorithms* (Feb. 28, 2021) (unpublished manuscript) (on file with the Berkeley Technology Law Journal) (exploring activist group calls for greater public participation in decisions surrounding the use of AI and machine-learning technologies in legal institutions).

66. *See, e.g.*, Todd Ehret, *SEC's Advanced Data Analytics Helps Detect Even the Smallest Illicity Market Activity*, REUTERS (June 30, 2017), https://www.reuters.com/article/bc-finreg-data-analytics-idUSKBN19L28C (describing SEC use of data analytics to detect insider trading).

67. *See generally* Jonathan Manes, *Secrecy & Evasion in Police Surveillance Technology*, 34 BERKELEY TECH. L.J. 503 (2019) (generally describing the anti-circumvention rationale for law enforcement secrecy).

68. *Cf.* United States v. Reynolds, 345 U.S. 1, 9–10 (1953) (warning that in national security context, "judicial control over the evidence in a case cannot be abdicated to the caprice of executive officers"); Herring v. United States; 424 F.3d 384, 386 (3d Cir. 2005) (discussing, though ultimately rejecting, the allegation that the government had fraudulently asserted the state secrets privilege in *Reynolds*).

69. *See, e.g.*, United States v. Pirosko, 787 F.3d 358, 365–67 (6th Cir. 2015).

Nonetheless, the risk of error or misconduct that arises from secrecy surrounding government algorithms can be illustrated by a case in which the secret was ultimately deemed illegitimate and the information was fully disclosed. In 2008, the New York City Office of Chief Medical Examiner (the OCME) began developing the Forensic Statistical Tool (FST), a software program designed to statistically analyze complex mixtures of DNA found at crime scenes.[70] To gain regulatory approval for FST, the OCME conducted over a multi-year development process repeated presentations to the DNA Subcommittee of the New York State Commission on Forensic Science, before beginning to use the program in criminal cases in 2011.[71] New York State criminal courts subsequently relied on the approval by the Forensic Science Commission, along with internal validations conducted by the OCME, to determine that outputs of the FST software program met the requirements for admissibility in court.[72] Meanwhile, the OCME adopted a secretive stance surrounding the program, calling it "proprietary" and refusing to disclose the source code to criminal defense counsel, or even to share an executable copy of the program for independent testing by defense expert witnesses.[73]

In 2016—five years after the implementation of FST in criminal cases—Judge Valerie Caproni of the United States Court for the Southern District of New York held that the OCME did not have a legitimate "proprietary" interest in withholding the FST source code from criminal defense counsel and ordered the OCME to reveal the code under a protective order.[74] A defense expert witness examining the code then identified an undisclosed function that discarded data in certain circumstances without notice to the user, a function not clearly described in publications detailing the FST methodology.[75] The defense witness argued that the undisclosed function must have been added

70. STATE OF N.Y. OFF. OF THE INSPECTOR GEN., INVESTIGATION INTO THE NEW YORK CITY OFFICE OF CHIEF MEDICAL EXAMINER: DEPARTMENT OF FORENSIC BIOLOGY 27–28 (2013), https://ig.ny.gov/sites/g/files/oee571/files/2016-12/OCMEFinalReport.pdf.

71. *Id.* at 28. The New York State Commission on Forensic Science is the regulatory oversight body for the OCME. *See* N.Y. EXEC. LAW § 995-b (McKinney 2018).

72. *See* People v. Williams, 35 N.Y.3d 24, 147 N.E.3d 1131, at *35–36 (Mar. 31, 2020).

73. *See* Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1362 n.80 (2018).

74. Order, United States v. Johnson, No. 1:15-cr-00565-VEC (S.D.N.Y. July 6, 2016).

75. Decl. of Nathaniel Adams at 20, United States v. Johnson, No. 1:15-cr-00565-VEC, (S.D.N.Y Oct. 27, 2016).

after the internal validation studies and regulatory approval on which the courts had relied.[76] Thousands of cases were potentially thrown into disarray.[77]

Ultimately, there was no good reason to keep the FST algorithm secret, and hence the OCME's secretive stance was not a true verification dilemma. When secrecy surrounding government information is unjustified or outweighed by countervailing transparency interests, there should be no secret at all.[78] But similar concerns as applied in the FST case also appear in cases where the secrecy interests ca be more credible. In the FST case, the issue at hand boiled down to a relatively simple and well-defined one: Is the evidence presented in court the result of applying the certified (but secret) algorithm to the relevant (known) data?[79] Similar questions can arise with algorithms that must remain secret to stay effective. In addition, there can also be questions about the validity and appropriateness of the algorithms themselves.

2. Imperfect Legal Responses

The example of FST described *supra* illustrates how keeping algorithms secret exacerbates the risk of mishandling algorithms that are critical to the legal process, especially in high stakes scenarios like the criminal legal system: A government actor might obtain regulatory approval to use a software system but then alter the software prior to implementation without revalidating the system or seeking additional regulatory approval. Guidance and standards documents for forensic software systems recommend that "significant" or "core" changes to software should require additional validation.[80] But even

76. Decl. of Nathaniel Adams at 5–6, United States v. Johnson, No. 1:15-cr-00565-VEC, (S.D.N.Y Feb. 12, 2017).

77. *See* Lauren Kirchner, *Traces of Crime: How New York's DNA Techniques Became Tainted*, N.Y. TIMES (Sept. 4, 2017), https://www.nytimes.com/2017/09/04/nyregion/dna-analysis-evidence-new-york-disputed-techniques.html.

78. For instance, Amy Kapczynski offers a powerful critique of corporate claims to secrecy in information disclosed to regulators but not to the public, including information about fracking chemicals, drug prices, and pharmaceutical clinical trial data. *See* Amy Kapczynski, *The Public History of Trade Secrets*, 55 U.C. DAVIS L. REV. 1367, 1373–76 (2022).

79. This authentication question of whether a software program used in a particular case was the program previously approved by regulators is a recurring issue in a variety of contexts. Deven Desai and Joshua Kroll have identified similar issues concerning regulatory accountability for software used in automobiles and voting machines and have also proposed zero-knowledge proofs as a solution to these problems. *See* Deven R. Desai & Joshua A. Kroll, *Trust But Verify: A Guide to Algorithms and the Law*, 31 Harv. J. L. & Tech. 1, 14–16, 47 (2017).

80. *See* SCI. WORKING GRP. ON DNA ANALYSIS METHODS, GUIDELINES FOR THE VALIDATION OF PROBABILISTIC GENOTYPING SYSTEMS 11 (2015), https://1ecb9588-ea6f-4feb-971a-73265dbf079c.filesusr.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf; M. D. Coble, J. Buckleton, J. M. Butler, T. Egeland, R. Fimmers, P. Gill, L. Gusmão, B. Guttman, M. Krawczak, N. Morling, W. Parson, N. Pinto, P. M. Schneider, S. T. Sherry, S. Willuwiet & M. Prinz, *DNA Commission of the*

experts often disagree on what constitutes a "significant" or "core" change,[81] so employees tasked with routine maintenance of a software system might be unaware of the significance of an alteration.

The current legal check on employee misjudgment in these types of cases is, as occurred with FST, to disclose the source code to a defense expert witness in ex post litigation. However, a perceived necessity to keep the algorithm itself hidden can render this legal check powerless and ineffective. In the FST example, as soon as algorithmic secrecy was no longer perceived as necessary, the legal system's existing check regained its power and the undisclosed function in the FST algorithm was exposed.

This of course raises the question of what happens in other cases where algorithms critical to legal cases are kept hidden. As described above, and in contrast to FST, there can be circumstances where disclosing an algorithm's source code might risk substantial harm, such as by enabling future wrongdoers to evade detection. That is, the government and other entities will often keep algorithms secret because they are used in an adversarial process, often in a security or regulatory context, where the perceived danger is that exposing the underlying information would allow for regulated parties to circumvent the algorithms by knowing where the tripwires lie.[82] It is likely that in those kinds of "anti-circumvention" cases, the type of undisclosed code alteration that seemingly occurred with FST could go undetected indefinitely. Such circumstances are not limited to criminal cases but appear across an array of government accountability contexts.

Where ex post remedies are deemed insufficient to mitigate the risk of post-disclosure misuse, a frequent legal solution to the verification problem in cases associated with an anti-circumvention rationale for secrecy is to have no disclosure at all. For instance, when criminal defendants raise doubts about the legality or reliability of a confidential law enforcement investigative tool, courts frequently hold that information about how the tool works is entirely shielded

International Society for Forensic Genetics: Recommendations on the Validation of Software Programs Performing Biostatistical Calculations for Forensic Genetics Applications, 25 FORENSIC SCI. INT'L: GENETICS 191, 196 (2016).

81. Decl. of Nathaniel Adams at 7, United States v. Johnson, No. 1:15-cr-00565-VEC, (S.D.N.Y Feb. 12, 2017).

82. *See, e.g.*, Andrew Moshirnia, *No Security Through Obscurity: Changing Circumvention Law to Protect our Democracy Against Cyberattacks*, 83 BROOK. L. REV. 1279 (2018) (explaining that the government often keeps national security information secret under an anti-circumvention justification).

from disclosure by a governmental privilege.[83] Often criminal defense experts are not even permitted to test an executable version of the software program.[84] Hence, the law presumes a tradeoff between verification and competing values, and sacrifices the trust and accountability that greater transparency could facilitate.

On the other hand, in the less-common scenario where courts find that criminal defense rights necessitate disclosure of information about law enforcement algorithms, the prosecution sometimes elects to drop criminal charges and permit a suspected criminal to evade punishment rather than comply with a court-ordered disclosure that risks undermining the efficacy of future investigations.[85] This scenario, often called the disclose-or-dismiss dilemma, risks harms to public safety, justice, and fairness.[86] For classified information specifically, the Classified Information Procedures Act mitigates the disclose-or-dismiss dilemma in criminal cases by specially permitting partial and protected disclosures.[87] However, those procedures are unavailable for sensitive but unclassified information and for civil and regulatory proceedings.

D. TRADE SECRET LITIGATION: VERIFICATION DILEMMAS IN THE ADVERSARY PROCESS

1. Verification Dilemmas

Adjudicating an ex post legal remedy through litigation itself creates verification dilemmas which, in current practice, once again often require costly overdisclosure. Litigation verification dilemmas can affect all sorts of cases, in all stages of proceedings, to the detriment of plaintiffs, defendants, and even nonparties. The following discussion reviews and highlights these litigation verification problems, noting their particular salience in trade secret misappropriation lawsuits.

---

83. *See generally* Stephen W. Smith, *Policing Hoover's Ghost: The Privilege for Law Enforcement Techniques*, 54 AM. CRIM. L. REV. 233 (2017) (describing the development of the evidentiary privilege for police investigative techniques).

84. *See, e.g.*, United States v. Pirosko, 787 F.3d 358, 366 (6th Cir. 2015).

85. *See, e.g.*, Michael Nunez, *FBI Drops All Charges in Child Porn Case to Keep Sketchy Spying Methods Secret*, GIZMODO (Mar. 6, 2017), https://gizmodo.com/fbi-drops-all-charges-in-child-porn-case-to-keep-sketch-1793009653.

86. *See generally* Charles M. Bell, Note, *Surveillance Technology and Graymail in Domestic Criminal Prosecutions*, 16 GEO. J.L. & PUB. POL. 537 (2018).

87. Specifically, section 4 of the Classified Information Procedures Act establishes a court's authority, "upon a sufficient showing" by the prosecution, to permit the prosecution to satisfy its discovery obligations by disclosing redacted documents, disclosing summaries of documents, stipulating to facts in lieu of full disclosure, and advocating for these protections via an in camera ex parte proceeding. 18a U.S.C. § 4 (2018).

To start, sensitive information is often disclosed during pre-litigation settlement negotiations. Current legal protections for such information include contractual NDAs[88] and Federal Rule of Evidence 408's bar on admitting "conduct or statements made during compromise negotiations" into evidence.[89] Nonetheless, parties sometimes attempt to circumvent these protections and abuse the settlement information. For instance, in a recent dispute alleging that StubHub, Inc. misappropriated trade secret source code from Calendar Research, LLC, the parties disclosed confidential business information to a neutral settlement expert; settlement was not reached, and StubHub then attempted to hire the settlement expert as an adversarial expert witness in the subsequent litigation.[90]

Once litigation begins, plaintiffs must disclose sufficient information to prove their claim by a preponderance of the evidence, including elements and damages. When that disclosure involves sensitive information, plaintiffs may choose to forego litigation altogether, leading to an under-vindication of legal rights. For instance, Omri Ben-Shachar and Lisa Bernstein explain that the costs of revealing the breadth of private business information necessary to prove expectation damages in breach of contract cases can exceed the expected recovery.[91] "As a consequence," they show, "the aggrieved party may not file suit and may therefore receive no compensation."[92]

Meanwhile, defendants often must reveal sensitive information in order to disprove a claim. Doing so can impose risks and costs on entities who have done nothing wrong. If the costs of revealing sensitive information exceed the costs of settlement, defendants may choose to settle unsound or even frivolous lawsuits, leading to an over-vindication of legal rights. A common (albeit unverified)[93] anecdote in trade-secret lore about a Coca-Cola case illustrates this risk. The Coca-Cola Bottling Company sued Coca-Cola for breach of a contract that required Coca-Cola to sell them "Coca-Cola Bottler's Syrup" at a certain price.[94] The case turned on a dispute over product identity,[95] leading the judge to order Coca-Cola to reveal—under a protective order—trade

88. *See, e.g.*, Calendar Rsch. LLC v. StubHub, Inc., No. 2:17-cv-04062, 2017 WL 10378337, at *1 (C.D. Cal. Sept. 22, 2017).

89. FED. R. EVID. 408(a)(2).

90. *Calendar Research*, 2017 WL 10378337, at *1. The court ultimately disqualified the witness. *Id.* at *5.

91. Omri Ben-Shahar & Lisa Bernstein, *The Secrecy Interest in Contract Law*, 109 YALE L.J. 1885, 1888 (2000).

92. *Id.*

93. *See infra* note 97.

94. Coca-Cola Bottling Co. v. Coca-Cola Co., 107 F.R.D. 288, 290 (D. Del. 1985).

95. *Id.* at 296.

secret ingredients and data used in developing certain beverages.[96] According to Miller,-Cola settled the claim rather than comply with the disclosure order.[97] If Coca-Cola did indeed settle, there is a distinct (though admittedly unconfirmable) possibility that Coca-Cola had not breached the contract, yet suffered consequences anyway. This story illustrates the general concern that participating in court adjudication might not be worth the information disclosures it can entail. Even disclosures to a litigation adversary under a protective order entail some risks, as protective orders can, and have, been violated.[98]

Although litigation verification problems appear across legal doctrines, they are particularly salient in trade secret misappropriation lawsuits.[99] These suits involve a special "identification" procedure whereby plaintiffs must inform the defendant and the court what the defendant allegedly misappropriated.[100] Plaintiffs must undertake identification with sufficient specificity to enable the defendant to challenge both whether the information qualifies as a valid trade secret[101] and whether the defendant improperly accessed, used, or distributed it.

At the same time, the risks of disclosing trade secrets during litigation are particularly high because trade secrets must remain secret to maintain their

96. *Id.* at 300.

97. *See, e.g.*, Arthur Miller, *Confidentiality, Protective Orders, and Public Access to the Courts*, 105 HARV. L. REV. 427, 470 (1991) ("Coca-Cola settled the dispute privately and thereby relinquished its right to seek complete vindication."). Miller provides no citation for his assertion that Coca-Cola settled. *Id.* However, Miller informed one of the authors that his source for the assertion was likely a lawyer in the case. Email from Arthur R. Miller to Rebecca Wexler (Feb. 22, 2022, 7:49 AM) (on file with author). Further verification is challenging because the district court docket for the case is not available in online databases and the related appellate dockets lack any clear indication of the aforementioned settlement. *See* Coca-Cola Bottling v. Coca-Cola Co., No. 1:83-cv-00095 (D. Del. docket filed Feb. 22, 1983); Coca-Cola Bottling v. Coca-Cola Co., Nos. 91-03496, 91-03497, 91-03498 (3d Cir. July 31, 1991).

98. *See, e.g.*, Bradford Techs., Inc. v. NCV Software.com, No. C 11-04621 EDL, 2013 WL 75772, at *3 (N.D. Cal. Jan. 4, 2013); MobileMedia Ideas LLC v. Apple Inc., No. 10-258-SLR/MPT, 2012 WL 5379056, at *2 (D. Del. Oct. 31, 2012).

99. *See generally* THE SEDONA CONF., COMMENTARY ON PROTECTING TRADE SECRETS IN LITIGATION ABOUT THEM (2022); *see also* Elizabeth A. Rowe & Nyja Prior, *Procuring Algorithmic Transparency* 3-4 (Feb. 26, 2022) (unpublished manuscript) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4044178 (describing a criminal case in which prosecutors appear to have withdrawn evidence rather than reply to a defense motion that may have required disclosure of trade secrets).

100. *See, e.g.*, Xerox Corp. v. Int'l Bus. Machs. Corp., 64 F.R.D. 367 (S.D.N.Y. 1974).

101. *See, e.g.*, William M. Corrigan, Jr. & Jeffrey L. Schultz, *Trade Secret Litigation—an Updated Overview*, 63 J. MO. BAR 234, 235–37 (2007) (collecting misappropriation cases in which defendants challenged whether the plaintiff's qualified as a valid trade secret).

status as protectable intellectual property.[102] If a plaintiff over-discloses a trade secret in a public court filing and then loses their misappropriation claim, the plaintiff will have destroyed their trade secret as well as lost their case. This is not merely a theoretical concern. Litigants have repeatedly disclosed trade secrets in public court filings by accident.[103] Meanwhile, litigation disclosures will, almost by definition, take place between untrusting business competitors.[104] A bad faith competitor might not only use the information obtained during litigation; they could also leak the information publicly and thereby destroy its value as a form of intellectual property.[105] Furthermore, adjudicating whether future actions of a competitor amount to misuse of information obtained during litigation might be a complex matter in and of itself, thereby burdening even law-abiding competitors that are encumbered by such information.[106]

To add further complications, the identification requirement is susceptible to gaming.[107] If the plaintiff is permitted to proceed with the claim based on too general or conclusory a description of its trade secret, this will both impede the development of a defense and permit the plaintiff to calibrate and customize its claim based on the information it learns about the defendant in discovery.[108] As a result, plaintiffs may try to withhold information for as long as possible while obtaining discovery about the defendant's business and commercial information. The gaming risks can be compounded by the fact that, due to both practical and strategic concerns, many companies do not maintain regular written records of their trade secrets. Lack of a preexisting record enhances litigants' opportunity to craft and recraft the definition of an

102. *See* THE SEDONA CONF., *supra* note 99, at 1.

103. The Texas Supreme Court recently considered whether such oversights in initial filings waive a party's ability to correct and seal after the fact, which would leave the information permanently in the public domain. *See* Title Source, Inc. v. HouseCanary, Inc., 603 S.W.3d 829, 832 (Tex. App. 2019), *aff'd in part, rev'd in part*, 622 S.W.3d 254 (Tex. 2021).

104. *See* THE SEDONA CONF., *supra* note 99, at 2–12 (providing guidance to courts on when and how to limit discovery disclosures to various opposing party representatives in order to minimize the risks of disclosure, even under a protective order).

105. The leakiness of trade secret protection is one of its defining features. *See* Annotation, *Disclosure of Trade Secrets as Abandonment of Secrecy*, 92 A.L.R.3d 138 (1978); *see* Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 475 (1974) (observing that one of the defining features of trade secret law is its lack of protection against discovery "by independent invention, accidental disclosure, or by so-called reverse engineering").

106. Whether use of information qualifies as misappropriation of a trade secret is a fact-intensive jury question. *See generally* JAMES POOLEY, TRADE SECRETS § 6.03 (2021) (describing various types of evidence that may be relevant to proving misappropriation).

107. *See, e.g.*, The Sedona Conference, *Commentary on the Proper Identification of Asserted Trade Secrets in Misappropriation Cases*, 22 SEDONA CONF. J. 223, 253 (2021).

108. *See, e.g.*, POOLEY, *supra* note 106, § 11.02(2)(c).

alleged trade secret as the case evolves. In contrast, defendants will want to nail down the plaintiff's claim with specificity early on so as to prevent claim morphing, narrow their own discovery obligations, and—less legitimately—to force the plaintiff to undertake maximum disclosure risks, even with information that may ultimately be irrelevant to resolving the dispute.[109]

Finally, criminal prosecutions for trade secret misappropriation raise special, challenging tensions. The arguments for detailed disclosure are stronger because criminal defendants have unique constitutional entitlements to access relevant evidence and have their case proceed in public.[110] Yet the risk of destruction of trade secrets through judicial publication[111] or leakage falls on the alleged victim of misappropriation—a nonparty who did not elect to initiate the case and lacks control over the government's litigation decisions.[112]

2. Imperfect Legal Responses

Existing legal solutions to these litigation verification problems are, once again, imperfect. Courts often rely on protective orders, limited discovery, sealing orders, and courtroom closures to limit the redistribution of sensitive information shared in litigation. But protective orders are vulnerable to leaks, whether through mistakes, negligence, or malicious intent. They also introduce new problems. For instance, protective orders can impede effective representation by limiting what attorneys can communicate to their clients or by unduly restricting access to expert witnesses. In the criminal context, protective orders can interfere with prosecutors' *Brady* due process disclosure obligations.[113] Limited discovery risks impeding the judicial truth-seeking process of adjudication and obstructing the regulatory goals that private discovery can serve.[114] And judicial sealing orders and courtroom closures can

109. *Id.*

110. *See* Kenneth Rosenblatt, *Criminal Law and the Information Age: Protecting Trade Secrets from Disclosure in Criminal Cases*, 8 COMPUT. L. 15, 15 (1991).

111. For a discussion of judicial authority to compel even public disclosures of trade secret information that is "indispensable" for the adjudication of the case without triggering constitutional takings claim, see Kapczynski, *supra* note 78, at 1437 & n.308.

112. Of course, prosecutors in these cases often try to accommodate the victims' interests, in part because they want or need victim cooperation to successfully bring the charge. *See, e.g.*, Brian L. Levine & Timothy C. Flowers, *Your Secrets Are Safe with Us: How Prosecutors Protect Trade Secrets During Investigation and Prosecution*, 38 AM. J. TRIAL ADVOC. 461, 464 (2015).

113. *See* Jonathan Abel, *Brady's Blind Spot*, 67 STAN. L. REV. 743 (2015).

114. *See* Diego Zambrano, *Discovery as Regulation*, 119 MICH. L. REV. 71 (2020).

raise First Amendment and broad democratic governance concerns about public access to court records.[115]

***

Part II has introduced verification dilemmas in law—that is, the problem that verifying facts about information often requires undertaking risky disclosures. It has shown that the problem is a recurring and transdoctrinal issue in law, appearing in the information privacy and security, dealmaking, government oversight, and trade secret litigation contexts. Each of the examples discussed above also illustrates limitations in current legal solutions to the verification problem in law. Ex post remedies for post-disclosure misuse of information are available solely after misuse has been detected and can be prohibitively costly to pursue. Moreover, the very process of pursuing such a remedy through litigation can create new verification problems that in turn require new, costly disclosures of sensitive information in court proceedings. And in certain circumstances, such as those associated with anti-circumvention rationales for secrecy, the law may err on the side of *not* permitting entities to disclose information, thereby sacrificing some quantum of trust and verifiability.

Having laid out the verification problem in law and some limitations of the current legal solutions to it, the Article now turns to new technological developments that may offer a radical alternative approach to help solve these recurring legal issues.

III. INTRODUCING ZERO-KNOWLEDGE PROOFS

A. THE IDEA OF ZERO-KNOWLEDGE PROOFS (OR, *THE TALE OF THE MATHEMATICIAN'S FRIEND*)

First presented in 1985, zero-knowledge proofs (ZKPs) enable one party (the "prover") to prove to another party (the "verifier") assertions regarding properties of secret information known only to the prover, without revealing an secret information. ZKPs address the insight that often we are unnecessarily exposed to an entire corpus of data for the sole purpose of verifying limited properties of the corpus. For example, when a police officer stops a driver on the road to verify that they hold a valid driver's license, the officer need not learn the driver's date of birth or other private information that is disclosed

115. For a discussion of the public harms of suppressing public access to purported trade secrets in both the regulatory state and judicial proceedings, see Kapczynski, *supra* note 78, at 1428-41. For a discussion of First Amendment rights of access to judicial documents, see generally Hannah Bloch-Wehba, *Exposing Secret Searches: A First Amendment Right of Access to Electronic Surveillance Orders*, 93 WASH. L. REV. 145 (2018).

when presenting a driver's license. Using a ZKP, the driver could prove to the police (the verifier in this case) that her license (which would contain encrypted identifying information such as age) is valid (which implies a valid driving age) *without revealing an exact age or any other identifying information.* In a more complex example, an employer can, without disclosing specific salaries, prove to its employees that the salaries of its employees (stored, say, in an encrypted salary database) are equitable with respect to gender. In an even more complex example, a government agency (acting as a prover) can prove to the public that a certain sensitive forensic (or, data-gathering) algorithm operates as claimed without disclosing the algorithm itself.

A bit more abstractly, a pair of algorithms (one for the prover and one for the verifier) qualifies as a ZKP for a given public assertion regarding some hidden data that is known only to the prover if the following three properties hold.[116] The first is *completeness*: if the assertion (about the hidden data) is true, the verifier's algorithm, after interacting with the prover's algorithm, will accept the assertion's validity. The second is *soundness*: if the claimed assertion is false, the verifier will reject the assertion. Crucially, this holds even if the prover tries to cheat and does not follow its prescribed algorithm. (Technically, the situation where a verifier will accept a false claim can happen. However, the probability that this is the case depends only on random choices made by the verifier, and can be set to be sufficiently small so as to be both mathematically and realistically insignificant.) This probabilistic aspect is necessary to enable the third characteristic, *zero-knowledge*, which is where the novelty lies: the verifier will learn no new information about the undisclosed data besides the validity of the assertion regarding the data. Again, this holds even if the verifier tries to cheat and does not follow its prescribed algorithm. Mathematically, this is formulated as the requirement that for any given verifier, the verifier herself could generate the probability distribution over the information seen when engaging in a ZKP. This means that the verifier could have obtained the same information just by knowing only that the statement is true, and without ever interacting with the prover. In other words, nothing new is gained beyond the validity of the assertion.

We emphasize that this last characteristic is what makes ZKPs unique: rather than demonstrating the correctness of the assertion via making the data public, ZKPs enable convincing a distrustful verifier without exposing anything about the hidden data—other than the very fact that the assertion is correct.

116. *See* Goldwasser et al., *supra* note 4, at 293, 295.

To show how this is done, it is helpful to illustrate a basic example of the mathematical principle underlying ZKPs before providing more detail on how the ZKPs can be implemented. The following beautiful illustration of the concept of an indirect mathematical proof, taken from Ron Aharoni's book, *Mathematics, Poetry and Beauty*, supplies a rudimentary illustration:[117]

> A mathematician and his friend are walking in the forest. The friend boasts: "In a flash, I can tell how many needles are on this pine tree." "How many?" the mathematician asks. "143,547" says the friend, without batting an eyelash. The mathematician takes a handful of needles and asks: "And how many now?"[118]

All the mathematician (the "verifier" here) needs to do is to calculate by a simple subtraction that the number of needles in her hand equals the difference between her friend's two answers. If the calculation succeeds, then the mathematician will accept her friend's boast, otherwise she will reject it. This story demonstrates how it is possible to verify statements about having information (or, in this case, computational ability to verify the ability to compute the number of pine needles), without having full access to the information (or computational ability). In this story, the *completeness* property is satisfied because the mathematician will accept if her friend does have the magical ability to count pine needles. *Soundness* is satisfied as well: if the friend cannot truthfully count the number of needles on a pine tree, it is extremely unlikely that he will be able to guess the exact number of needles the mathematician holds in her hand. If this happens, the calculation will fail, and the mathematician will reject her friend's boastful assertion. The exact probability of rejection is one of the possible choices for the number of needles that the mathematician holds. Thus, the verifying mathematician controls the probability of rejection, regardless of the prover's strategy.

But what about zero-knowledge? In Aharoni's story, if the friend is telling the truth then the mathematician learns the number of pine needles on the tree! Can the skeptical mathematician be convinced of her friend's magic powers without learning the number of needles on the pine tree? Yes, but we need to slightly augment the story: Instead of asking the friend to report the number of needles before and after picking the random handful of needles, the mathematician will only ask the friend to guess the number of needles hidden in the mathematician's hand. The friend (who can count needles on trees) will then simply report the difference between the number of needles on the pine tree before and after the mathematician picked the random handful

117. RON AHARONI, MATHEMATICS, POETRY AND BEAUTY (2015).
118. *Id.* at 78.

of needles. The mathematician will then count the number of needles in her hand and accept the number if and only if it equals the friend's guess.

Now, all three properties are satisfied: completeness, soundness, and zero-knowledge. If the friend is a truth teller, the mathematician will verify that the number of needles in her hand equals the friend's guess. If the friend is a liar, the guess is likely to be wrong, just as before. Importantly, all the mathematician learns is the number of needles in her hand she picked herself. This is a random number which can be picked without ever interacting with the prover. Thus, the mathematician learns nothing from the interaction with the friend, except for being convinced in the friend's ability to count pine needles on trees.[119]

B. THE GENERAL APPLICABILITY OF ZERO-KNOWLEDGE (WITH SOME MATH FOR GOOD MEASURE)

The pine-needle story might seem irrelevant for real-life applications. However, the ideas that underlie it, namely the use of randomness to challenge the proving party as a replacement for asking for more details about the proof, turn out to be extremely powerful and generalizable. In particular, these ideas are readily applicable to digital information, which is where ZKPs are most powerful.[120]

ZKPs' wide applicability to digital information stems from the following deeply insightful observation: it is possible to transform any conventional mathematical proof for an assertion regarding digital information into a ZKP of the same assertion.[121] In fact, many such transformations have been devised over the past three decades.[122]

---

119. We thank Tal Canetti for proposing the current, simplified version of the zero-knowledge proof of the ability to count pine needles on trees.

120. It should be stressed, however, that the applicability of zero-knowledge proofs transcends the digital domain. In particular, ZKPs for physical properties have been proposed in a number of settings and for multiple purposes, from educational and recreational to international relations. *See, e.g.*, Ben Fisch, Daniel Freund & Moni Naor, *Physical Zero-Knowledge Proofs of Physical Properties*, 2014 ADVANCES IN CRYPTOLOGY – CRYPTO: PART II 313 (Juan A. Garay & Rosario Gennaro eds.); Glaser, *supra* note 11; Ronen Gradwohl, Moni Naor, Benny Pinkas & Guy N. Rothblum, *Cryptographic and Physical Zero-Knowledge Proof Systems for Solutions of Sudoku Puzzles*, 4475 FUN WITH ALGORITHMS, LECTURE NOTES COMPUTER SCI. 166 (2007).

121. The first such general transformation was devised in 1986 in two works: Oded Goldreich, Silvio Micali & Avi Wigderson, *All Languages in NP Have Zero-Knowledge Proof Systems*, 38 J. ACM 691 (1991) and Gilles Brassard, David Chaum & Claude Crépeau, *Minimum Disclosure Proofs of Knowledge*, 37 J. COMP. SYST. SCI. 156 (1988).

122. *See, e.g.*, ODED GOLDREICH, FOUNDATIONS OF CRYPTOGRAPHY 184 (2001) ("The main result presented in this chapter is a method for constructing zero-knowledge proof systems for every language in NP . . . . Specifically, almost all statements one may wish to prove in practice can be encoded as claims concerning membership in languages in NP."); See

The rest of this section provides a high-level overview of how such transformations work (with more details provided in the Appendix). It is stressed that the legal analysis in this paper holds regardless of the particular transformation used. Furthermore, understanding how these transformations work is not needed to evaluate and build on the legal analysis. The goal of this overview is to demystify ZKPs for a legal audience and provide a (largely) nonmathematical understanding of how such transformations work.

A preliminary step towards transforming a conventional mathematical proof into a ZKP is to view the conventional process of verifying a mathematical proof as a computer program that takes the text of the proof as input, outputs "1" if the verification succeeds, and "0" otherwise. (For instance, if the assertion is "the number 77 is a product of two prime numbers," then the proof-text would consist of two numbers, and the verification program would first check that the two input numbers are primes, and then multiply the two numbers and check that the result is 77. Finally, the verification program will output 1 if both checks succeed, and 0 otherwise. Alternatively, if the assertion is, "There exists a value *W* such that the plaintext obtained by decrypting the ciphertext 12345678 using AES with key *W* is a number between 18 and 120," the proof-text would consist of a value *W*, and the verification program will first decrypt 012345678 using AES with key *W*. It will then output 1 if the result is a number between 18 and 120 and 0 otherwise.)

Viewed this way, a ZKP's goal is to enable the prover to convince the verifier that it holds a proof-text *W* such that *if the verification program were to be run on W then the output would be "1*." Furthermore, the prover should be able to do so without disclosing the proof-text itself. It is stressed that the verification program is public and known to all. Only the proof-text (namely, the numbers 7 and 11 in the first example above, or the key *W* in the second example) is to remain hidden.

As the above examples suggest, verification programs can express a broad range of properties that a hidden data set might or might not have. The Article now turns attention to demonstrating how to design a zero-knowledge proof for a given verification program. Specifically, two alternative (and very different) methods for designing zero-knowledge proofs for *any given verification program* are described. (It is noted that ether one of these two methods would suffice for any of the applications mentioned in this work. Presenting both will

also the ZKProof Resources Repository for an extensive list of currently available ZKP systems: https://docs.zkproof.org/schemes/

hopefully help the reader separate the concept of zero-knowledge proofs from a particular algorithmic way to realize the concept.)

1. Two methods for realizing ZKPs

a) Method 1 (The Boxes)

The idea underlying this method is to transform the underlying verification program into another computer program that can perform the same computational steps (i.e., the same sequence of manipulations of the proof-text) even when the proof-text is given only in a "veiled" way.[123]

The following two concepts are helpful to understanding how this transformation works. First, any computer program (including the verification program at hand) can be written as a sequence of very simple basic steps, where each basic step consists of choosing two pieces of data from either the input or the memory, computing a simple function of the two pieces, and writing the result back in memory. Such functions are called *complete*. For instance, a piece of data can be as small as one binary value, namely either 0 or 1, and the complete function can be the NAND logical gate.[124] The first step is thus to transform the verification program to such a format.

Second is the concept of a cryptographic *commitment* to data. Metaphorically, a cryptographic commitment is the digital analog of a lockable box: *Committing* to a piece of data (say, the proof input) is tantamount to writing this piece of data on paper, putting the paper inside the box, locking the box with a key, and handing the locked box to the recipient of the commitment. *Opening* the commitment is tantamount to handing the key to the recipient of the commitment, thus allowing the recipient to open the box and read the data. The salient properties here are: (a) the committer is guaranteed that until such time that she hands the key to the recipient, the data remains completely hidden, and (b) once the recipient obtains the box it is guaranteed that the data inside it are immutable (even though she might not yet know what they are).

The commitments used in this method have an additional *homomorphism* property. *Homomorphic* commitments are commitments which allow for the following "magic" to happen: Assume the committer hands two boxes to the recipient: box c_1 contains the value m_1, and box c_2 contains the value m_2. The committer keeps the corresponding keys r_1, r_2. Homomorphic

123. This methodology roughly follows the approach in Gilles Brassard & Claude Crépeau, *Zero-Knowledge Simulation of Boolean Circuits*, 1986 ADVANCES IN CRYPTOLOGY – CRYPTO '86 PROC. 223 (A.M. Odlyzko ed.).

124. *See, e.g.*, Even, G., & Medina, M. (2012). Propositional Logic. In Digital Logic Design: A Rigorous Approach (pp. 68-93). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139226455.007.

commitments allow the receiver to "mashup" box c_1 and box c_2 and obtain a single box (let us call it c_3) that contains the number $m_1 * m_2$, where $*$ is some agreed upon function *and nothing else*. Furthermore, the committer can now "mashup" the keys r_1 and r_2 to obtain a key r_3 that can be used to open the new box c3—but neither c1 nor c2. The function * to be used here is some complete one, such as the NAND operation mentioned above.

Armed with the concepts of complete functions and homomorphic commitments, one can now turn to creating the ZKP itself. The prover (given a public verification program and a secret proof-input) and the verifier (given only the verification program) proceed as follows:

1. The prover commits (using homomorphic commitment) to the proof-input by splitting the proof-input into small pieces, putting each piece in a box as described above and sending all the boxes to the verifier.

2. The verifier runs the verification program on the proof homomorphically (namely, "in boxes"). Recall that the verification program is now only a sequence of applications of the function * to the values in specific input (or memory) locations. These applications of * are now realized by mashing up the corresponding boxes—either boxes obtained from the prover, or previously mashed-up boxes—until the verifier obtains a box c^* that corresponds to the output value of the verification program. At this point, the verifier knows that the value inside box c^* is the result of the verification process.

3. The prover mashes up the keys for the boxes it sent to the verifier, in the same way that the verifier mashes up the boxes. Finally, the prover obtains the key r^* that can enable opening the box c^* and sends r^* to the verifier.

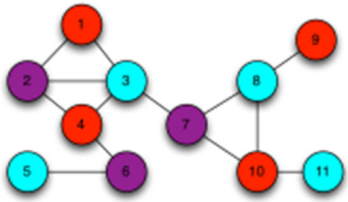
4. The verifier opens the box c^* using the key r^* obtained from the prover and accepts if the value in the box is 1.

Completeness and soundness of this proof follow from the correctness of the homomorphic commitment scheme. Zero knowledge follows from the fact that all that the verifier sees is a collection of identical-looking opaque "boxes," where only one box is opened. Furthermore, as long as the prover follows its prescribed algorithm, the value in the opened box is always 1. The verifier can sample this same information without ever interacting with the prover. Appendix A holds a more mathematical description of this approach.

b) Method 2: (Graph Coloring)

This method (invented in the landmark work of Goldreich et al.[125]) uses the theory of NP-completeness,[126] which demonstrates that some families of combinatorial objects (e.g., graphs) have the following remarkable property: it is possible to represent the execution of any given program on a given input by way of some combinatorial object (e.g., a graph) from the family in such a way that the combinatorial object (the graph) has a certain combinatorial characteristic *if and only if* the given program, running on the given input, outputs "1."

The combinatorial object of choice in Goldreich et al. is a graph, and the characteristic is that the graph has a valid "3-coloring": each vertex is assigned one of three possible colors such that no edge has its two endpoints assigned the same color. The following structure ensues: any proof verification program can be translated into a graph known to both prover and verifier, and any purported proof-text, known to the prover, can be translated into a coloring of the nodes of the graph, such that the coloring is a valid 3-coloring if and only if the verification program, given the proof-text as input, outputs "1." This means that proving that the verification program running on the input-proof outputs "1" is now equivalent to proving that the coloring held by the prover is valid.



A valid 3-color graph. Each node is colored in either blue, red, or purple, and no two nodes connected by an edge are the same color.[127]

Proving that a graph is 3-colorable is done as follows: The prover starts by randomly renaming the three colors—the prover chooses a random

125. *See* Goldreich et al., *supra* note 121.

126. *See* Stephen A. Cook, *The Complexity of Theorem-Proving Procedures*, 3 PROC. ACM SYMP. ON THEORY OF COMPUTING 151 (1971); Leonid Levin, *Универсальные задачи перебора* ["Universal Search Problems"], 9 PROBS. INFO. TRANSM'N 115 (1973).

127. Matthew Green, *Zero Knowledge Proofs: An Illustrated Primer*, A Few Thoughts on Cryptographic Engineering (Nov. 27, 2014), *available at* https://blog.cryptographyengineering.com/2014/11/27/zero-knowledge-proofs-illustrated-primer/.

permutation of {blue, red, purple} and reassigns the colors accordingly. Next, the prover sends a series of commitments, each to the color of a different vertex in the graph. The verifier then chooses a random edge among all edges in the graph and requests the prover to open the commitments corresponding to the colors of the two endpoints of the chosen edge. If the open commitments indicate that the two endpoints are colored with the same color, the verifier rejects the ZKP. If the open commitments indicate that the two endpoints were colored with two different colors, the process is repeated. After some number of successful iterations in which the verifier did not reject, the verifier accepts the ZKP. Evidently, the probability that the verifier will accept when the graph is not 3-colorable exponentially vanishes with the number of iterations. (In particular, 200n*log(n) iterations, where *n* is the number of edges in the graph, will be plenty.) Furthermore, the protocol is zero-knowledge: if the coloring is valid and the prover follows the instructions of the protocol then, in each iteration, all that the verifier sees is that the two endpoints of a random edge are colored in two different random colors; this is information that the verifier could have generated on her own, without any interaction with the prover.

Regardless of which of the two methods is used, any verification program can be turned into a ZKP. In other words, for *all* mathematical assertions for which there exists a traditional proof, there exists a protocol where one party who knows a traditional proof can convince another party that the assertion is valid, revealing nothing else.

2. Specialized Assertions and Constructions

Although the above two methods are very general, they often require an excessive amount of computational resources. Luckily, there are many constructions of ZKPs in the literature that are tailor-made for specialized assertions and require significantly less computational resources. Examples include the equality (or inequality) of two encrypted documents, numerical assertions about encrypted numerical data sets such as the value of the average, verifying membership in a list, or verifying the results of search functions.[128] Two questions that must be asked in the context of the present paper are (1) how to design assertions that would be of value in legal contexts such as the ones discussed in this work and potentially others, and (2) how efficient can ZKPs be made for such assertions. In Part IV and the appendices, these two

128. *See* ZKPROOF, ZKPROOF COMMUNITY REFERENCE 29, 75 (D. Benarroch, L.T.A.N. Brandão, E. Tromer eds., 2019), https://docs.zkproof.org/pages/reference/reference.pdf.

questions are discussed at length in the context of the verification dilemmas presented in Part II.

3. Noninteractive Zero-Knowledge

Originally, ZKPs were conceived as "interactive" algorithms, in the sense that the verifier interacts directly with the prover and chooses random challenges in the course of the interaction. Furthermore, they were nontransferable: the verifier had no way to convince third parties that did not witness the interaction of the prover's assertion verity. (Recall the example of the mathematician and his friend: it was crucial for the mathematician to choose the number of pine-needles to remove from the tree at random, and furthermore do so only *after* the friend announced the original count.)

It is often beneficial to have ZKPs where the interaction is limited to having the prover send to the verifier only a single message, where the verification process requires the verifier to make no secret random choices. Such ZKPs are called "noninteractive" zero-knowledge proof (NIZK) systems.[129] The great advantage of NIZKs over interactive ZKPs is that the prover and the verifier do not need to interact directly with each other. In fact, the proof need not be associated with a specific verifier: the prover can just post the proof once and for all to be verified by anyone at any time (much like a standard written proof). This *public verifiability* property is particularly useful in a setting where a single entity (say, a government agency) wishes to make public assertions about hidden data, while avoiding the need to separately convince each member of the public. In Part IV we discuss specific legal settings where this property may be useful.

Achieving non-interactivity comes at a price. In order to make the mechanism work, the prover and the potential verifiers need to *a priori* agree on a value, called the common reference string (CRS) that they both trust to have been randomly sampled. (The CRS may have been randomly sampled by some trusted physical process, some trusted authority, or the parties themselves in a preliminary interactive stage. Either way, a CRS can be reused indefinitely for as many proofs as needed.)

Interestingly, the first method for constructing general zero-knowledge proofs presented above ("The Boxes") can be transformed in a straightforward fashion to a NIZK. In fact, the above intuitive explanation is already a NIZK, since the communication between prover and verifier consists of a single message from the prover to the verifier containing the commitments to the

129. Manuel Blum, Alfredo De Santis, Silvio Micali, and Giuseppe Persiano: *Noninteractive Zero-Knowledge*, 20 SOC'Y FOR INDUS. & APPLIED MATHEMATICS 1084, at 1091 (1991) [hereinafter Blum et al., *Noninteractive Zero-Knowledge*].

proof-text and the key for the final mashed-up box. (The CRS in this case contains the initial random values that are required to set up the mathematical representation of the boxes.) Transforming proofs that use the Graph Coloring method to a NIZK is a bit more involved. One intuitive idea here is to encode the verifier's random challenges in the CRS, so that they need not be sent by the verifier "in real time."

C. THE CASE OF SPLIT SECRETS: ZKPS AND MULTI-PARTY COMPUTATION

This subpart describes a related and important technique, *Secure Multiparty Computation* (MPC), which contributes to ZKPs broader applicability to legal verification dilemmas.

In their basic form, ZKPs are designed for a setting where one party (the prover) has some secret information and wishes to convince one or more other parties of the correctness of some assertion pertaining to the secret, while keeping the secret otherwise hidden. However, there exist situations where the secret information is split into two or more pieces, where each piece is exclusively known by different parties. For instance, consider a trade secret litigation where the court wishes to verify the plaintiff's assertion that the trade secret formula used the defendant used is the same as the one the plaintiff holds. Furthermore, the parties want to do so without having any of the trade secrets disclosed to either the court or to the other party. One may be tempted to let the proof-text be the two formulas and apply one of the above methodologies to obtain a ZKP. But this straightforward attempt would fail since there is no single party who knows the entire proof-text. Instead, such contexts require a generalization of ZKPs called secure multiparty computation (MPC).

MPC is a class of cryptographic techniques that address the following type of setting. Consider two (or more) mutually distrustful parties, each holding a piece of sensitive information. The parties wish to jointly compute some agreed-upon function of all their secrets put together. Figuratively, the parties wish to emulate a situation where each one hands their secret to an imaginary trusted party who computes the agreed-upon function, informs each party of their function value, and then disappears. Importantly, the protections should hold even when some or all of the other parties deviate from the protocol instructions. This includes guaranteeing preserving the secrets of the parties that follow the protocol instructions as well as guaranteeing that the output values these parties obtain are computed according to the agreed-upon function as applied to the parties' secret data. Furthermore, each party that follows the protocol is guaranteed that the secrets contributed by other parties

are well defined and do not depend on the party's own contributed secret. This holds even if the other parties do not follow the protocol.

In the context of the above trade secret litigation example, the parties to the MPC computation would be the plaintiff, defendant, and court. The function to be evaluated would take a description of trade secret *A* from the plaintiff, a description of trade secret *B* from the defendant, and a description of a test algorithm *T* from the court. The function will then run algorithm *T* on secrets *A* and *B* and announce the result to the court. (The test algorithm *T* will apply a set of agreed-upon tests and subsequently output a value indicating whether the two input trade secrets are "close enough.") More generally, in the language of ZKPs, here the two litigants act as provers and the court acts as a verifier. Alternative configurations are of course possible as well.

In a different example, each party's secret is their client database, and the court wishes to learn whether the two lists are sufficiently similar (or if one list is contained in the other). Alternatively, each party's secret is a computer program, and the court wishes to learn whether the respective programs are similar, according to some agreed-upon measure of similarity.

To realize an MPC computational task, the participants are provided an *MPC protocol*, namely a set of instructions to be followed by each participant in the joint computation. These instructions enable each participant to process its local data and information received from other parties, so as to send new information to others, and eventually determine the desired outcome of the computation. MPC protocols exist for securely evaluating any desired function of the secret values held by the parties.[130]

A general and ubiquitous paradigm in constructing MPC protocols consists of two main conceptual steps: (1) construct MPC protocols whose security guarantees hold only as long as all parties adhere to their protocol instructions; and (2) have the parties run the protocol from the previous step, and, in addition, have each party prove to the other parties that its messages were computed correctly. That is, each message of the protocol from the first step will be accompanied by a ZKP that the message was computed correctly

130. The area has been extensively studied in the past three decades, since the first groundbreaking works of Andrew C. Yao, *Protocols for Secure Computations*, FOUND. COMPUT. SCI. 160 (1982); Oded Goldreich, Silvio Micali & Avi Wigderson, *How to Play Any Mental Game*, 19 PROC. ACM SYMPOSIUM ON THEORY OF COMPUTING 218 (1987); Michael Ben-Or, Shafi Goldwasser, Joe Kilian & Avi Wigderson, *Multi-Prover Interactive Proofs: How to Remove Intractability Assumptions*, 20 PROC. ACM SYMPOSIUM ON THEORY OF COMPUTING 113 (1988).

given the messages received so far, the local (hidden) randomness, and initial hidden input.[131]

Recalling the terminology introduced earlier, a traditional ZKP with some verification program *V* and proof-input *pf* can be cast as an MPC protocol for two parties— prover and verifier— where the prover's input is *a purported proof-text,* the verifier has no input, and the agreed function is the proof verification program (applied to the prover's input). The MPC's guarantees imply that the verifier learns the (0 or 1) output of the proof verification program and nothing more. In the same vein, a ZKP for the case where the purported proof-text has several components, each held by a different party, is an MPC protocol for the proof verification function applied to all the components of the proof-text put together.

Part III has provided a detailed overview of the idea of zero-knowledge proofs and how ZKPs work. The following Part will develop examples of how this technique can help resolve legal verification dilemmas across multiple legal contexts.

IV. ZERO-KNOWLEDGE PROOFS APPLIED TO VERIFICATION DILEMMAS

The newfound ability (supported by ZKPs) to verify assertions about information without learning the underlying information itself, can potentially upend existing understanding and common legal practice regarding the balance between disclosure and secrecy and it can lead to fresh understanding and new forms of balance. This Part explores the ways that ZKPs might be used in the four doctrinal contexts discussed in Part II and offers specific-use cases demonstrating how they ZKPs help to resolve recurrent verification dilemmas.

Our prototypical workflow for using ZKPs for verification dilemmas consists of three steps:

> (a)The prover and the verifier agree on a digital *base document* that uniquely identifies the corpus of data under consideration. The information in this base document must preserve the secrecy of the data under consideration (much like the concept of a mathematical commitment discussed in Part III). Simultaneously, the information

131. MPC protocols that guarantee security only when all parties adhere to the protocol instructions are often called protocols for the "honest-but-curious model." When the parties trust each other to adhere to the protocol (say, due to contractual agreements enforced by ex-post legal remedies), it suffices to run protocols for the honest-but-curious model, without additional protections. However, such protocols are not sufficient for the applications considered here. *See also* Yehuda Lindell, *Secure Multiparty Computation*, COMMC'NS OF THE ACM VOL 64 NO 1, PAGES 86-96, 89 (2021).

must provide assure the verifier that the document uniquely and unequivocally pins down the data that pertains to the case at hand. It is stressed that this guarantee is twofold: First, the base document must uniquely pin down an entire dataset to be considered (while still keeping this dataset unknown to the verifier). Second, the verifier must be provided guarantees that this unknown dataset is the one that pertains to the case at hand. While the first guarantee (uniqueness) can be provided via purely mathematical means (such as encryption, one-way hash, or cryptographic commitment), the second guarantee would typically be obtained via social or legal means (such as third-party attestation, existing records, audit, or facing punishment for perjury).

(b) Once this base document is in place, the prover and verifier agree on the set of properties (of the hidden dataset) of which the prover will convince the verifier. These properties would typically take the form of a set of checks, or, more concretely, a computer program *C* that reads the dataset and outputs "ok" if all checks passed.

(c) Now, the prover and verifier each run their ZKP programs. (These programs, ZKP-Prover and ZKP-Verifier, respectively, would typically be fixed and known ahead of time.) The verifier's program, ZKP-Verifier, is provided the base document and the check-program *C*. The prover's program is provided the same base document and check-program, plus the hidden data. The two programs exchange messages until the verifier's program outputs its decision. (In the case of noninteractive zero knowledge, a single message from the prover's program to the verifier's program suffices. In general, however, more messages might be needed.) If its program accepts, then the verifier is assured that the hidden data the base document uniquely determined passes all the agreed-upon checks.

It is important to again stress that a ZKP does not provide, in and of itself, any guarantee regarding whether the hidden data identified by the base document pertains to the actual case. Such guarantee must be provided in case-specific ways.

A. INFORMATION PRIVACY AND SECURITY: USING ZKPS TO SEVER VERIFICATION FROM IDENTIFICATION

ZKPs can help reduce the need for existing, imperfect legal solutions to verification dilemmas. Recall the recurring verification problem whereby the process of verifying eligibility to access digital systems forces eligible access-seekers to expose facts about themselves, thereby risking broader aggregations of data and the creation of a permanent "digital person," and exacerbating the

risk of theft, misuse, or fraud. As described in Section II.A, existing legal solutions to this problem have had little success in reflecting the insight that, "[i]f your personal data is never collected, it cannot be stolen."[132]

ZKPs can help to solve those verification problems that concern private or sensitive information by enabling individuals to verify the specific information required for authentication without disclosing their identity. Perhaps the most straightforward application of ZKPs—and the context in which its applications have advanced furthest—involve what is often referred to in the cryptography literature as "anonymous credentials."[133] By providing anonymous credentials, ZKPs enable digital "gatekeepers" that keep no private information about users and still recognize when an access-seeker is eligible for accessing the sought service.

More specifically, ZKPs allow users to keep control of their private information (e.g., birthdate, credit card and other financial information, passwords for access to various systems or services), storing them exclusively on their own computing devices or private spaces. When seeking to access a digital service, the user's computing device U will interact with the gatekeeping device G of the service, providing it with a ZKP that the user whose private information is recorded on the device U is one of the users allowed to access the service. Here the zero-knowledge property guarantees that neither G nor anyone else learns anything from the interaction with U other than the mere fact that U can access the service.

For example, ZKPs can permit the verification that an individual falls into a qualified "range," such as the age required to stream an R-rated film or order alcohol.[134] ZKPs can also be used to prove "set membership"—such as citizenship—in digital interactions without disclosing identity.[135] Stable noninteractive proofs of different personal attributes can be used to prove the statement that one is an EU citizen, for example, without even revealing one's member state. The European Commission's EU Blockchain Observatory &

132. *See* Dubovitskaya, *supra* note 15.

133. The concept of anonymous credentials has been studied extensively over the past two decades. *See* Jan Camenisch & Anna Lysyanskaya, *An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation*, 2001 ADVANCES IN CRYPTOLOGY – EUROCRYPT 93 (Birgit Pfitzmann ed.); Non-Transferable Anonymous Credentials, U.S. Patent No. 7,222,362 (issued May 22, 2007).

134. *The Sovrin Network and Zero Knowledge Proofs*, SOVRIN (Oct. 3, 2018), https://sovrin.org/the-sovrin-network-and-zero-knowledge-proofs/ ("It's as if you're creating a carbon copy of your driver's license that is every bit as reliable, and conveys the same personal identifiable information, as the real thing; but, based on who is asking, you control what information actually appears to them on that particular copy.").

135. *See id.* ("ZKPs can prove if a value is contained in a set without revealing with [sic] value.").

Forum heralded the promise of these approaches in promoting GDPR compliance, concluding that "ZKP applications hold great promise when it comes to privacy-by-design and self-sovereign ownership of personal data."[136] Some governments have already begun to develop public-based digital identification systems enabling "self-sovereign" selective disclosures of information.[137] Estonia has progressed farthest with a partially blockchain-based national identity system. This system facilitates managed information disclosure, enabling EU travel, national health benefits, bank account access, and medical-record administration.[138] A public-private partnership in the Dutch province of Groningen has implemented a digitized social service provision system that allows parents to receive funding for children who require financial aid. The system employs ZKP mechanisms to limit the exchange of raw personal data and permit the use of cryptocurrency as an additional privacy-protecting mechanism.[139]

ZKP-enabled data comparisons are useful also for preserving privacy in biometric applications by freeing gatekeeper servers from having to store users' private biometric features, such as fingerprints, iris scans, or face prints.[140] Instead, it is only the user-controlled device that keeps the users' private biometrics, and the gatekeeper only needs to store public, nonidentifying information.[141]

The use of ZKPs for verification of eligibility has another important advantage: it allows controlling the extent to which gatekeepers are able to link between different access attempts by the same user (or related users).[142] When

136. THE EUROPEAN UNION BLOCKCHAIN OBSERVATORY & F., BLOCKCHAIN AND THE GDPR 23 (2018), https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf.

137. ANDREJ J. ZWITTER, OSKAR J. GSTREIN & EVAN YAP, DIGITAL IDENTITY AND THE BLOCKCHAIN: UNIVERSAL IDENTITY MANAGEMENT AND THE CONCEPT OF THE "SELF-SOVEREIGN" INDIVIDUAL 8–10 (2020), https://www.frontiersin.org/articles/10.3389/fbloc.2020.00026/full.

138. *e-identity*, E-ESTONIA, https://e-estonia.com/solutions/e-identity/id-card; *see* ZWITTER ET AL., *supra* note 137, at 8.

139. ZWITTER ET AL., *supra* note 137, at 9 (citing Pim Van der Beek, *Blockchain Kindpakket Zuidhorn Wint Prijs*, COMPUTABLE (Mar. 30, 2018), https://www.computable.nl/artikel/nieuws/digital-transformation/6329958/250449/blockchain-kindpakket-zuidhorn-wint-prijs.html).

140. *Cf.* Andrea Roth, *Spit and Acquit: Prosecutors as Surveillance Entrepreneurs*, 107 CALIF. L. REV. 405, 407–08 (2019) (describing large database of DNA collected through deals of prosecutorial leniency).

141. It is noted that ZKPs are typically applied to the digital rendering of the biometrics, rather than to the biometrics themselves. Still there do exist ZKPs that are applied directly to biometric data. *See, e.g.*, Fisch et al., *supra* note 120, at 314.

142. *See infra* note 160.

optimizing for privacy, systems can be set so that gatekeepers will not be able to link between access attempts by any sets of users. Alternatively, systems can be designed so as to enable gate keepers to collect agreed-upon statistics on the characteristics of users attempting access. For example, with a ZKP, sign-in systems that use government-provided IDs can be designed to allow for the gatekeeper to collect the states of all participants but verify the ID's validity without collecting a name; all that is required is agreeing on this schema ahead of time.

This Section has detailed use cases for ZKPs to protect privacy by helping to solve verification dilemmas in law without requiring overdisclosure. As the following Sections will show, in other instances, the ability to shield attributes of personal information in big data sets can further information protection in both the private transactional (Section IV.B.), and public governance (Section IV.C.) contexts.

B. DEALMAKING: ZKPS AND AVOIDING ARROW'S PARADOX

1. ZKPs and Information Partitioning

Developing successful use cases for ZKPs in dealmaking would provide important means for avoiding the all-or-nothing disclosure choice faced by participating parties, reducing disclosure risk that make negotiations costly. ZKP's capacity to partition information could, in Michael Burstein's words, allow "sufficient information to be transferred to link ideas with capital and development partners" while also "ensuring that enough value remains in the original information holder so that she still has an incentive to disclose."[143] In certain contexts, ZKP's capacity to partition information could permit the opportunity for more limited disclosures, protecting proprietary information by minimizing the amount of information subject to Arrow's disclosure paradox and by avoiding the threat of mandated disclosure to legal or regulatory authorities down the line—a potentiality reflected in the exceptions contained in standard confidentiality agreements.[144] In other contexts, it could allow more extensive disclosures, such as when information is not shared because of the threat of antitrust or contract liability arising from sharing secrets with rivals or revealing confidential contract terms; or, similarly, because the disclosure would violate privacy mandates. In both cases, meaningful partition of information could facilitate a more "optimum level of appropriability."[145]

143. Burstein, *supra* note 38, at 254.

144. *See* Practical Law Corporate & Securities, Confidentiality Agreement (US-Style): Cross-Border Acquisitions 3 (2021), Westlaw w-002-6486 (discussing "Required Disclosure").

145. Burstein, *supra* note 38, at 254.

Such partitioning would facilitate the exploration and consummation of beneficial deals. Enabling limited information sharing—especially at the initial phases of negotiation—would allow parties to get a sense of potential partners' motives early on, facilitating ongoing negotiations in contexts where jeopardizing proprietary information is at a premium. This could prove particularly helpful in the start-up financing context, where specialized funds frequently focus on targeted market segments and engage in simultaneous discussion with numerous firms in the same business space, fostering suspicion among innovators about the funds' use and sharing of their information, creating a drag on the entrepreneurial system.

More generally, the successful development of meaningful applications for ZKPs in dealmaking may well have transformative effects on market structure itself. Such successes could transform the "boundaries of the firm" by expanding firms' choices regarding whether they need to protect innovation by keeping knowledge of the information within firm boundaries or, alternatively, profit from that innovation through partnerships or market transactions. Economic understandings of the ways firms organize, building on Ronald Coase's theories,[146] suggest that when the costs of transactions between firms exceed the benefits of those transactions, business functions will be kept or brought within the company—in other words, firms will vertically integrate.[147]

Intellectual property scholars have extended these insights to the information context, pointing to the disclosure costs resulting from Arrow's paradox as additional constraints on a firm's decision whether to perform functions internally or to contract with others to perform them. Robert Merges argues that strong intellectual property rights alleviate many of the costs

146. *See generally* Ronald H. Coase, *The Nature of the Firm*, 4 ECONOMICA 386 (1937) (exploring what accounts for the boundaries of firms, and why some production functions are executed within the firm, while other functions are executed outside the firm on the market).

147. *See id.* at 394–96; *see also* Burstein, *supra* note 38, at 245 ("The theory of the firm suggests that in the absence of other solutions to transaction costs, firms will vertically integrate"); Dan L. Burk & Brett H. McDonnell, *The Goldilocks Hypothesis: Balancing Intellectual Property Rights at the Boundary of the Firm*, 2007 U. ILL. L. REV. 575, 579 (2007) ("First, they must search for and identify each other as potential partners. Once they have found each other, they must negotiate with each other as to the terms that will govern their relationship in making the widget. Once they have reached agreement and entered into a contract, each party must monitor the performance of the other to ensure that it is doing as promised. Disputes may arise as to whether one or both has performed as promised. Each of these steps may generate costs that reduce the value that the transaction creates.").

associated with interfirm market transactions.[148] According to Merges, intellectual property rights resolve the disclosure paradox by making information excludable and eliminating the need for firms to integrate production functions into their hierarchies.[149] Consequently, Merges argues, stronger intellectual property rights, especially patents, facilitate efficient interfirm transactions,[150] and functions that were traditionally in-house will rather be executed on the market, ultimately resulting in "smaller, nimbler, and more specialized firms."[151]

Yet the calculus changes in contexts in which confidential or sensitive business information or trade secrets lack strong legal protection, as is frequently the case in transactional due diligence. By this logic, writes Michael Burstein, "the absence of property rights in information that firms need to transfer should lead those firms to integrate in order to accomplish the transaction"[152] rather than achieve their goals through joint ventures or market exchange. The development of ZKP methods for partitioning information so that less of it needs to be transferred, accordingly, could limit transaction costs in a way that expands firm choices, permitting verification of material elements without full revelation.

2. Use Cases

The range of sensitive information relevant to transactions will differ by case. Still, it is worth exploring a range of use cases that frequently arise in deals. Certain paradigmatic types of trade secrets that might be significant in determining a transaction's worth, such as manufacturing processes or characteristics or emergent qualities of recipes (like the uniqueness, or lack thereof, of various Coca-Cola beverages produced using slightly different ingredients) are not easily digitized. Accordingly, such types are not readily amenable to traditional ZKPs on digital data. But where information is already or can be digitized, statements about it are immediately amenable to verification with zero-knowledge.

148. *See* Robert P. Merges, *A Transactional View of Property Rights*, 20 BERKELEY TECH. L.J. 1477, 1513–14 (2005); *see also* Ashish Arora & Robert P. Merges, *Specialized Supply Firms, Property Rights, and Firm Boundaries*, 13 INDUS. & CORP. CHANGE 451 (2004).

149. *See* Merges, *supra* note 148, at 1503.

150. *See id.* at 1488–89, 1512–13.

151. Burk & McDonnell, *supra* note 147, at 615; *see* Merges, *supra* note 148, at 1507. *But see* Burk & McDonnell, *supra* note 147, at 615 (agreeing that overly weak intellectual property rights offer less utility in overcoming the disclosure paradox, leaving firms little choice but to turn to integration to protect their innovations, but also arguing that overly strong protections can also result in a "situation where property rights are fragmented or too finely divided, impeding or preventing desirable projects that entail such rights").

152. Burstein, *supra* note 38, at 245.

Easier use cases might include verifying customer numbers and characteristics without revealing full customer lists. Easier use cases could also include verifying contract attributes and terms—such as length, assignability, contingency, or even profit margins. Furthermore, information can be revealed with any agreed-upon level of granularity. For instance, contract terms or profit margins can be either fully disclosed or else asserted to be within a certain range. So long as both sides in a transaction agree on which elements would satisfy their need for diligence material to the transaction, and the information was encoded and stored digitally along these variables, using a ZKP would address privacy concerns and protect proprietary information while obviating the need for clunky, often imperfect, analog methods, such as creating and providing redacted versions of volumes of documents.

ZKPs offer an even greater promise in more complex situations where the partitioning of sensitive information to "exposed" and "unexposed" portions incurs additional challenges. In some of these cases, without ZKPs we are currently limited to either sharing all of the data or none of it, and neither option is desirable.

One such situation is the case where there is contention regarding the form, precision, and scope of the information to be disclosed, and some creative middle-ground solutions might be necessary. Here, ZKP's generality and flexibility greatly facilitate finding such middle ground. For instance, potential acquirers or venture capital funds may legitimately wish to assess the financial models that potential targets, or start-up firms, have used in projecting future performance. At the same time, the latter often hold that these models reveal confidential and proprietary information. ZKPs allow the parties to explore compromises by having the target firm assert certain partial information about their financial models (e.g., asserting that certain salient parameters are within a given range or disclosing only partial information on the outcome of the model).

Another complex situation is mergers where both parties have proprietary secrets they are reluctant to disclose, and at the same time each party wishes to learn certain partial information about the other party's secrets. In such situations, even disclosing to the other party the type of partial information one is interested in might compromise one's own secrets, so the above method of partitioning proprietary information to disclosed and undisclosed portions might run into a wall.

ZKPs can get around this seemingly inherent difficulty with the help of multiparty secure computation (MPC) technology, introduced in Section II.A.ii. MPC technology allows the parties of a merger discussion to (a) agree on which partial information each party should obtain, where that partial

information may depend both on one party's own secrets and on the other party's secrets, and (b) engage in an interactive protocol which implements the disclosures agreed upon in stage (a). Adding ZKPs on top the MPC protocol allows each party to verify that the information it obtained from the MPC protocol was computed as agreed and that the other party did not learn anything beyond what was agreed upon initially.

For instance, suppose two health companies, with two sets of patient data, wish to explore a merger or partnership. In assessing the value of the patient data held by the other firm, each wishes to know whether their respective customer pools are similar in characteristics (permitting synergies by expanding a particular approach to care across similar populations) or different (suggesting that the two do not overlap). They may even wish to perform a computation on the combined data to explore whether a merger might improve outcomes or aid in research. For a range of reasons—privacy, antitrust, protection of proprietary assets—they could not allow each other access to the data sets themselves. Moreover, data anonymization—which in some circumstances would permit its transfer—might eliminate the very characteristics (e.g., where the subjects live)—that make it valuable in the first place. Today, the only way to perform such checks is to engage a trusted third party and disclose all secrets to it. However, this is an expensive and risky solution that might well render the potential merger not worth pursuing. ZKPs along with MPC could potentially turn such negotiations to mundane routine.

Finally, perhaps the most challenging application—yet one that holds substantial promise—arises in verifying aspects of proprietary code, software, or algorithms. This is the type of issue at the heart of the *TargetSmart* allegation discussed above,[153] and often the core innovative element in a deal—hence often the most sensitive.

A potential partner, acquirer, or funder might want to verify both the correctness and validity of a program (*i.e.*, whether it "operates as intended") as well as its novelty. The challenge in verifying correctness is finding agreement on what its intended operation entails and codifying this agreement specifically enough to enable a clear resolution of the question. ("Correctness" can have multiple interpretations, in different contexts.) Still, once such agreement is reached, the actual test of correctness can be done via ZKPs, with the guarantee that nothing else is disclosed other than the result of the agreed-upon test of correctness.

The following two are worth highlighting in concluding this Section: First, deploying ZKPs in settings where the properties asserted about the hidden

153. *See* Part I.

information may be cumulative requires extra care (e.g., when the parties may agree on new properties to be asserted as part of an ongoing interaction or negotiation). In particular, it must be assured that all the properties are asserted with respect to *the same corpus of hidden data,* namely with respect to the same initial base document. Furthermore, the parties must be aware of the potential leakage of information from the aggregate of all the checks considered together.

A second point pertains to the inherent difficulties of formalizing potential checks as computer programs applied to the hidden data. As a telling example, asserting a property such as "novelty" in zero knowledge might prove to be a tricky business. Indeed, the crux of the difficulty is in translating the "novelty" claim to rigorously verifiable assertions. (This is the case even regardless of the need to keep the innovation secret.) One potential path for such translation is to turn the statement on its head and instead prove dissimilarity of the hidden algorithm to some known and plausible prior art candidates similar to the new algorithm. Here, again, a set of concrete, quantifiable measures of similarity between algorithms would need to be agreed upon. Once such agreement is obtained, and the measures have been encoded in a sufficiently specific way, ZKPs can be employed to assert that the hidden algorithm is sufficiently dissimilar to any one of the candidate prior-art algorithms.

C. GOVERNMENT OVERSIGHT: ZKPS IN ALGORITHMIC AND DATA ACCOUNTABILITY

As laid out in Section III.C, there are a host of situations where government accountability and the transparency of administrative and judicial decision processes stand in direct conflict with a perceived need to keep certain information secret. This Section outlines how ZKPs can be a game-changer in this domain, enabling transparency and accountability, while at the same time preserving (and even improving on) secrecy and privacy protections when they are justified and legitimate. Following the lead of Section II.C.1, this Section outlines potential uses of ZKPs to alter the all-or-nothing baseline choice regarding disclosure, in the case of algorithmic identification, algorithmic accountability, and privacy-preserving data verification.

1. Verifying the Identity of Algorithms

ZKPs can be used to assert that a value is the result of running a specific hidden algorithm on some data without revealing the code or other information about the algorithm. Determining how to specify the algorithm and how to make it amenable for a ZKP would need to depend on the case at hand. Appendix B presents a potential process for using ZKPs in a case like

that of the FST algorithm, outlined in Section 1.C., provided that—unlike with FST—the algorithm was of a sort that had a legitimate reason for secrecy.

In the context of financial stress tests, the benefit of ZKPs can be twofold: First, the stress tests themselves can deploy ZKPs to allow the Federal Reserve to verify assertions regarding financial information (such as volatility of investment portfolios) of the tested financial institution without violating the statutory confidentiality of same information.

Second, as in the IRS example, a public review board can determine a process for certifying financial stress tests. The certification process can take into account sensitive secret information known only to the Federal Reserve, as well as sensitive secret information known only to the tested financial institution. Still, the certification process itself would be transparent and open to public scrutiny. When performing a stress test, the Federal Reserve will provide the tested institution with a ZKP that the test has passed the certification process.

2. Verifying Characteristics of Algorithms

Another set of challenging situations outlined in Section II.C.1 involve the conflict between the need to keep certain government-run algorithms secret and the need to provide public evidence that these same algorithms behave in certain ways. This is one of the manifestations of the algorithmic accountability challenge.[154] Algorithmic accountability can manifest in many ways, including reviewing the algorithm's code directly or proving particular assertions about the code by subjecting it to tests like static analysis. However, these methods typically require full access to the analyzed algorithm.

ZKPs have the capacity to transform the field by again cutting the Gordian knot of secrecy: using ZKPs, governments—and even private entities that may introduce fairness concerns, like banks and credit agencies—can prove assertions about their algorithms without revealing them. This capability creates privacy-preserving accountability without resorting to full disclosure.

154. *See* Deirdre K. Mulligan & Kenneth A. Bamberger, *Procurement as Policy: Administrative Process for Machine Learning*, 34 BERKELEY TECH. L.J. 773, 791–96 ("Scholars have identified a number of ways that these systems operate as black boxes, inscrutable from the outside: (1) corporate secrecy, by which the design details are kept secret by private developers; (2) technical illiteracy—the impenetrable nature of system rules to non-engineers even where they are shared; and (3) the inability of humans, even those who design and deploy machine learning systems, to understand the dynamic models learned by complex machine learning systems."); *see generally* Kroll et al., *supra* note 14; Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54 (2019); David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. ON REG. 800 (2020).

Furthermore, ZKPs can do so ex ante, as a matter of process, rather than as part of a costly ex post litigation.

The benefit of using ZKPs is amplified in common situations where government agencies rely on contractors who may produce code that is too complex for the government agency itself to understand or that is kept secret for intellectual property reasons.[155] In such cases, the onus would be on the contractor to generate the ZKPs proving the agreed-upon assertions regarding the algorithm in use.

Returning to the challenge of asserting even-handedness in IRS auditing algorithms while preserving secrecy of the auditing algorithm itself, a ZKP-based solution could take the form of the following two-step process:

1) An IRS review board will determine a process for certifying auditing algorithms. The process might specify, say, criteria for selecting keywords to be used in identifying entities to be audited, as well as other limitations. Or it could instruct that the auditing algorithm be run on some benchmark sample of cases to detect potential bias. The certification procedure itself can be public and transparent.

2) Any IRS audit will be accompanied by a ZKP that the audit decision was made by an algorithm that passed the certification process that was approved by the review board. Both the algorithm and the data algorithm uses to determine the audit decision itself will be kept secret.

3. Privacy-Preserving Verification of Data

Another domain where transparency and government accountability stand in contrast with the need to keep salient information hidden is that of determining government policy based on data collected from or about individuals. Such data is often subject to use and disclosure restrictions to protect the privacy of the individuals whose data is used.

To allow for meaningful use of data about individuals (say, medical, economic, or social data) without violating these privacy constraints, mathematical methods have been developed for disclosing collected data in aggregate (and often perturbed) forms that prevent reidentification of individuals while still maintaining much of the data's utility for inferring salient properties of the population. Most studies and policy decisions can use the

155. *See* Mulligan & Bamberger, *supra* note 154, at 789 ("On the one hand, private developers keep much of the relevant code secret. On the other hand, agency staff frequently have few technical skills, so they can neither assess technology design shared with them nor participate in design themselves."); Kroll et al., *supra* note 14, at 647, 662, 685.

aggregated privacy preserving data in lieu of the original, raw data that compromises individual privacy.[156]

However, using privacy-preserving data aggregation methods incurs a potentially significant drawback: in and of itself, the aggregated data is not obviously tied to any individual or the raw data. A suspicious critic of a study or policy decision, then, has no way to verify whether the posted aggregate data corresponds to the actual raw data collected from individuals. Instead, the critic must trust the entity that presents the aggregate data to perform the aggregation and perturbation process as claimed. This situation is a bit unsettling, as there is no way to verify correctness of the aggregation.[157]

Using ZKPs, the entity that performs the data aggregation and perturbation can first provide a digital commitment to the actual raw data. Then, the entity would provide a proof that the aggregated data is the result of applying a prescribed and certified aggregation and perturbation method to the committed data. This can be done without exposing the raw data any further.

D. TRADE SECRET LITIGATION: ZKPS AND THE ADVERSARY PROCESS

Beyond mitigating the need for ex post legal remedies, ZKPs could also improve the function of those remedies that remain necessary. Here we concentrate on the case of trade secret litigation, where the ability to prove claims about secrets, while preserving their secrecy and value, is key.

There are several use cases, described below, where ZKPs can help solve litigation verification dilemmas while avoiding overdisclosing sensitive information.[158] We start with some straightforward cases and make our way to more complex ones. We consider how the level of complexity of deploying ZKPs increases along the following two axes: first, the level to which the alleged trade secret, as well as the other relevant secret information the parties hold, can be rendered as well-defined digital documents; and second, the ease of mechanizing the process of determining whether the defendant's documents constitute an alleged trade secret misappropriation.

156. One salient class of methods for aggregating and perturbing data to preserve privacy are *differential privacy* methods. *See* Cynthia Dwork, Frank McSherry, Kobbi Nissim & Adam D. Smith, *Calibrating Noise to Sensitivity in Private Data Analysis,* 3 THEORY OF CRYPTOGRAPHY 265, 265 (Shai Halevi & Tal Rabin eds., 2006).

157. This is the stated reason for the new EPA rule. *See* 86 FED. REG. 469 (Jan. 6, 2021).

158. Importantly, the claim is not that ZKPs could *determine* the legal status of a claim of trade secret misappropriation. Rather, ZKPs can help the parties analyze evidence that might support or negate a legal finding of misappropriation, without risking disclosure of their trade secrets to their litigation adversaries.

1. Case I (Customer Lists)

The first case is where the information provided by the parties exists in well-defined digital documents, and evidence of misappropriation, or lack thereof, is assessed by a purely mechanical process, namely the number of equal entries in the documents.

Consider a plaintiff alleging that the defendant misappropriated the plaintiff's secret database of valuable customers. Assume that the customer databases of both the plaintiff and the defendant are digitized and stored in well-defined locations, and it has been established that the only way for the defendant to obtain the names of these customers is by misappropriation of the plaintiff's database. Accordingly, the only remaining question is whether there is any sizable intersection between the plaintiff and defendant's respective databases. ZKPs then allow for any one the following interactions to take place:

1) Privacy for plaintiff:
    a. The plaintiff computes a cryptographic commitment C_p to its own database.
    b. The defendant reveals its own database D_d to the plaintiff.
    c. The plaintiff generates a zero-knowledge proof for the following statement: "The database that is committed to in C_p and the database D_d have x records in common."
2) Privacy for defendant:
    a. The defendant computes a cryptographic commitment C_d to its own database.
    b. The plaintiff reveals its own database D_p to the defendant.
    c. The defendant generates a Zero-knowledge proof for the following statement: "The database that is committee to in C_d and the database D_p have x records in common."
3) Privacy for both parties:
    a. The plaintiff computes a cryptographic commitment C_p to its own database.
    b. The defendant computes a cryptographic commitment C_d to its own database.
    c. The plaintiff and defendant engage in a two-party secure computation where they jointly generate a ZKP of the statement: "The database that is committed to in C_d and the database that is committed to in C_p have x records in common."

Clearly, each one of these three methods carries a different burden for each party. The first method gives an advantage to the plaintiff, in that it allows the plaintiff to keep the secrecy of its trade secret while requiring the defendant to expose its secret information to the plaintiff and the court. The second method gives the same advantage to the defendant: it requires the plaintiff to expose its trade secret to the defendant and to the court while allowing the defendant to keep the secrecy of its trade secrets from the plaintiff and from the court. The third option allows both parties to keep their secrets secret. It would be up to the court and the parties to determine which method to use.

2. Case II (Annotated Customer List)

Assume that the plaintiff's customer database also contains notes with additional (legitimate and worthwhile) information that the plaintiff collected about its customers, say the type of products they prefer. To assess whether the defendant misappropriated the plaintiff's secret database one may wish to determine the level of similarity between the plaintiff's notes and the defendant's notes. This might involve several context-sensitive aspects, including textual and semantic features of the notes.

To determine whether the notes are similar enough to support—or disparate enough to negate—an inference of misappropriation the parties could proceed in a similar way to the one described above. However, in this context, with a crucial additional first step whereby an algorithmic process should be set for determining whether the notes are similar enough. That is, an examining expert could first determine the criteria for whether the databases are similar enough. Next, the parties would agree upon a mechanized process for determining whether these criteria hold, given the plaintext databases. Importantly, the criteria and process for determining whether the criteria hold is determined without access to the databases themselves. Rather, they constitute an algorithm that would evaluate similarity of any potentially relevant pair of databases.

Once such a mechanized process is in place, the parties would run this process "in zero knowledge." This can be done in any one of the three alternative ways described above, with the difference that the criterion "the databases have *X* records in common" is replaced by "the agreed algorithm determines that the databases are similar enough or not similar enough" (whichever is the case).

Note that there may be multiple reasons for parties to want to keep information such as notes in a customer database secret. In addition to privacy concerns, companies sometimes lace their databases with "easter eggs"—fake or non-operational data laced into databases or code—that can be used to

identify trade secret theft. Here, the easter egg is a trade secret, and if one reveals what it is, then its value for detecting future misappropriation will reduce. Using a ZKP to determine the presence or absence of an easter egg in the opposing parties' database can keep the easter egg itself secret and thus operational.

3. Case III (Computer Programs)

Assume the alleged misappropriated trade secret is a computer program rather than a database. That is, the plaintiff claims that a certain computer program *P* sold or used by the defendant misappropriates a computer program *P'* that is the plaintiff's trade secret. Assume further that both programs *P* and *P'* are written in well-defined digital documents, and that the only remaining question is whether the programs are similar enough to support an inference of misappropriation. In this case, the parties can determine similarity of the programs while keeping them secret in very much the same way as in Case II. The only difference is that the algorithmic process for determining the programs' similarity might be somewhat more technically involved and require the assistance of an expert in programs of the relevant character. Importantly, as with Case II, the similarity-determining algorithm would be developed without access to the secret programs themselves. In particular, the expert is not encumbered with any secret information. Hence, a ZKP could have avoided the type of misconduct in the example described earlier where StubHub sought to hire a settlement expert as its own witness in subsequent litigation.[159]

4. Case IV (Mixed Media)

Finally, consider Case II (annotated customer list) again, but assume that either the plaintiff's list or the defendant's list appears in a variety of forms—say, some items appear on hand-written notes, others on voice recordings, yet others in multiple separate documents. Here, the process described in Case II will need to be augmented by two initial processes: (1) a process of pinning down a digital rendering of the information that the plaintiff claims as a trade secret, and (2) a process of pinning down a digital rendering of the information that allegedly misappropriates the secret. This should be done while preserving both parties' secret information. As per our example, process (1) can amount to the plaintiff providing digital commitments (see Section II.A) to the audio of relevant voice recordings, digital photocopies of the handwritten notes, and all the relevant digital documents. Process (2) is a bit trickier, and may be context-dependent: for instance, the defendant can be asked to provide digital

159. *See supra* Section II.D.1.

commitments to its database of customer information and provide assurance that it uses no other source of information on customers.

V. LESSONS FOR ZKP INFORMATION GOVERNANCE

This Part develops a framework for evaluating the policy implications of substituting ZKP technology as an information governance tool in lieu of, or in addition to, existing legal, technological, or institutional solutions to verification dilemmas. Adopting new governance technologies risks disrupting background presumptions against which existing law and practice have developed. Failing to understand these disruptions with specificity, in turn, threatens to render invisible the policy decisions that adopting new governance technologies can enact.[160] Being clear about these disruptions, by contrast, can provide what Deirdre Mulligan and Kenneth Bamberger have called "political visibility"—surfacing "the very existence and political nature of questions being resolved by design choices," which in turn makes them "visible to stakeholders and the broader public" and more amenable to purposive resolution.[161]

Accordingly, this Part begins by explaining some key technical prerequisites to implementing ZKPs. Next, it examines how meeting these technical prerequisites and deploying ZKPs would change information-protection practices as compared to existing legal rules. We argue that ZKPs carry policy implications along five broad axes: better enforcement, technological self-reliance, efficiency, stickiness, and specificity. Finally, we suggest a series of policy questions that decisionmakers considering adopting ZKP governance tools should consider.

A. TECHNICAL PREREQUISITES TO IMPLEMENTING ZKPS

First, it will be helpful to clarify certain technical prerequisites to the implementation of any ZKP. The fact that ZKPs take the form of mathematical proofs requires, for their operation, data in a certain form, operated on by a certain type of rule. Regarding the data, ZKPs require that the data must be *well-defined* and *unambiguously interpretable*. (Typically, the data will be digital. Non-digital data may be acceptable if it can be digitized via an effective and unambiguous mechanism.) In addition, two key elements characterize the implementation of the proof. First, the *information to be disclosed* about the hidden data must be specified. (Typically, a ZKP will involve disclosing a digital commitment to the private data. The commitment

160. Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-by-Design*, 106 CALIF. L. REV. 697, 772 (2018).

161. *Id.*

preserves the secrecy of the data while making it unequivocal. In addition, some characteristics of the data might be disclosed—for example, some bounds on the size of the data.) Second, *the assertion to be verified* about the hidden data must be decided and agreed upon beforehand. (Typically, the assertion will involve the hidden data, along with the disclosed information.)

These requirements of ZKP implementation affect the type of "translation" challenges relevant to considering the use of ZKPs to help resolve legal verification dilemmas. In particular, the case must be one where legal statements can be rendered in code and human judgment can be reduced to design requirements.[162] In particular, some of the cases discussed are "easy" precisely because attributes of the information being verified are already amenable to being encoded into variables (such as age range or data characteristics already captured in digital form). The more difficult cases present increasing complexity for encoding information into variables. Physical data, for example, must be measured and those measurements made digital. Rich data, by turn, may need to be reduced to simpler data, such as the earlier example of simplifying contractual terms. It also means that the measurements that form the ground truth must be agreed upon and unambiguous.

Finally, recall that a ZKP is not complete without an additional guarantee that the data considered in the mathematical proof are the same as the objects considered in the litigation or other legal verification dilemma. In particular, where the prover applies the ZKP to committed data, making sure that the committed data relate to the actual object at issue in the verification dilemma must be handled using other mechanisms that would be context specific. Such mechanisms could include a public hash, contract law, third party auditors, or court orders punishable by contempt. Indeed, if the prover runs the ZKP on false data, whether as a result of user error or malicious cheating, the outcome of the proof will be meaningless for the legal verification dilemma. As mentioned previously, this point of failure falls outside the scope of what a ZKP can address.

B. FIVE AXES FOR EVALUATING ZKP POLICY

Having laid the technical foundation, this Section explores the ways that ZKPs can change the nature of information governance as compared to the

162. *See* Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, 676 (2010) ("Computer code . . . operates by means of on-off rules, while the analytics it employs seek to 'quantify the immeasurable with great precision.'"); Danielle Keats Citron, *Technological Due Process*, 85 Wash. U. L. Rev. 1249, 1303 (2008) ("Automated systems inherently apply rules because software predetermines an outcome for a set of facts.").

baseline of current legal practice. We argue that these changes occur along five axes—better enforcement, self-reliance, efficiency, stickiness, and precision—each of which carries policy tradeoffs.

1. Better Enforcement

Supplementing legal remedies with ZKP technology can better enforce existing legal rules. For instance, as detailed in Section II.A, existing legal solutions to verification dilemmas in information privacy and security rely on imperfect notice-and-consent regimes, unreliable anonymization mandates, and often-prohibitively expensive ex post litigation remedies that fail to correct for unidentified misappropriation or other harms. In contrast, as detailed in Section IV.A, ZKPs offer a new alternative of lesser-disclosure, which makes ex ante consent more meaningful. ZKPs can also eliminate the need to collect, duplicate, and aggregate personal identification data, which in turn avoids problems of unreliable anonymization. ZKPs can also make ex post litigation remedies unnecessary, thereby avoiding their costs and oversights. Hence, ZKPs can better enforce existing legal safeguards.

Notably, scholars have long debated policy preferences for either more or less comprehensive enforcement of law in different circumstances.[163] Although ZKP-enabled enforcement may well be optimal for some privacy and security protections, it could also eliminate beneficial leakiness in status quo legal safeguards. For instance, the U.S. Supreme Court has identified leakiness as a key feature of substantive trade secret law that distinguishes it from, and prevents its preemption by, federal patent law.[164] Here, we identify enforcement as a key axis that the adoption of ZKPs can alter, noting some positive use cases in the protection of privacy. The policy consequences of other applications should be evaluated on a case-by-case basis.

2. Technological Self-Reliance

As a second axis of alteration, ZKPs offer a technological self-reliance mechanism to extend information protections beyond existing legal rules. In other words, ZKPs can completely seal information in circumstances where even perfect enforcement of existing law would have permitted "knowledge spillovers."

For example, consider the disclosure of trade secrets under a contractual NDA or judicial protective order. As discussed in Sections I.B and III.B, even if the recipient fully abides by the terms of that agreement or order, merely

163. *See, e.g.*, EDUARDO M. PENALVER & SONIA K. KATYAL, PROPERTY OUTLAWS: HOW SQUATTERS, PIRATES, AND PROTESTERS IMPROVE THE LAW OF OWNERSHIP (2010).

164. *See* Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 490 (1974).

viewing the information might advantage them to the detriment of the discloser or in a manner that implicates antitrust concerns. Such "knowledge spillovers" are inherent in the sharing of information that cannot be unseen. Existing law often may provide no remedy against these more abstract knowledge transfers, sometimes even encouraging them. Trade secret law, for instance, encourages abstract knowledge transfers by explicitly exempting an employee's "knowledge, skill, and experience" from trade secret protection[165] and by disfavoring injunctions based solely "on the information the person knows"[166] (rather than on evidence of actual or threatened misappropriation). If firms deploy ZKPs to limit what information employees receive, then employees may not gain the same level of knowledge, skill, or experience to transfer with them to a future employer.

ZKP technology can extend privacy protections beyond existing legal rules by eliminating certain information transfers entirely and, as a result, their derivative knowledge spillovers. Such technological self-reliance "use constraints"[167] above and beyond even perfectly-enforced legal entitlements could benefit ZKP adopters yet simultaneously impose societal drawbacks. For instance, ZKP-enabled lesser-disclosures of information could impede auditing, making it harder to identify and correct mistakes. Lesser-disclosures may also eliminate serendipitous, unanticipated discoveries that could be gained from reviewing broader swaths of information. Scholars have debated similar issues in relation to Digital Rights Management technologies and copyright law. For instance, Julie Cohen has identified the potential for Digital Rights Management to "automatically enforce limits on user behavior" and create a governance mechanism that does more than existing legal regimes.[168] Similarly, ZKPs could permit trade secret owners to preempt the policy balancing built into current legal systems by imposing broader technological self-reliance protections than existing legal regimes would permit.

3. Efficiency

Supplementing law with ZKP technology can also improve efficiency in comparison to both legal and technical baselines—our third axis. ZKPs can substitute cheaper ex ante technical protections for costly litigation remedies. When ex post litigation remains necessary, ZKPs can make it cheaper to

165. *See* Camilla Hrdy, *The General Knowledge, Skill, and Experience Paradox*, 60 B.C. L. REV. 2410, 2410 (2019).

166. 18 U.S.C. § 1836(b)(3)(A)(i) (2018).

167. *See* Mulligan & Bamberger, *supra* note 160, at 717 (discussing the phenomenon); Margaret Jane Radin, *Regulation by Contract, Regulation by Machine*, 160 J. INST. & THEORETICAL ECON. 142 (2004).

168. Julie Cohen, *DRM and Privacy*, 18 BERKELEY TECH. L.J. 575, 580 (2003).

protect information during legal procedures. Consider the example of redacting voluminous records described in Section IV.B. Even setting aside the issue of leaks through sloppy redactions that either overlooked key information or redacted in an unsecure manner,[169] redactions take time and resources. Using ZKPs instead may be a more efficient, as well as a more reliable, solution.

Of course, as with each axis that ZKPs alter, increasing efficiency may sometimes produce undesirable policy consequences. For one illustrative example, courts and scholars grappling with the relationship between the Fourth Amendment and technological change have debated whether the ease and affordability of new surveillance technologies might undermine prior de facto privacy safeguards produced by cost.[170] So too, the substitution of manual redaction with easier and cheaper ZKPs might lead to excessive secrecy. Whereas the resource-intensive nature of manual redaction could encourage verifiers to err on the side of over-disclosure, cheap and easy ZKP-enabled micro disclosures could encourage verifiers to rely on ZKPs to limit disclosures of relevant information even when broader disclosures would have been acceptable.

4. Stickiness

Supplementing law with ZKP technology can increase the *stickiness* of assertions. Recall ZKP's requirement for precise, unchangeable pre-specification and pre-commitments. Stickiness means that this requirement pins down representations concerning information and choices about decisional rules earlier than existing legal regimes. This pinning down also limits a user's capacity to evolve over time. Stickiness may be good or bad. It prevents subsequent nefarious tampering but also makes it harder to correct initial mistakes.

For example, in the context of government algorithmic oversight, applying ZKPs would require ex ante commitments about government processes and compel the government to stick by that process.[171] Similarly, ZKPs require the

169. *Cf.* Tucker Higgins, *Justice Department Mistakenly Reveals Indictment Against Wikileaks' Julian Assange*, CNBC (Nov. 16, 2018), https://www.cnbc.com/2018/11/16/doj-mistakenly-reveals-indictment-against-wikileaks-julian-assange.html (recounting the accidental filing of sealed information on a public docket).

170. *See, e.g.*, Kiel Brennan-Marquez & Stephen E. Henderson, *Search and Seizure Budgets*, 13 U.C. Irvine L. Rev. at 9-11 (forthcoming 2023) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3910743 (proposing numerical caps to artificially constrain the state's search and seizure capacity given that technological change can make the production of surveillance data "trivial").

171. *See* Kroll et al., *supra* note 14, at 668-69.

government to commit to pinning down a piece of data, and making all subsequent determinations depend on that pinned-down data.[172] This characteristic of sticky representation of information is also evidenced by the potential use case for ZKPs in trade secret litigation. Applying ZKPs would require a mechanism for a plaintiff to commit to an early "identification" of their alleged trade secrets without having to disclose the information at the time of commitment. As the litigation proceeds and the plaintiff learns more information from the defendant in discovery, the plaintiff could amend its complaint to add new relevant details about its trade secrets. Unlike the status quo, where these modifications might lead to accusations of gaming, plaintiffs could rely on a ZKP to establish that the subsequent amended claims match the early identification, all the while maintaining secrecy.

The stickiness of ZKP assertions, of course, may be counterproductive, particularly in rapidly changing circumstances. For instance, consider our discussions of the encumbered witness in litigation[173] or the allegations of bias in IRS algorithms.[174] If the procedure that the parties and expert witness commit to in advance is imperfect, producing an unanticipated result at "run time," then reliance on ZKPs may make it harder to identify the mistake and to exercise judgment to correct it after the fact.[175] To remedy these issues, ZKP users might have to design procedures to create alerts and opportunities to revisit the pre-commitments in cases of unanticipated situations or error.

5. Specificity

Finally, supplementing law with ZKP technology can increase the specificity of rules, legal or otherwise. That is, the technical specificity required to implement a ZKP, whereby rules and assertions must be predefined as a sequence of basic operations on data, forces the resolution of policy decisions that may be implicit in status quo legal rules and practice. For example, status quo legal rules and practice may presume that all-or-nothing disclosure choices are the sole means to resolve verification dilemmas. ZKPs create a new, intermediate disclosure option that can force a debate about how much disclosure law should encourage. In this sense, ZKPs can also create the possibility of devising new laws and regulations mandating intermediate disclosures that were previously impractical.

172. *See id.*

173. *See supra* Section IV.D.

174. *See supra* Section IV.C.

175. *See* Mulligan & Bamberger, *supra* note 160, at 715 (discussing the ways that "the implications for values occur—and can shift—at design, configuration, *and* run time" (citing David D. Clark, John Wroclawski, Karen R. Sollins & Robert Braden, *Tussle in Cyberspace: Defining Tomorrow's Internet*, 13 IEEE/ACM TRANSACTIONS NETWORKING 462, 463 (2005))).

Specificity may introduce context-specific translational challenges, some of which might be less difficult to resolve than others. For instance, testing a claim that an algorithm had a certain level of accuracy in classifying a dataset might be relatively simple to encode and prove. And even translational problems that require more judgment may sometimes be resolved by simple agreement in transactions involving private parties. Take the dealmaking context, for example, where the relevant attribute of information sought to be proven in a zero-knowledge setting is subject to definitional judgment—such as how to measure the "similarity" of two parties' trade secrets or how an algorithm or model is "intended to operate." In this context, the parties can come to mutual agreement on a methodology for comparing those trade secrets or assessing an algorithmic function. The parties can then reduce the secrets to a digital formula.

In other settings, translational considerations might be a bigger hurdle. For example, where translation implicates public policy—such as in deciding the appropriate level of privacy or security protection to mandate or using ZKPs in governmental oversight—the construction of ZKPs might require a broader discussion involving transparency about the issues at stake and stakeholder involvement. This is particularly true the more the elements being proven involve standards that are subject to definition—for instance, with suggestions that ZKPs might be used to prove compliance with regulatory mandates, such as the Federal Reserve's capital adequacy requirements demanding "sufficient" reserves or the example discussed above of auditing the IRS for bias.

Yet although the prerequisites for ZKP implementation often require such translation and the judgment inherent in it, the characteristic of specificity offers a logical step for ex ante transparency about the ways such decisions involve choices about policy—not simply one-to-one reduction. Surfacing those policy questions creates new opportunities to foster debate about them. As computer scientists Michael Kearns and Aaron Roth explain in a related context, the process of specificity "has great merit in its own right—both because it is necessary in the algorithmic era" and also because it surfaces policy implications and decisions in the use of on-off algorithmic application. In essence, specificity "often reveals hidden subtleties, flaws, and trade-offs in our own intuitions . . . ."[176] Thus the technical prerequisite of specificity forces an opportunity to have a policy discussion.

176. MICHAEL KEARNS & AARON ROTH, THE ETHICAL ALGORITHM: THE SCIENCE OF SOCIALLY AWARE ALGORITHM DESIGN 18 (2020).

This policy debate-forcing function of ZKP-enabled specificity is especially salient[177] in opening new questions as to whether disclosure, or lack thereof, is desirable as a policy choice. Consider questions about whether to collect data identifying race. Current law, for example, prevents the use of certain categories of information in lending decisions.[178] So, one might think that using ZKPs to exclude this information from data collection would promote privacy and fairness. However, converting legal prohibitions on use into technical mechanisms for non-collection may be counterproductive for broader policy goals. It might significantly hinder access to the data necessary to explore disparate impact concerns in lending.[179] It might also hinder current exploration of the use of artificial intelligence algorithms to correct for historic bias by integrating goals such as forward-looking expansion of mortgage access.[180]

There are also other types of policy choices that are implicit in current legal practice and that ZKP-enabled precision can foreground and open to debate. Take the case study of the FST algorithm the New York City Office of Chief Medical Examiner uses. In the existing legal baseline, the regulator approved use of the FST software system without specifying precisely how to define what that system was—for instance, whether fixing a bug or recompiling the program for a different operating system should vitiate the regulatory approval. Ambiguity in the legal definition of what had been approved by the regulator apparently left to unidentified employees within the Office of Medical Examiner the crucial policy decision about which system alterations amounted to core or substantive changes requiring revalidation and new regulatory approval. Using a ZKP could have enabled the regulator to set that policy choice by selecting the level of specificity or generality for the signed

177. *Cf.* Jack Balkin, *The Path of Robotics Law*, 6 CALIF. L. REV. CIRCUIT 45, 46 (2015) ("When we consider how a new technology affects law, our focus should not be on what is essential about the technology but on what features of social life the technology makes newly salient.").

178. *See, e.g.*, Abbye Atkinson, *Borrowing Equality*, 120 COLUM. L. REV. 1403, 1407 (2020).

179. Mulligan & Bamberger, *supra* note 160, at 728 ("Reducing the collection of data about protected class status can constrain its intentional use to discriminate. But it removes data that is useful if not essential for identifying the latent, redundant encoding of protected traits that algorithms are so adept at finding."); *see also* Ignacio N. Cofone, *Antidiscriminatory Privacy*, 72 SMU L. REV. 139, 141 (2019) (highlighting examples where privacy is helpful, such as identifying facially neutral screening rules, and examples where it is not helpful, such as affirmative action cases).

180. Sian Townson, *AI Can Make Bank Loans More Fair*, HARV. BUS. REV. (Nov. 6, 2020), https://hbr.org/2020/11/ai-can-make-bank-loans-more-fair (discussing methods to prevent lending bias by "regulariz[ing]" an algorithm "so that it aims not just to fit historical data, but also to score well on some measure of fairness," which requires "including an extra parameter that penalizes the model if it treats protected classes differently").

commitment that could later be validated with zero knowledge. The ZKP would thus have surfaced the policy choice about whether, and how much, the grant of regulatory approval included flexibility to alter the system and shifted the decision from the employee within the Office of Medical Examiner to the regulator itself.

Similarly, the use of ZKPs may force explicit engagement with the question of what variables might indicate an algorithm's fairness, or a lack of bias in its choice of data set.[181] Moreover, there is an inherent tension in the deployment of algorithms, accountability, and secrecy. Regulation favors transparency for several reasons, including classical arguments of fairness and accountability to the public. Yet, regulators do sometimes have legitimate interests in limiting the disclosure of regulatory information, whether it is the rules themselves or the results.[182] These interests should be protected, but they also introduce the opportunity for regulators to overclaim secrecy. Because implementing ZKPs requires specificity about what the ZKP will be able to prove, the process of implementing the tool will force policy debates around these crucial questions concerning the costs and benefits of transparency.

Considering our discussion throughout this Part, ZKP's technical and policy attributes lead to four questions that should be discussed before implementing ZKP governance tools in any given circumstance:

> *First*, what is the value in keeping information undisclosed as long as said properties have been verified?
>
> *Second*, where the status quo baseline is full disclosure, could substituting that disclosure with ZKPs cause a loss of serendipitous value?
>
> *Third*, is using a ZKP worth the complexity and burden of implementation? If so, then which of the parties should carry the burden of performing the ZKP? In cases where the information is held by only one of the parties, it is natural that this party will be the one carrying the burden because it is the only party with a secret to keep. In cases where both parties have secret information the determination might be less clear.

181. *See* Kroll et al., *supra* note 14, at 633 (suggesting a "technological toolkit to verify that automated decisions comply with key standards of legal fairness").

182. Limiting the disclosure of regulatory or investigatory results also implicates fairness under the law. This is the same reason that, for example, the Department of Justice is exempt from having to confirm nor deny the existence of an investigation in response to a FOIA request. *See* 5 U.S.C. § 552(c)(1)(A) (2018).

Fourth, does the specificity requirement of ZKPs force new policy choices, and if so, what is the best way to resolve those choices in each context?

VI. CONCLUSION

Status quo legal rules and practice often presume an all-or-nothing choice between costly verification through overdisclosure or costly secrecy through underdisclosure. Private actors must choose either to forego the benefits of verification or to undertake risks from disclosure of sensitive information. Public policy, in turn, often faces a binary option between a full transparency and extreme opacity. Full transparency imposes untenable policy costs of its own—whether by undermining distinct public concerns such as information privacy, or subverting the very efficacy of the operations rendered visible, as with an IRS investigatory algorithm. Yet extreme opacity is in tension with public accountability norms. Existing legal solutions to this conundrum mitigate the risks of overdisclosure but cannot entirely eliminate them. New technologies often shape the environment in which they are used.[183] ZKPs are uniquely capable of protecting data not just from unauthorized parties but between communicating parties themselves, and it is possible that as ZKPs continue their development they will catalyze a new paradigm built not on unlimited exposure but instead on controlled disclosure.

This Article has developed case studies that demonstrate the possibility of using ZKPs to help solve verification dilemmas across multiple areas of law. ZKPs offer the occasion to disrupt the presumption that verification dilemmas present an all-or-nothing disclosure choice, and to address some of the costs those dilemmas frequently impose in legal contexts. By changing understandings of the quanta in which information can be disclosed and the possibility of separating discrete qualities of that information from the underlying data for purposes of verification, ZKPs offer the promise, in certain circumstances, of previously unavailable ways to sever portions of data for sharing. The result is more efficient means in contexts where limited disclosures are currently attempted and more effective ways of reducing or eliminating disclosure risk that the law is currently unable to achieve. At the

183. For example, Marc Andreessen, the inventor of the web browser, called the lack of payment technology on the early internet its "original sin," arguing that the internet primarily uses advertising to monetize because it was not originally technically feasible to process payments online. Marc Andreessen, *From the Internet's Past to the Future of Crypto*, A16Z PODCAST, at 17:00 (Aug. 29, 2019), https://a16z.com/2019/08/29/internet-past-crypto-future-crypto-regulatory-summit/ ("Because we were unable to build payments into the browser . . . as a consequence, that is why the internet today, at least in the U.S., is predominantly based on advertising.").

same time, by transforming the background against which existing disclosure norms and practices have developed and the law has evolved, ZKPs raise important challenges for the future of law and policy.

APPENDIX A: MORE ON CONSTRUCTING ZERO-KNOWLEDGE PROOFS

This Appendix expands on method I (the boxes method) for constructing ZKPs, described in Section II.B.1.a:

To commit to data *m*, the committer chooses a random number *r* in a predefined range (which corresponds to the key of the locked box) and applies a special algorithm COM on inputs m and r, to obtain a value c. (In shorthand, the committer obtains $c = COM(m, r)$.) The value c, henceforth called the "commitment value," is given to the recipient. To reveal m, the committer sends m and r to the recipient, who verifies that *c* = COM(*m, r*). The commitment value c represents the box and the random number r represents the key. The algorithm must satisfy two guarantees:

1) The data m remains secret even knowing c; and
2) the committer cannot feasibly find a commitment value c, two values m_1, m_2 and two keys r_1, r_2 such that $c = COM(m_1, r_1)$ and at the same time $COM(m_1, r_1)$.

Traditional cryptographic commitments satisfy (a) and (b). For ZKPs, we will need COM to satisfy yet an additional property which is called *homomorphism with respect to a mathematical operation on pieces of data*. This mathematical operation is denoted by *. The property is stated as follows:

3) Given two commitment values $c_1 = COM(m_1, r_1)$ and $c_2 = COM(m_2, r_2)$, the recipient should be able to compute a third commitment value c_3, and the committer should be able to compute a value r_3, such that:

I. $c_3 = COM(m_3, r_3)$, for $m_3 = m_1 * m_2$. Namely, c_3 is now a commitment to the value $m_1 * m_2$. Furthermore, the committer who knows r_1 and r_2 can compute r_3 that can be used to open c_3.

II. the original values $m_1$and m_2 remain hidden (aside from what is revealed about them from knowing m_3), even when c_1, c_2, c_3 and r_3 are known. This means that the committer can now open $c_3$to m_3 (by exposing m_3, r_3) while still keeping m_1, m_2 hidden.

The "homomorphism" property allows evaluating the "*" operation directly on committed data without learning the data itself. This "veiled evaluation" operation has no immediate physical analog, other than being somewhat akin to "mashing" a box that holds data m_1 with key r_2 with a box

that holds data m_2 with key r_2 into a third box that contains $m_1 * m_2$ and that is openable by a key r_3 that's constructed from r_1 and r_2.[184]

ZKPs require commitments that are homomorphic with respect to all the operations that the verification program employs. Fortunately, there exist relatively simple operations on data that are "universal": any computer program with any instruction set, including our verification program, can be rewritten as a sequence of applications of only the universal operation on different portions of the data. The operation $*$ will be such a universal operation.[185]

Assuming a homomorphic commitment scheme as described above, the ZKP protocol for an assertion is now straightforward:

0. Both the prover and the verifier agree on the verification program V which consists of a sequence of $*$ operations. In addition, the prover has the input proof *pf* of the assertion, written in binary (i.e., a sequence of 0's and 1's).

1. The prover commits each binary number in the input proof *pf*. Namely, for $= m_1, \dots, m_n$, the prover chooses n random keys $r_1, \dots, r_n$, and gives the verifier values $c_1, \dots, c_n$ such that $c_1 = COM(m_i, r_i)$.

2. The verifier homomorphically evaluates the verification program on commitment values $c_1, \dots, c_n$. That is, for each operation $m_i * m_j$ in the verification program, the verifier performs the corresponding operation on the commitment values c_i and c_j to obtain c_{i*j}. At the end of this process, the

184. To further illustrate the concept, we sketch the commitment algorithm proposed by Pedersen. *See* Torben Pryds Pedersen, *Non-Interactive and Information-Theoretic Secure Verifiable Secret Sharing*, *in* ADVANCES IN CRYPTOLOGY – PROC. OF CRYPTO '91, LECTURE NOTES IN COMPUT. SCI. 129, 130 (Joan Feigenbaum ed.). Assume that a large prime number p is known to all, along with a number g which is a generator of the multiplicative group Z_p, and a random group element h. Then, $COM_{p,g,h}(m,r) = g^m \cdot h^r$. (Here we assume that m can be represented as a number in $1..p-1$, and $\cdot$ denotes multiplication modulo p. Observe that this commitment algorithm is homomorphic with respect to addition modulo $p-1$. If $c_1 = g^{m_1} \cdot h^{r^1}$ and $c\_2 = g\^{m\_2}\cdot h\^{r\^2}$ then it holds that $c_3 = c_1 \cdot c_2 = g^{m_1} \cdot h^{r^1} \cdot g^{m_2} \cdot h^{r_2} = \mathrm{g^{m1+m2} \cdot h^{r1+r2}}$. This means that c_3 is a commitment to $m\_1 + m\_2$ with key $r\_1 + r\_2$ (mod p-1). Security of this commitment protocol holds under a widely believed mathematical conjecture (Decisional Diffie Hellman in certain prime order groups.)

185. There are many universal operations. For example, we can choose the NAND operation: 0 NAND 1 = 1 NAND 0 = 1 NAND 1= 0, whereas 1 NAND 1 = 0. NAND is a *universal operation*. It is possible to write any computer program using only NAND operations, applied to different parts of the input data and the program's memory. Professors Groth, Ostrovsky, and Sahai have designed commitments which are homomorphic with respect to NAND. The security of these commitments relies on another widely believed mathematical conjecture (subgroup indistinguishability in certain composite-order groups that enable bilinear maps). *See generally* Jens Groth, Rafail Ostrovsky & Amit Sahai, *New Techniques for Noninteractive Zero-Knowledge*, 59 J. ACM 1 (2012).

verifier obtains a commitment value c_{out}, which is guaranteed to be a commitment to the output value of V.

3. The prover performs a similar sequence of operations with respect to the keys $r_1, \ldots, r_n$.That is for each operation $m_i * m_j$ in the verification program, the prover performs the corresponding operation on the keys r_i, r_j to obtain r_{i*j}.At the end of this process, the prover obtains a key r_{out} that corresponds to the output value of V. The prover then sends $r_{out}$to the verifier.

4. The verifier verifies that $c_{out} = COM(1, r_{out})$. If the verification succeeds, the verifier agrees that the original verification program accepts the original (committed) input proof. (The fact that c_{out}opens to 1 implies that $V(pf) = 1$, hence the proof is correct.)

There are several ways to design an algorithm COM which satisfies properties (a)–(c) defined above. One method has the prover and verifier engage in a preliminary three-round (back and forth messages exchanged) cryptographic protocol in which they agree on a randomized choice of COM which neither one can control so as to violate soundness or zero-knowledge: the verifier needs the randomness guarantee to ensure soundness and the prover needs the randomness guarantee to ensure zero knowledge.

APPENDIX B: USING ZKPS IN AN FST-LIKE CASE

Here is how a ZKP-based solution might work in the case of an algorithm similar to the New York City Office of Chief Medical Examiner's Forensic Statistical Tool, discussed in Part IV.C.1, if, hypothetically, there were a legitimate reason to keep that other algorithm secret:

1) The regulator prepares a document that specifies the approved algorithm. Here the level of detail by which the algorithm is specified is of central importance: the algorithm should be specified at a level of detail that suffices for guaranteeing the properties that the regulator sees as critical to the adequacy of the algorithm to the stated use case. To maximize usability and minimize the need to re-accreditation, the regulator might leave out details that are deemed irrelevant to those critical properties. (For instance, the regulator might choose to specify the approved algorithm by way of a higher-level, safe programming language, such as Rust, which provides explicit functional consistency guarantees for programs, regardless of the specific execution environment. Alternatively, the regulator might specify the algorithm in a more flexible language such as C++, Java, or Python, and, in addition, specify the allowed "program libraries" that the algorithm might link to at runtime. If the regulator chooses to be more specific

in the accreditation, then it might sign the algorithm in the form of a specific executable program, thus specifying the program down to a specific configuration of "virtual machine" or an actual computer and forcing the prosecution to obtain a new accreditation for each new computer or virtual machine that the prosecution may use.)

2) The regulator augments the document A holding the algorithm with a "cover letter" that asserts that the signed algorithm has passed the regulator's test. Next, the regulator digitally signs the augmented document A, Let V_R denote the regulator's public signature verification key, and let S_A denote the signature of the regulator on the document A. (Recall, signature schemes come with a verification procedure Ver such that $Ver(V_R, X, S) = 1$ only if the regulator signed the document X.)
3) When the prosecution presents the result, T, of running the algorithm specified in document A on the relevant data D , it will be required to present also a ZKP of the following statement: "There exists a document X and a signature S such that:
    a. Document X includes a cover letter asserting that the regulator approved the algorithm described within.
    b. $Ver(V_R, X, S) = 1$. In essence, the signature verification procedure, when given public verification key V_R, document X, and signature S. outputs 1.
    c. $A(D) = T$. In essence, when executing the algorithm described in document A on data D, the output is T.

This proof will be computed using a special software tool for ZK proof generation, run by the prosecution.

4) The court and the defendant will then verify the assertion made by the prosecution. For that purpose, they will run a special software tool for ZK proof verification.