
THE RUTGERS UNIVERSITY LAW REVIEW
VOLUME 74 | ISSUE 1 FALL 2021

CREDIBILITY IN AN AGE OF ALGORITHMS

Julia Simon-Kerr[*]

ABSTRACT

Evidence law has a "credibility" problem. Artificial intelligence creators will soon be marketing tools for assessing credibility in the courtroom. While credibility is a vital concept in the United States legal system, there is deep ambiguity within the law about its function. American jurisprudence assumes that impeachment evidence tells us about a witness's propensity for truthfulness. Yet this same jurisprudence focuses fact-finders on a distinct inquiry: whether a witness has the status or outward appearance of a person who is worthy of belief. In the face of this equivocation about what credibility in the legal system is or should be, the terms of engagement will be set by the creators of algorithms in accordance with their interests.

This Article illuminates the actual and purported function of credibility in the law through analogies to two existing algorithmic products. One is the U.S. financial credit score. The other is China's experiment with a "social credit" scoring system. These analogies show that a predictive approach to credibility is structurally distinct from a worthiness-centered one. They also deepen critiques of both approaches as they appear in current practice and as we contemplate the credibility of the future.

* Professor of Law, The University of Connecticut School of Law. Many thanks to Steven Wilf for early conversations about social credit and credibility. Thanks to Kiel Brennan-Marquez, Teneille Brown, Bennett Capers, Virginia Kerr, Jamelia Morgan, Anna Roberts, Richard Wilson, and Carly Zubrzycki for comments on this and other drafts. Thanks also to participants at the Rutgers and Pepperdine Faculty Workshops, the Evidence Summer Workshop, and attendees at the Wesleyan Center for the Humanities Lecture Series for helpful feedback on versions of this work. Thanks to Cait Barrett for her outstanding research assistance.

TABLE OF CONTENTS

I. INTRODUCTION

What would it look like to impeach witness credibility with algorithms? This Article takes up this as yet unexplored question for two related reasons. First, big data and algorithmic tools are an increasing part of the law.[1] Second, this burgeoning use of big data makes it imperative to clarify a duality in how the law approaches credibility.

1. Legal decisionmakers have already welcomed big data and the algorithms that interpret it into adjudicative spaces. *See, e.g.*, Megan Stevenson, *Assessing Risk Assessment in Action*, 103 MINN. L. REV. 303, 314–16 (2018) (explaining the different kinds of risk assessment tools currently being used throughout the criminal justice system, including "checklist-style" and machine learning tools); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2221–22 (2019) (noting the heightened concern of algorithmic bias in criminal justice risk assessment) [hereinafter *Bias In, Bias Out*]; Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242, 243–47 (2019) (arguing that automated judging is imminent and the government must shape the new market for AI justice). Algorithms that claim to detect character traits, including deception are used now in the corporate world. *See* Hannah Devlin, *AI Systems Claiming to 'Read' Emotions Pose Discrimination Risks*, GUARDIAN, (Feb. 16, 2020, 12:00 PM), https://www.theguardian.com/technology/2020/feb/16/ai-systems-claiming-to-read-emotions-pose-discrimination-risks; Andrea Murad, *The Computers Rejecting Your Job Application*, BBC (Feb. 8, 2021), https://www.bbc.com/news/business-55932977 (describing increasing use of AI to screen job applicants, including by multinational companies such as McDonald's, Kraft Heinz, and J.P. Morgan). These practices have huge inequality implications because of the discrimination inherent in AI tools. *About*, ALGORITHMIC JUST. LEAGUE, https://www.ajl.org/about (last visited Oct. 28, 2021) (describing their mission of combating AI discrimination and pushing for more ethical AI); *see* Joy Buolamwini, *Artificial Intelligence Has a Problem with Gender and Racial Bias. Here's How to Solve It*, TIME (Feb. 7, 2019, 7:00 AM), https://time.com/5520558/artificial-intelligence-racial-gender-bias/. Artificial intelligence ("AI") creators have already targeted

As I have argued in past work, legal credibility has two main meanings: it can refer either to a witness's propensity for truthfulness or her *worthiness* of belief.[2] Without acknowledging this duality, the jurisprudence of credibility has long embodied, and indeed freely manipulated, these distinct conceptions of what is meant by the term. Indeed, the Supreme Court recently reiterated this duplexity, writing that a witness may be "lacking credibility—in the sense that she [i]s lying *or* not 'worthy of belief.'"[3]

This Article clarifies what is at stake in these two credibility paradigms by contemplating how algorithms might be brought to bear on each of them. It explores these possibilities by analogy to two existing algorithmic products. One is our own system of gauging financial creditworthiness through the credit score. The other is China's experiment with a social credit scoring system.[4] These examples confirm the fundamental, yet unheeded, distinction between a worthiness-centered approach to credibility and one that tries to offer a risk-predictive or probabilistic account of a witness's truthfulness. In doing so, they heighten critiques of the current system that I have offered in

the market for credibility evaluation. Lie-detection and other demeanor-based algorithms are being advertised to governments for use in border security settings. *See, e.g.*, Camilla Hodgson, *AI Lie Detector Developed for Airport Security*, FIN. TIMES (Aug. 2, 2019), https://www.ft.com/content/c9997e24-b211-11e9-bec9-fdcab53d6959; Ryan Gallagher & Ludovica Jona, *We Tested Europe's New Lie Detector For Travelers—and Immediately Triggered a False Positive*, INTERCEPT (July 26, 2019, 5:00 AM), https://theintercept.com/2019/07/26/europe-border-control-ai-lie-detector/; Jeff Daniels, *Lie-Detecting Computer Kiosks Equipped with Artificial Intelligence Look like the Future of Border Security*, CNBC (May 15, 2018, 8:17 AM), https://www.cnbc.com/2018/05/15/lie-detectors-with-artificial-intelligence-are-future-of-border-security.html.

2. As shown in more depth below, evidence admitted to impeach witness credibility is described as relevant because it helps jurors or judges learn something about a witness's propensity for truthfulness. At the same time, credibility jurisprudence focuses fact-finders on external qualities such as demeanor and prior convictions, which are poor indicators of truthfulness itself. As the common law has long recognized, these features are instead probative of a witness's *worthiness* of belief. Julia Simon-Kerr, *Credibility by Proxy*, 85 GEO. WASH. L. REV. 152, 207 (2017) [hereinafter Simon-Kerr, *Credibility by Proxy*]; *see* Julia Simon-Kerr, *Uncovering Credibility*, *in* THE OXFORD HANDBOOK OF LAW AND HUMANITIES 583 (Simon Stern et al. eds., 2019) [hereinafter *Uncovering Credibility*]; Julia Simon-Kerr, *Unmasking Demeanor*, 88 GEO. WASH. L. REV. ARGUENDO 158, 161 (2020) [hereinafter *Unmasking Demeanor*].

3. Garland v. Ming Dai, 141 S. Ct. 1669, 1681 (2021) (emphasis added).

4. *See* Xin Dai, Toward a Reputation State: The Social Credit System Project of China 1–2, 14 (June 10, 2018) (unpublished manuscript) (SSRN), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3193577 (describing the social credit program's "overarching theme" of enhancing "lofty moral virtues" across all societal interactions).

prior work.[5] These analogies also track the two focal points of credibility doctrine, offering a first glimpse at how algorithms attempting to replicate that doctrine might be tailored.

China's social credit scoring system is explicitly designed to create a credibility metric.[6] Under a recent central government initiative, local governments are creating a so-called social credit score for their citizens.[7] Unlike a typical U.S. credit score, these scores may take account of a wide range of behavioral inputs, like volunteerism or failing to pay child support.[8] They are made possible in large part by the central government's control of vast amounts of data on its citizens, which it is making accessible to local governments.[9] The central government hopes that these publicly-available scores will shape behavior and influence interpersonal and business practices as people strive to behave in ways that confer credibility in the form of a higher score.[10] Social credit scoring, in other words, will control whether a person is worthy of belief.

While credibility impeachment in U.S. courtrooms cannot claim the same reach, it shares something essential with China's system. Credibility impeachment focuses on external metrics of credibility, most importantly demeanor and prior convictions. In the U.S. legal system, being found worthy of belief is equivalent to having credibility, and credibility itself, like a good social credit score, is an end rather than a means to something else.[11] If witnesses want to be believed, they must conform to the fact-finder's vision of what makes a person believable. Many disparate elements of our evidentiary system reinforce the importance of these external metrics of credibility, cementing credibility's function as a reflexive question of social performance.[12]

5. *See supra* note 2; *see also* Julia Simon-Kerr, Note, *Unchaste and Incredible: The Use of Gendered Conceptions of Honor in Impeachment*, 117 YALE L.J. 1854, 1893–97 (2008) [hereinafter *Unchaste and Incredible*].

6. *See* TRIVIUM CHINA, UNDERSTANDING CHINA'S SOCIAL CREDIT SYSTEM: A BIG-PICTURE LOOK AT SOCIAL CREDIT AS IT APPLIES TO CITIZENS, BUSINESS AND GOVERNMENT *passim* (2019), https://socialcredit.triviumchina.com/wp-content/uploads/2019/09/Understanding-Chinas-Social-Credit-System-Trivium-China-20190923.pdf [hereinafter TRIVIUM REPORT].

7. *Id.* at 3.

8. *Id.* at 11.

9. *Id.* at 10.

10. *Id.* at 3, 24.

11. *See Credibility*, BLACK'S LAW DICTIONARY (11th ed. 2019) ("The quality that makes something (as a witness or some evidence) worthy of belief.").

12. As elaborated in Part III.A, many facets of procedural law work in concert to emphasize how central demeanor is to assessing credibility. For example, the Supreme Court has held that part of the reason for deference to lower court decisions enshrined in Federal Rule of Civil Procedure 52 stems from the fact that "only the trial judge can be aware of the variations in demeanor and tone of voice that bear so heavily on the listener's

By contrast, courtroom credibility assessment is often described as analogous to the project of financial credit scoring – a way to predict or identify risk. As historian Josh Lauer and others have shown, our credit scoring system evolved from one focused on the same markers of worthiness that have long been used in the courtroom to the one familiar today, which uses algorithmic prediction to measure the likelihood of default. [13] Because credit scores were initially based primarily on reputation, they were higher for those who behaved *as if* they were trustworthy and a good bet. [14] But in the early twentieth-century, credit agencies began attempting to disentangle the question of creditworthiness from the broader concept of credibility. [15] Over time, credit scoring agencies refined their algorithms to focus on measuring a borrower's likelihood of default. [16]

As with many risk-prediction algorithms, credit scores reflect the unequal world from which they cull their data. [17] Still, modern credit

understanding of and belief in what is said." Anderson v. City of Bessemer City, 470 U.S. 564, 575 (1985). Similarly, one justification for the prohibition on hearsay evidence is the need for the fact-finder to assess credibility by seeing witnesses in person. *See* KENNETH S. BROUN ET AL., MCCORMICK ON EVIDENCE 589–90 (Robert P. Mosteller ed., 8th ed. 2020) (defining hearsay statements as "statements whose evidentiary value depends upon the credibility of the declarant without the assurances of oath, presence, or cross-examination"). Congress has endorsed this vision of credibility in the Immigration and Nationality Act. *See* 8 U.S.C. § 1158.

13. *See* Josh Lauer, *From Rumor to Written Record: Credit Reporting and the Invention of Financial Identity in Nineteenth-Century America*, 49 TECH. & CULTURE 301, 307 (2008) [hereinafter *From Rumor to Written Record*]; Josh Lauer, *Your Credit Score Isn't a Reflection of Your Moral Character*, SLATE (Nov. 23, 2018, 8:09 AM), https://slate.com/technology/2018/11/dhs-credit-scores-legal-resident-assessment.html [hereinafter *Your Credit Score Isn't a Reflection of Your Moral Character*].

14. *See* Ingrid Jeacle & Eamonn J. Walsh, *From Moral Evaluation to Rationalization: Accounting and the Shifting Technologies of Credit*, 27 ACCT., ORGS. & SOC'Y 737, 743 (2002) (referencing the "preoccupation with character and social standing" associated with early credit reporting, citing to an example where a creditor took into account a family's "characteristic pride in square dealing" to indicate "that any obligation will be faithfully kept to the letter").

15. *See From Rumor to Written Record*, *supra* note 13, at 307 (tracing the evolution of credit-reporting from the watchful neighbor to the establishment of large firms that hired traveling credit reporters); JOSH LAUER, CREDITWORTHY: A HISTORY OF CONSUMER SURVEILLANCE AND FINANCIAL IDENTITY IN AMERICA 250 (2017) [hereinafter CREDITWORTHY] (noting the rise of credit-ranking, with risk models being used to predict "everything from response rates to attrition" by the early 1990s).

16. ROWENA OLEGARIO, THE ENGINE OF ENTERPRISE: CREDIT IN AMERICA 209–11 (2016).

17. *See* Noel Capon, *Credit Scoring Systems: A Critical Analysis*, 46 J. MKTG. 82, 86–87 (1982); Lisa Rice & Deidre Swesnik, *Discriminatory Effects of Credit Scoring on Communities of Color*, 46 SUFFOLK U.L. REV. 935, 936–37 (2013); Barbara Kiviat, *Credit Scoring in the United States*, 21 ECON. SOCIO. 33, 37–38 (2019) ("The U.S.'s shameful history of racial segregation and discrimination looms large in credit markets."); Kaveh Waddell,

scores are now relatively good at predicting the likelihood that a borrower will default on a loan within a given period of time.[18] In this way, financial ability and willingness to pay loans have been unbundled from the larger package of "credibility," and algorithms have homed in on the risk of default, which is itself susceptible to data-driven assessment.[19]

The jurisprudence of credibility shows a similar desire within the legal system for credibility judgments to predict or identify a witness's "propensity for truthfulness."[20] Although this claim for credibility in the

How Algorithms Can Bring Down Minorities' Credit Scores, ATLANTIC (Dec. 2, 2016), https://www.theatlantic.com/technology/archive/2016/12/how-algorithms-can-bring-down-minorities-credit-scores/509333/ (describing the "guilt by association" problem of taking family and friend's actions into account in credit scoring, disproportionately affecting low-income communities); Michelle Singletary, *Credit Scores Are Supposed to be Race-Neutral. That's Impossible.*, WASH. POST (Oct. 16, 2020), https://www.washingtonpost.com/business/2020/10/16/how-race-affects-your-credit-score/ ("[F]actors that are included or excluded in the algorithms used to create a credit score can have the same effect as lending decisions made by prejudiced White loan officers."); Jennifer Streaks, *Black Families Have 10 Times Less Wealth than Whites and the Gap is Widening—Here's Why*, CNBC (May 18, 2018, 1:04 PM), https://www.cnbc.com/2018/05/18/credit-inequality-contributes-to-the-racial-wealth-gap.html (noting the modern problem of "credit invisibility" which eliminates Black, Latino, and young Americans from credit scoring models). These problems also persist in the AI context. *See, e.g.*, Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1027, 1040–43 (2017) (calling for affirmative action in algorithmic design in order to combat "discrimination lurking in the data" in a "society where discrimination affects opportunities in innumerable ways").

18. Will Dobbie et al., *Bad Credit, No Problem? Credit and Labor Market Consequences of Bad Credit Reports*, 75 J. FIN. 2377, 2392–96 (2020).

19. While credit scores are relatively effective at predicting the risk of default, they have increasingly been used for purposes well beyond the one for which they were designed, namely lending. For example, in 2019 the Trump administration made credit scores a factor in immigration decisions on the premise that they are relevant to whether an immigrant might become a "public charge." AnnaMaria Andriotis, *New Trump Administration Rule Will Look at Immigrants' Credit Histories*, WALL ST. J. (Aug. 16, 2019, 12:46 PM), https://www.wsj.com/articles/new-trump-administration-rule-will-look-at-immigrants-credit-histories-11565973971. This is one way in which the credit score itself has been folded back into the larger enterprise of using credibility to constitute some desired end in and of itself, one which must be performed in order to reap social benefits, such as a green card. *See Your Credit Score Isn't a Reflection of Your Moral Character*, *supra* note 13 (suggesting that credit scores are only useful proxies for an immigrant's likelihood of becoming a public charge "if one believes credit scores reveal something about a person's character").

20. ROGER PARK & TOM LININGER, THE NEW WIGMORE. A TREATISE ON EVIDENCE: IMPEACHMENT AND REHABILITATION § 3.4 (1st ed. 2021) (suggesting that "prior convictions may add little to the assessment of the credibility of an accused" because "information about prior misconduct says little about his propensity to lie in the case at bar"). *See also, e.g.*, United States v. Lipscomb, 702 F.2d 1049, 1062 (D.C. Cir. 1983) (interpreting F.R.E. 609 by discussing which crimes bear on a witness's credibility by showing a "propensity for truthfulness" and concluding that "Congress believed that all felonies have some probative value on the issue of credibility); United States v. Fearwell, 595 F.2d 771, 776–77 (D.C. Cir. 1978) (holding that "like multifarious others of a similar nature, [the crime of petit larceny] simply has no bearing whatever on the "accused's propensity to testify truthfully" and is

legal realm has never been subjected to study or data-driven assessment, courts and commentators regularly treat credibility as a quasi-scientific indicator of the likelihood that a witness will lie.[21]

In grappling with credibility in the age of algorithms, a first intervention is to understand the choices presented by current doctrine. What are the implications of embracing either a risk-predictive model of credibility focused on a propensity for lying or an "interventionist" model focused on socially-constructed worthiness? [22] As the predominant justification for current credibility jurisprudence, predicting a witness's propensity to lie will be a natural focus for developers of evidentiary risk-prediction tools. Alternatively, as with China's social credit scoring, the system could assign positive or negative scores based on algorithmic assessments of features such as a witness's demeanor, lack of prior convictions, reputation for untruth or apparent candor. In this regime, those shaping the algorithm could create what it is to be a trustworthy or credible person in the eyes of the law. If neither approach seems promising, that has implications for current jurisprudence as well as for the credibility we might hope to see in an algorithmic future. Either way, this study is one way to safeguard against a future in which the terms of algorithmic engagement with credibility in the law are set by the creators of such algorithms in accordance with their interests.[23]

A final consideration is why credibility practice should be a focus when so many structural features of our criminal or civil legal systems, such as money bail, punitive damages, or incarceration itself, are in urgent need of reform. [24] The answer is simple: credibility matters.

thus inadmissible "for the purpose of attacking the credibility of a witness"); United States v. Estrada, 430 F.3d 606, 617 (2d Cir. 2005) (holding "all Rule 609(a)(1) felonies are not equally probative of credibility but . . . many are significantly probative of a witness's propensity for truthfulness"). Leading evidence treatises discuss the question of credibility using the idea of a propensity for truth. *See* PARK & LININGER, *supra* note 20, at 125–27, 150.

21. *Credibility by Proxy*, *supra* note 2, at 207–12.

22. Daithí Mac Síthigh & Mathias Siems, *The Chinese Social Credit System: A Model for Other Countries?*, 82 MOD. L. REV. 1034, 1036, 1054 (2019).

23. For example, Emily Berman argues that translating policy goals into algorithmic tools "can only be described as policymaking," and points out that in the context of COMPAS, the commercial pretrial risk assessment tool, "the policymakers were a team of programmers engaged in the for-profit venture that developed" the tool. Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. REV. 1277, 1280–81, 1329 (2018).

24. Abolitionist groups, such as the Movement for Black Lives, argue, in fact, that reform is not enough and that the system should be reconceptualized and restructured. *See* Amna A. Akbar, *Toward a Radical Imagination of Law*, 93 N.Y.U. L. REV. 405, 460–61 (2018). They call for abolition of the carceral state and for putting "an end to police in schools; mass surveillance by police; privatization of police; capital punishment; money bail,

Mistaken credibility judgments may convict the innocent, assign fault where none is due, rob a domestic abuse victim of a remedy, or deny asylum to someone suffering persecution. Credibility judgments are made every day in our legal system, and so long as we continue to offer state-sponsored or formal dispute resolution they will continue to matter.

This Article proceeds in four Parts. The second establishes why we should be thinking about algorithms for credibility at all. The third shows how, as a doctrinal matter, the law of credibility is focused on both worthiness of belief and a witness's propensity for truth. It argues that social credit scoring in China offers a vision of a worthiness-centered approach to algorithmic credibility assessment while American credit-scoring provides an example of a path that uses algorithms to focus on risk-prediction in the credibility arena. The final Part unpacks the distinct promise and perils highlighted by these analogies for each vision of credibility in an age of algorithms.

II. TAKING ALGORITHMIC CREDIBILITY SERIOUSLY

Pressure to automate lie detection is not new. Indeed, mechanized approaches to lie detection have been knocking on the courtroom doors for at least a century. In 1922, defense attorneys for James A. Frye, who had confessed to murder and then recanted, attempted to introduce the testimony of lawyer and psychologist William Moulton Marston.[25] As Jill Lepore reports in her account of the *Frye* case, Marston hoped to offer an expert opinion that Frye's recantation was truthful based on the results of a device that purported to identify lies based on a subjects' systolic blood pressure.[26] The trial court rejected this precursor to the polygraph as being insufficiently accepted in the scientific community.[27] While polygraph evidence is now admissible in some state court systems and used in police interrogations,[28] among other applications, it is still

fines, and fees; the use of criminal history as relevant to determining access to housing," among other things. *Id.* at 461.

25. Jill Lepore, *On Evidence: Proving* Frye *as a Matter of Law, Science, and History*, 124 YALE L.J. 1092, 1092, 1120–1122, 1126 (2015).

26. *See id.* at 1126, 1131–32 (reproducing partial trial transcript).

27. *Id.* The notion that credibility requires observation of the witness's demeanor is part of the story of the polygraph's place in the legal system. In upholding a bar on the use of polygraph evidence in military court martials, the Supreme Court suggested that the polygraph might "diminish the jury's role in making credibility determinations." United States v. Scheffer, 523 U.S. 303, 313–14 (1998).

28. *See, e.g.*, Saul M. Kassin et al., *Police-Induced Confessions: Risk Factors and Recommendations*, 34 L. & HUM. BEHAV. 3, 28–29 (2010) (describing police use of polygraph results to induce suspects to confess); State v. Harrison, 7 P.3d 478, 489 (N.M. 2000) ("In New Mexico, the trial court has discretion to admit results of polygraph tests into evidence

deemed insufficiently reliable to be admitted in most United States' courts of law.[29]

In recent years, the field of scientific—or pseudo-scientific—lie-detection has been reinvigorated by the advent of artificial intelligence. Several research groups have marketed "artificial intelligence-driven" lie detection tools to border patrol services and airport security agencies, among other places.[30] One such "deception detection tool" is named the Avatar, or "Automated Virtual Agent for Truth Assessments in Real-time".[31] The Avatar is a virtual border guard that asks a series of scripted questions to travelers.[32] Its creators assert that it can more accurately identify those who pose a danger than human guards.[33] It makes these identifications by looking for "deception signals" or "microexpressions" that are theoretically triggered by the cognitive load of telling lies.[34] The Avatar then assesses whether the traveler is telling the truth.[35] Proponents of these tools make the unsubstantiated claim that the algorithms can identify "micro" clues to deception that humans can neither control nor identify.[36]

A similar program called the "virtual policeman" was being tested at border checkpoints in Europe in the summer of 2019.[37] Its creators suggested that its proprietary AI analysis of a subject's "facial expressions, gaze, and posture" can pick up on "micro gestures" that connote deception.[38] The Avatar's creators claim between eighty and eighty-five percent accuracy,[39] while one tiny study offered a seventy-five

if certain conditions, designed to ensure the accuracy and reliability of the test results, are met." (quoting State v. Sanders, 872 P.2d 870, 877 (N.M. 1994))).

29. *See Scheffer*, 523 U.S. at 310–11 ("Most States maintain *per se* rules excluding polygraph evidence.").

30. *See* Hodgson, *supra* note 1; Gallagher & Jona, *supra* note 1; Devlin, *supra* note 1; *see also* Jake Bittle, *Lie Detectors Have Always Been Suspect. AI Has Made the Problem Worse*, MIT TECH. REV. (Mar. 13, 2020), https://www.technologyreview.com/2020/03/13/905323/ai-lie-detectors-polygraph-silent-talker-iborderctrl-converus-neuroid/ (describing recent commercialization of lie detection algorithms for use by border patrol agencies and others).

31. Hodgson, *supra* note 1.

32. *Id.*

33. *Id.*

34. *Id.*

35. *Id.*

36. *See* Bittle, *supra* note 30 (quoting lie-detection algorithm developer's description of program development that involved translating "high level cues into our own set of micro gestures and trained AI components to recombine them into meaningful indicative patterns").

37. Gallagher & Jona, *supra* note 1.

38. *Id.*

39. Hodgson, *supra* note 1.

percent accuracy rate for the virtual policeman.[40] Both of these tools were developed by academic researchers using public funding with an eye to government contracts.[41] In Spain, police have already used a related tool—an algorithm that could allegedly detect insurance fraud when people falsely claimed to have had their smartphones stolen.[42] The tool analyzed statements given to law enforcement by looking for "suspicious wording."[43] Private corporations are developing similar algorithmic products as aids to hiring.[44] To give one example, Unilever recently announced that it was using software to analyze videos of prospective employee interviews looking for traits "considered to be correlated with job success."[45] As these instances show, governments and for-profit corporations are already harnessing big data to predict lies or other personality failings.

Researchers have not ignored the courtroom as a space ripe for the application of AI-based lie-detection. Scientists at Dartmouth and the University of Maryland developed a tool they called "DARE" specifically for use in courtrooms.[46] Like the "virtual policeman," this system is also trained to identify so-called "micro-expressions" that are supposedly indicative of lying.[47] Importantly, though, its creators suggested that if the system included more than just visual information, "deception prediction [could] be further improved."[48]

Indeed, incorporating big data is an obvious next step in the evolution of algorithmic character or deception prediction tools. As Aziz Huq has argued in reference to sentencing and bail decisions, "[n]ewer tools will combine powerful computational instruments with large volumes of data to enable prediction."[49] In the United States, companies like Facebook, Google, and Amazon have been collecting and selling consumer data for

40. Gallagher & Jona, *supra* note 1.

41. *See* Hodgson, *supra* note 1; Bittle, *supra* note 30; Gallagher & Jona, *supra* note 1.

42. *See* Editorial, *False Testimony: A Lie-Detection System Being Used by Spanish Police Highlights Concerns About Algorithms.*, 557 NATURE 612, 612 (2018).

43. *Id.*

44. *See, e.g.*, Devlin, *supra* note 1.

45. *Id.*

46. Alexandra Richards, *Artificial Intelligence System Could Be Used to Detect if People Are Lying in Court*, EVENING STANDARD (Dec. 20, 2017), https://www.standard.co.uk/news/world/artificial-intelligence-system-could-be-used-to-detect-if-people-are-lying-in-court-a3724221.html.

47. *Id.*; Zhe Wu et al., *Deception Detection in Videos*, *in* THE THIRTY-SECOND AAAI CONF. ON A.I. 1695, 1695 (2018).

48. Wu et al., *supra* note 47, at 1701.

49. Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1062–63 (2019).

years.[50] These companies know everything about us from who we call, to the comments we leave on YouTube videos, to what we buy.[51] They may share information with advertisers, researchers, data aggregators, and law enforcement agencies.[52] As one author found in 2017, $23 paid to a data aggregation company could buy access to "a person's contact information and age, organizational memberships, links to social media accounts, business interests, and known associates."[53] While states are increasingly looking to regulate companies' data collection practices, Congress has thus far failed to intervene in this area.[54] Thus, as it stands, vast troves of consumer information are collected daily and available commercially with only minimal policy oversight. With access to this type of Big Data, algorithmic tools could be built that would consider a mind-boggling array of personal information, ranging from political affiliations, credit scores, and group memberships to consumer product preferences.[55]

While the legal system's response to such tools might mirror its long-held skepticism of the polygraph, this is a different time and algorithms are a different tool.[56] Prediction tools analogous to the lie-detection algorithms described above are already in use in the legal system. So-called "risk-assessment algorithms" are now being used to quantify the risk of re-offending in the charging, sentencing, and bail contexts,[57]

50. *See* Caitlin Dewey, *98 Personal Data Points That Facebook Uses to Target Ads to You*, WASH. POST (Aug. 19, 2016), https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/ (describing ways in which Facebook collects and sells personal information for advertising); Aliza Vigderman & Gabe Turner, *The Data Big Tech Companies Have on You,* SECURITY.ORG, https://www.security.org/resources/data-tech-companies-have/ (last updated Aug. 23, 2021) (describing data collected on consumers by Google, Amazon, and Facebook).

51. *See* Vigderman & Turner, *supra* note 50.

52. *Id.* (describing third-party sharing policies of Amazon, Facebook, and Google).

53. Theodore Rostow, *What Happens When an Acquaintance Buys Your Data?: A New Privacy Harm in the Age of Data Brokers*, 34 YALE J. ON REGUL. 667, 668 (2017).

54. *See* Rita W. Garry, *Where is the Consumer in Consumer Privacy Legislation?*, LAW.COM (June 25, 2021, 12:02 PM), https://www.law.com/2021/06/25/where-is-the-consumer-in-consumer-privacy-legislation/.

55. I am not the first to see these developments headed to the courtroom. *See, e.g.*, Bittle, *supra* note 30 (suggesting that widespread adoption of deception-prediction algorithms may lead to their use in court, or in other contexts, such as "before renting a car or taking out a loan").

56. Indeed, the "age of algorithms is upon us." *Bias In, Bias Out*, *supra* note 1, at 2221. Or, as Professor Huq writes, "[e]ven if machine-learning and deep-learning tools are not now omnipresent, they will be soon." Huq, *supra* note 49, at 1068.

57. *See* Ngozi Okidegbe, *When They Hear Us: Race, Algorithms and the Practice of Criminal Law*, 29 KAN. J.L. & PUB. POL'Y 329, 329–32 (2020) (both critiquing and finding promise in criminal law's use of algorithms to set bail, to inform sentencing, and in parole determinations); *see also Bias In, Bias Out*, *supra* note 1, at 2222.

where they are used to make life-altering determinations for criminal defendants. [58] Much like China's various social credit scoring experiments, these algorithms take account both of previous legal entanglements and of other information, like marital status or place of residence.[59] Also, as in China, United States state and local governments are "integrating their data systems," making it possible to access and analyze data on individuals ranging from their earnings to their visits to public hospitals to their contacts with the DMV.[60]

With the growing embrace by both the legal system and the corporate world of algorithms as productive tools for guiding human decision-makers, proponents of algorithmic measures of credibility can be expected to further develop tools for use in the courtroom. And if we are unclear what credibility does or should mean in the courtroom, it is the AI developers who will answer those questions as they bring their own preconceptions to bear on the creation of algorithms.

III. WHICH CREDIBILITY?

There are two primary ways to understand credibility in the law. The first is the Black's Law Dictionary definition of credibility as "worth[iness] of belief."[61] Worthiness is signaled by factors including an appropriate demeanor or the absence of prior convictions. The second is the doctrinal insistence that credibility judgments measure a witness's propensity for truthfulness.[62] At present, these concepts coexist quietly but illogically in evidence law. Today's credibility doctrine simply points the fact-finder's attention to markers of a witness's worthiness of belief while claiming at the same time that the markers are predictive of truthfulness.

This Part elaborates this point and then explores two algorithmic credibility tools that implement predictive and worthiness-focused visions of credibility. The first is China's social credit system, which shapes and creates credibility in a way that I argue is analogous to credibility's actual function in today's legal system. The second is the

58. *See* Erin Collins, *Punishing Risk*, 107 GEO. L.J. 57, 86 (2018) (describing sentencing consequences for defendants of actuarial risk prediction tools, some of which directly contradict courts' directives for appropriate use of these tools).

59. Stevenson, *supra* note 1, at 315 (describing socio-economic factors along with criminal history, age, and gender as common inputs to criminal justice risk assessment tools).

60. Sandra G. Mayson, *Dangerous Defendants*, 127 YALE L.J. 490, 552 (2018).

61. *Credibility*, *supra* note 11; *see also* Garland v. Ming Dai, 141 S. Ct. 1669, 1681 (2021) (quoting Black's definition).

62. *See supra* note 20 and accompanying text.

United States credit score, which originated as a fuzzy measure of social worthiness like legal credibility but which has since been refined as a predictive tool focused on a borrower's risk of default.[63] The credit score is analogous to an algorithm with real predictive capacity on the narrow question of a witness's propensity to lie.

While there are other ways that we might re-envision credibility, some of which I hope to explore in future work, conducting this particular thought experiment on how algorithms might replicate the function—real or imagined—that the law currently ascribes to credibility judgments has three major benefits. First, it trains attention on the two algorithmic approaches that would have the greatest doctrinal support, a logical point of departure given the law's fundamental conservatism and commitment to incremental change. Second, it shows beyond question that despite their elision in the law these two approaches are distinct. And finally, because it helps us see them as distinct, it raises important questions about the validity of these metrics in today's law and which, if either, version of credibility might we want to pursue in an algorithmic future.

A. Credibility Jurisprudence

"A judgment about credibility . . . measures the truthfulness of the speaker or the likelihood that what he says is true."[64] This claim of predictive capacity, made in this instance by the D.C. Circuit, is the standard explanation of the work credibility impeachment does in the courtroom.[65]

63. *See supra* notes 13–15 and accompanying text.

64. Latif v. Obama, 677 F.3d 1175, 1190 (D.C. Cir. 2011).

65. *Id.*; *see also* Urooj v. Holder, 734 F.3d 1075, 1078 (9th Cir. 2013) (noting that "impeachment" gets at "the basic aim of all credibility rules: to admit evidence that enables the trier of fact to determine whether or not the witness is telling the truth" (citing 4 JACK B. WEINSTEIN & MARGARET A. BERGER, WEINSTEIN'S FEDERAL EVIDENCE § 607.03 (Joseph M. McLaughlin, ed., Matthew Bender 2d ed. 1997))); State v. Russell, 625 S.W.2d 138, 141 (Mo. 1981) ("The purpose of impeachment is to impair or destroy a witness's credibility or render questionable the truth of his particular testimony."); Holland v. French, 848 S.E.2d 274, 282 (N.C. Ct. App. 2020) ("Impeachment evidence has been defined as evidence used to undermine a witness's credibility, with *any* circumstance tending to show a defect in the witness's *perception, memory, narration or veracity* relevant to this purpose." (quoting State v. Gettys, 777 S.E.2d 351, 356 (N.C. Ct. App. 2015))); State v. Brown, 233 A.3d 1258, 1265 (Conn. App. Ct. 2020) ("The purpose of impeachment is to undermine the credibility of a witness so that the trier will disbelieve him and disregard his testimony." (quoting State v. Valentine, 692 A.2d 727, 737 (Conn. 1997))); Sanders v. Buchanan, 407 F.2d 161, 162 (10th Cir. 1969) (writing that impeachment is necessary for witnesses because "[t]he credibility of the witness is always relevant in the search for truth" (quoting Creekmore v. Crossno, 259 F.2d 697, 698 (10th Cir. 1958))).

Despite such judicial breeziness, it is difficult to be precise about what it means to have credibility in the eyes of the law. All evidence in a case, to an extent, will reflect in some way on the credibility of witnesses.[66] Indeed, the sixteenth-century treatise cited by the D.C. Circuit to support its assertion that credibility judgments measure truthfulness offers a more realistic portrait of the concept as it functions in the law.[67] The treatise suggests that "[t]hings are made credible, either by the known condition and quality of the utterer, or by the manifest likelihood of truth which they have in themselves."[68] This breaks down into roughly two categories—information about the speaker and narrative plausibility, meaning what we think of the story the witness is telling.

Judges have debated whether there is a distinction between narrative plausibility and credibility because the two concepts are in some ways intertwined.[69] The story of a witness who seems credible may be read as more plausible.[70] Conversely, if her story seems implausible,

66. Reflecting the broad scope of what may influence our perceptions of credibility, the Supreme Court has explained that "[d]ocuments or objective evidence may contradict the witness' story; or the story itself may be so internally inconsistent or implausible on its face that a reasonable factfinder would not credit it." Anderson v. City of Bessemer City, 470 U.S. 564, 575 (1985).

67. *Latif*, 677 F.3d at 1190; RICHARD HOOKER, THE LAWS OF ECCLESIASTICAL POLITY 151–52 (George Edelen ed., Harvard Univ. Press 1977) (1594). The Hooker treatise provided one of the most influential English-language expositions of a theory of natural law of its day. *See generally* Torrance Kirby, *The 'Sundrie Waies of Wisdom': Richard Hooker on the Authority of Scripture and Reason*, *in* THE OXFORD HANDBOOK OF THE BIBLE IN EARLY MODERN ENGLAND, C. 1530–1700, at 164–75 (Kevin Killeen, Helen Smith, and Rachel Willie eds., 2015).

68. *Latif*, 677 F.3d at 1190 (quoting RICHARD HOOKER, THE LAWS OF ECCLESIASTICAL POLITY 151–52 (George Edelen ed., Harvard Univ. Press 1977) (1594)).

69. *See, e.g.*, *Latif*, 677 F.3d at 1190–91, 1219–20, 1226. In *Latif v. Obama*, the contentious Habeas case quoted at the outset of this Section, for example, the majority and dissenting judges argued about whether there was, in fact, a distinction between narrative plausibility and credibility. *Id.* The majority had held that the district court had made a plausibility finding about the petitioner's story, but not a credibility finding. *Id.* at 1190–91. Judge Tatel in dissent accused the majority of failing to "take seriously the notion that district courts are better at . . . determining credibility." *Id.* at 1226 (Tatel, J., dissenting). Judge Tatel wrote that the district court's finding was not limited to narrative plausibility and should have been understood to cover all aspects of credibility. *Id.* at 1219–20. The petitioner's account was "convincing enough, plausible enough, consistent enough, and corroborated enough" to overcome the government's attack on Latif's credibility. *Id.* at 1220.

70. The focus on plausibility has ancient roots. Humanist Rodolphus Agricola famously identified a triad of qualities essential to a "probable account." Paula Olmos, *ISSA Proceedings 2014—Story Credibility in Narrative Arguments*, ROZENBERG Q., https://rozenbergquarterly.com/issa-proceedings-2014-story-credibility-in-narrative-arguments/ (last accessed Oct. 28, 2021). The account should be sufficiently detailed, it should be free from contradiction, and it should be "consistent with how things are." *Id.*

she may be viewed as less credible. Nonetheless, narrative plausibility, which focuses on the story being told, is conceptually distinct from credibility, which considers features of the witness. Similarly, information about a witness's bias, while it may inform a fact-finder's belief about whether the witness is being truthful, is intertwined with the narrative of the case and distinguishable from acontextual information that goes to whether the speaker has credibility.[71]

This Article focuses on algorithms targeted to credibility in its narrower sense—information about speakers rather than their stories or motivations. In the United States legal system, two forms of information about witnesses have been given preeminent importance. The first is demeanor, or the outward appearance of witnesses. The second is whether they have prior convictions. The following subparts briefly canvass these two main foci of credibility doctrine. Like the social-credit scoring system described later in this Part, this body of law creates a performative and circular understanding of credibility in which having credibility depends on being able to meet legally entrenched expectations of how a believable, or worthy, person should appear or act. By emphasizing the "condition and quality" of witnesses who come before the legal system in this way, credibility jurisprudence establishes and reinforces credibility as a measure of a witness's worthiness of belief. At the same time, courts hold out this system as one that offers a probabilistic assessment of whether a witness is lying.[72]

1. Demeanor

For centuries, the common law of evidence has maintained that the demeanor, or outward bearing, of witnesses is central to credibility.[73] As

71. This is one reason that bias is generally treated differently from credibility under both evidentiary and constitutional regimes. *See, e.g.*, Boggs v. Collins, 226 F.3d 728, 736 (6th Cir. 2000) ("[T]he *Davis* Court distinguished between a 'general attack' on the credibility of a witness—in which the cross-examiner 'intends to afford the jury a basis to infer that the witness' character is such that he would be less likely than the average trustworthy citizen to be truthful in his testimony'—and a more particular attack on credibility 'directed toward revealing possible biases, prejudices, or ulterior motives as they may relate directly to issues or personalities in the case at hand.'" (quoting Davis v. Alaska, 415 U.S. 308, 316 (1974))).

72. *See Credibility by Proxy*, *supra* note 2, at 208–12. As I have shown elsewhere, there is no empirical evidence to support the ability to predict courtroom lies from past behavior in the manner suggested by credibility impeachment doctrine. *See id.*

73. *Unmasking Demeanor*, *supra* note 2, at 161–62 (arguing that mask-wearing in the courtroom affords an opportunity to reevaluate the role of demeanor in legal credibility determinations); *see also* Miller v. Fenton, 474 U.S. 104, 114 (1985) (noting that when an "issue involves the credibility of witnesses" it is understood to "turn[] largely on an evaluation of demeanor"); Mark W. Bennett, *Unspringing the Witness Memory and*

Judge Learned Hand put it, "the carriage, behavior, bearing, manner and appearance of a witness—in short, his 'demeanor'—is a part of the evidence."[74] In other words, fact-finders must, should, and do rely on external cues to evaluate credibility.

In addition to common law doctrine, procedural rules reinforce the importance of demeanor. For example, the rule against hearsay privileges live testimony because testimonial reliability can only be evaluated with reference to demeanor.[75] The "evidentiary value" of hearsay statements "depends upon the credibility of the declarant," which cannot be assessed "without the assurances of oath, presence, or cross-examination."[76]

The focus on demeanor is also reflected in longstanding rules and precedent that protect credibility judgments from appellate second-guessing on the theory that an appellate judge could not possibly assess the cues that make up the witness's demeanor. For example, the Supreme Court has explained that the clear error standard of appellate review in Rule 52(a) of the Federal Rules of Civil Procedure[77] is appropriate because the trial judge is the only one privy to "variations in demeanor and tone of voice that bear so heavily on the listener's understanding of and belief in what is said."[78] This, in turn, means that appellate courts will "overturn credibility determinations only where a witness's testimony is impossible under the laws of nature or incredible as a matter of law—an extraordinarily high standard."[79]

Demeanor Trap: What Every Judge and Juror Needs to Know About Cognitive Psychology and Witness Credibility, 64 AM. U.L. REV. 1331, 1338 (2015) ("The American justice system's longstanding reliance on evaluating a witness's credibility based upon his demeanor, so-called 'demeanor evidence,' has long been a pillar in jurisprudence.").

74. Dyer v. MacDougall, 201 F.2d 265, 268–69 (2d Cir. 1952).

75. *See* Laurence H. Tribe, *Triangulating Hearsay*, 87 HARV. L. REV. 957, 958 (1974) (explaining hearsay as making it difficult to "forg[e] a reliable chain of inferences" when an "utterance is not one made in court, under oath, by a person whose demeanor at the time is witnessed by the trier, and under circumstances permitting immediate cross-examination by counsel"); *see also* BROUN ET AL., *supra* note 12, at 589–90.

76. BROUN ET AL., *supra* note 12, at 589–90.

77. FED. R. CIV. P. 52(a)(6).

78. Anderson v. City of Bessemer City, 470 U.S. 564, 575 (1985).

79. Bennett, *supra* note 73, at 1350; Lukaneva v. Levy Rests. at McCormick Place, No. 05 C 6159, 2006 WL 1823169, at *5 (N.D. Ill. June 29, 2006) (stating that credibility determinations based on demeanor are "usually insulated from appellate review"); United States v. Beverly, 5 F.3d 633, 642 (2d Cir. 1993) ("We have recognized that assessing the credibility of witnesses is distinctly the province of the district court, and we will not lightly overturn such assessments."); United States v. Isaacson, 752 F.3d 1291, 1304 (11th Cir. 2014) ("We will upset a jury's decision to credit a witness's testimony only in the rare circumstance that the testimony is incredible as a matter of law." (citing United States v. Calderon, 127 F.3d 1314, 1325 (11th Cir. 1997))); United States v. Osum, 943 F.2d 1394, 1405 (5th Cir. 1991) ("[T]he jury is the ultimate arbiter of the credibility of a witness;

The cues that trial judges are thought to rely on when evaluating demeanor are telling. They include "facial expressions, eye contact, attitude, body language, length of pauses, hesitation, sincerity, gestures, candor, tone of voice, expression, dress, [and] grooming habits." [80] Although other factors, such as inconsistencies, responsiveness, or even some personal attributes might theoretically be assessed by a reviewing judge, trial and appellate courts set a high bar for parsing these out from demeanor-based credibility judgments.[81]

Juries are also instructed to focus on demeanor. For example, one leading treatise on federal jury instructions suggests the following instruction on the general topic of witness credibility:

> In making your assessment of that witness[,] you should carefully scrutinize all of the testimony given by that witness, the circumstances under which each witness has testified, and all of the other evidence which tends to show whether a witness, in your opinion, is *worthy of belief*. Consider each witness's intelligence, motive to falsify, state of mind, and *appearance and manner* while on the witness stand.[82]

testimony generally should not be declared incredible as a matter of law unless it asserts facts that the witness physically could not have observed or events that could not have occurred under the laws of nature."); United States v. Terry, 572 F.3d 430, 435 (7th Cir. 2009) ("Staring at the pages of a cold record, we are in no position to reassess the credibility of the sole witness who appeared at the suppression hearing."); United States v. Cabrera-Beltran, 660 F.3d 742, 754 (4th Cir. 2011) ("The jury has already assessed the credibility of the witnesses, and this court cannot do so on appeal."); United States v. Madison, 863 F.3d 1001, 1005 (8th Cir. 2017) ("[A] district court's credibility determinations are 'virtually unreviewable on appeal.'" (quoting United States v. Symonds, 260 F.3d 934, 936 (8th Cir. 2001))). For more examples, see Bennett, *supra* note 73, at 1350 n.108.

80. Bennett, *supra* note 73, at 1338.

81. *See, e.g.*, United States v. Crenshaw, 359 F.3d 977, 990–91 (8th Cir. 2004) ("[I]n considering the defendants' attacks on the testimony of the witnesses against them, we pay particular attention to whether other evidence supports the testimony in question, whether the alleged inconsistencies are collateral or central to the elements of the crime, whether the inconsistencies are amenable to explanation, and whether they have been exposed to the jury."); United States v. Valdez, 15 F. App'x 530, 534 (9th Cir. 2001) ("In the absence of any inherent inconsistencies or objective evidence to contradict the witnesses' testimony, we will not disregard testimony that supports the jury's verdict."); Chuan Feng Yu v. Sessions, 695 F. App'x 8, 11 (2d Cir. 2017) ("Given the initial lack of clarity in the question, this single example of a lack of responsiveness does not support the IJ's conclusion that Yu was testifying from a script rather than from actual memory or that his demeanor alone was sufficient grounds for the adverse credibility determination . . . .").

82. 1A KEVIN F. O'MALLEY ET AL., FEDERAL JURY PRACTICE AND INSTRUCTIONS § 15:01 (6th ed. 2021) (emphasis added).

This instruction makes explicit that the task of the fact-finder is to decide whether witnesses are *worthy* of belief, in part with reference to their external features.

Some jurisdictions offer a similar instruction but with an important difference. For example, the Sixth Circuit's Pattern Jury Instructions connect appearance not with worthiness but with honesty, explicitly trying to substitute one for the other.[83] Among other questions, they ask jurors to consider: "Did the witness appear honest? Or did the witness appear to be lying?"[84] Worthiness or appearance is still the focus of the credibility inquiry but in the guise of assessing truthfulness.[85]

In its overarching emphasis on demeanor, the law has adopted a cultural assumption that inner qualities manifest themselves externally.[86] Yet, as researchers in many fields have found, a focus on visual cues does not make us better at distinguishing lies from truth.[87] Indeed, the more we focus on superficial, nonverbal cues, the worse we may become at spotting actual untruths.[88] Such an emphasis may also make us more prone to suspect our interlocutors are lying.[89]

In a reflection of the degree to which armchair psychology governs this area of law, Judge Jerome Frank once aptly described credibility assessments as "un-ruly" in the sense that they "do not lend themselves to formulations in terms of rules.'"[90] Yet, assessing demeanor, Judge

83. SIXTH CIRCUIT COMMITTEE ON PATTERN CRIMINAL JURY INSTRUCTIONS, PATTERN CRIMINAL JURY INSTRUCTIONS 1.07(2)(D) (2019).

84. *Id.* Other jurisdictions share similar instructions. For example, the Eleventh Circuit asks jurors, "Did the witness impress you as one who was telling the truth?" JUDICIAL COUNCIL OF THE UNITED STATES ELEVENTH JUDICIAL CIRCUIT, ELEVENTH CIRCUIT PATTERN JURY INSTRUCTIONS (CRIMINAL CASES) B5 (2020). Some states, such as New Jersey in its Model Civil Jury Instructions, suggest that demeanor is a factor that jurors *may* consider, establishing that demeanor might be part of a credibility assessment without explicitly prioritizing it. COMMITTEE ON MODEL CIVIL JURY CHARGES, MODEL CIVIL JURY CHARGES 1.12L (1998).

85. It is worth noting that truthfulness can be further broken down into honesty and veracity. A person can be honest in the sense of believing her words to be true while lacking veracity because she has an inaccurate or modified memory of an event. Jurisprudence on demeanor does not tend to distinguish between these two, though at times, as in the example above, a focus on whether the witness "appears to be lying" seems to ignore the possibility that the witness could erroneously believe herself to be telling the truth.

86. Psychology researchers have found, for example, that in North American cultures people expect a person's outward bearing to reflect "reality without substantial discrepancy," even demanding that one's "core identity" be aligned with outward appearances. Albert Lee et al., *Fear Goliath or David? Inferring Competence from Demeanor Across Cultures*, 46 PERSONALITY & SOC. PSYCH. BULL. 1074, 1075 (2020).

87. *See Unmasking Demeanor*, *supra* note 2, at 166–67.

88. *See id.*

89. *See id.* at 167.

90. NLRB v. Dinion Coil Co., 201 F.2d 484, 488–89 (2d Cir. 1952).

Frank observed, is "one of the best guides available" to a witness's reliability,[91] a "preeminently important [judicial] power," one that can be "acquired only by experience" and that depends on the judge's "own nature and acquired sagacity."[92] At the same time, he acknowledged, such judgments cannot "be independently tested."[93] One might expect this built-in imperviousness to verification to have elicited attention from courts or commentators, but that attention has been rare.[94]

One notable exception is found in an opinion by Judge Posner in a 2005 asylum case, *Djouma v. Gonzales*.[95] In *Djouma*, Judge Posner wrote that cultural differences would make it "difficult for the immigration judge to 'read'" the demeanor of asylum applicants when assessing credibility.[96] To solve this problem, Judge Posner suggested that studies of the behavior of asylum applicants would help judges better evaluate the credibility of their claims.[97] "Without such systematic evidence," he explained, "immigration judges are likely to continue grasping at straws," such as "minor contradictions[,]" and the "patterns of behavior that would indeed be anomalous in the conditions prevailing in the United States but may not be" in other countries.[98]

91. Broad. Music, Inc. v. Havana Madrid Rest. Corp., 175 F.2d 77, 80 (2d Cir. 1949).

92. *NLRB*, 201 F.2d at 489 (citing JAMES FITZJAMES STEPHEN, THE INDIAN EVIDENCE ACT, WITH AN INTRODUCTION ON THE PRINCIPLES OF JUDICIAL EVIDENCE 41–43 (1872)).

93. *Id.*

94. In *United States v. Dean*, for example, the court explained its deference to the lower court's credibility determination as follows:

> [W]e will uphold the district court's choice of whom to believe unless the court credited exceedingly improbable testimony because the trial court is in the best position to make that judgment. As we have said, "[w]e do not second-guess the sentencing judge's credibility determinations because he or she has had the best opportunity to observe the verbal and non-verbal behavior of the witnesses focusing on the subject's reactions and responses to the interrogatories, their facial expressions, attitudes, tone of voice, eye contact, posture and body movements, as well as the confused or nervous speech patterns in contrast with merely looking at the cold pages of an appellate record."

550 F.3d 626, 630 (7th Cir. 2008) (alteration in original) (citations omitted) (quoting United States v. Woods, 233 F.3d 482, 484 (7th Cir. 2000)); *see also* Dyer v. MacDougall, 201 F.2d 265, 269 (2d Cir. 1952) ("He [the trial judge], who has seen and heard the 'demeanor' evidence, may have been right or wrong in thinking that it gave rational support to a verdict; yet, since the evidence has disappeared, it will be impossible for an appellate court to say which he was."); United States v. Zeigler, 994 F.2d 845, 849 (D.C. Cir. 1993) ("Demeanor evidence is not captured by the transcript; when the witness steps down, it is gone forever.").

95. 429 F.3d 685 (7th Cir. 2005).

96. *Id.* at 687.

97. *See id.* at 688.

98. *Id.*

Judge Posner's opinion recognizes the important truth that the fact-finder's cultural background will matter to how she perceives credibility, whether the subject is an asylum-seeker or someone from a different region or even a different neighborhood. A judge's "acquired sagacity" is simply an amalgamation of experiences he or she has or has not had—experiences that will often depend on the fact-finder's gender, ethnicity, race, and social privilege.[99] Indeed, in a system that conditions believability on being found worthy of belief based on a witness's appearance, even the wisest and most experienced fact-finders will have expectations or subconscious biases that influence their judgments.

As I have argued in past work, giving priority to demeanor does not, contrary to Judge Frank's suggestion, allow us to employ "one of the best guides available" to a witness's reliability.[100] Rather, by focusing—and instructing jurors to focus on—a witness's "manner, his intonations, his grimaces, his features," the law creates implicit behavioral prescriptions.[101] The notion that a person's features themselves or his "manner" should bear on whether we believe him will pressure witnesses—as guided by their attorneys—to perform credibility, to *appear believable*.[102] Some witnesses, for reasons unrelated to any inner qualities, will be unable to project the demeanor required to appear worthy of belief.[103] A person's outward bearing often inevitably suggests their degree of social privilege, level of education, comfort in speaking on the witness stand, whether they have a disability, and above all their race or ethnicity.[104] By insisting that credibility must and should be read externally, the legal system reinforces the normativity of those empowered to judge it.[105]

99. For example, psychology researchers have found that humans are hard-wired to attribute the characteristics of those we know to people we do not know who happen to have similar faces. ALEXANDER TODOROV, FACE VALUE: THE IRRESISTIBLE INFLUENCE OF FIRST IMPRESSIONS 133 (2017).

100. Broad. Music, Inc. v. Havana Madrid Rest. Corp., 175 F.2d 77, 80 (2d Cir. 1949); *Unmasking Demeanor*, *supra* note 2, at 169.

101. *Unmasking Demeanor*, *supra* note 2, at 169.

102. *Id.*

103. *See id.*

104. *Id.*

105. *See generally* Anna Offit, *Prosecuting in the Shadow of the Jury*, 113 NW. U.L. REV. 1071 (2019). Although most attention in this category is on judges and jurors, the process of assessing worthiness of belief begins with the attorneys tasked with deciding which cases to bring and how to present those cases. *Id.* For example, Anna Offit has shown in her anthropologic research on prosecutors that the jurors prosecutors imagine, and those jurors' imagined responses to witnesses, shape prosecutorial decision-making. *Id.* at 1092–96.

2. Prior Convictions

Legal actors tend to assert reflexively that credibility judgments are a measure of a witness's propensity to lie.[106] And indeed, there is plenty of support for this proposition in doctrine surrounding impeachment with prior convictions. The most influential pre-Federal Rules case on prior conviction impeachment, *Gordon v. United States*, explained that "the legitimate purpose of impeachment" is to elicit "background facts which bear directly on whether jurors ought to believe [the witness]."[107] Using this conception of credibility led the D.C. Circuit to the conclusion, which I have critiqued at length elsewhere,[108] that "convictions which rest on dishonest conduct relate to credibility whereas those of violent or assaultive crimes generally do not."[109] Once Congress enacted Rule 609, which provides a balancing test for admitting prior convictions as impeachment evidence, federal courts extended this logic, on the grounds that "Congress believed that all felonies have some probative value on the issue of credibility."[110]

What the courts claim to mean by credibility in this context is "the accused's propensity to testify truthfully."[111] Thus, in *Lipscomb*, another influential case, this one interpreting Rule 609, the court debated whether an accessory after the fact to manslaughter conviction was relevant to credibility in terms of what it might convey about the witness's "propensity to lie on the witness stand."[112]

106. For example, scholars often use this formulation when they refer to credibility. To offer just one example, in Andrea Roth's groundbreaking article on machine testimony, she argues that we may need what she terms "machine credibility testing." Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 2023 (2017). Professor Roth uses credibility here as synonymous with reliability or a propensity for truth, arguing that jurors need to have "the context they need to assess the reliability of [the] evidence." *Id.*

107. 383 F.2d 936, 940 (D.C. Cir. 1967).

108. *Credibility by Proxy*, *supra* note 2, at 192–96.

109. *Id.* at 193 (citing United States v. Estrada, 430 F.3d 606, 618 (2d Cir. 2005) (quoting *Gordon*, 383 F.2d at 940)).

110. United States v. Lipscomb, 702 F.2d 1049, 1057, 1061–62 (D.C. Cir. 1983). Today, almost all states also allow for impeachment with prior convictions. *See* Anna Roberts, *Conviction by Prior Impeachment*, 96 B.U. L. Rev. 1977, 1980–81 (2016) (identifying Kansas, Hawai'i, and Montana as the only three states not to allow the use of prior convictions).

111. *Lipscomb*, 702 F.2d at 1057.

112. *Id.* at 1072. Indeed, courts have endlessly parsed which prior acts or convictions have more or less to do with a witness's propensity for truth. In a representative example of this type of exercise, the Northern District of New York wrote:

> Convictions for murder, conspiracy, robbery, and weapons possession are generally not particularly probative as to honesty or veracity. However, "crimes requiring planning or preparation bear more strongly on veracity than violence alone

As I have documented in previous work, there is no empirical support for the notion that prior crimes or other past behavior can predict a *propensity* or *likelihood* that a witness will lie.[113] Social science researchers have shown that past behavior can help predict future actions only when all of the conditions surrounding the behavior are the same with a very high degree of particularity.[114] A prior conviction involving violence or theft is simply too vague an indicator to predict whether a witness will do something totally different, which is lie under oath.[115]

Like the emphasis on demeanor, impeachment with prior convictions is tailored to judging a witness's worthiness of belief rather than predicting future lies.[116] In order to be accorded credibility, to be someone who "ought to [be] believ[ed],"[117] a witness should avoid prior convictions, particularly for theft or crimes that are understood to involve deception.[118] A past free of entanglement with the law, like an outwardly

> suggests because planning indicates deliberate and injurious violations of basic standards rather than impulse or anger, and usually it involves some element of deceiving the victim." Moreover, "'theft' crimes, as well as 'crimes that involve evasions of responsibility or abuse of trust,' rank 'high on the scale of probative worth on credibility.'"

Somerville v. Saunders, No. 9:11-cv-556, 2014 WL 272415, at *8 (N.D.N.Y. Jan. 24, 2014) (citations omitted) (first quoting *Estrada*, 430 F.3d at 617–18; and then quoting Robinson v. Troyan, No. CV 07–4846, 2011 WL 5416324, at *2 (E.D.N.Y. Nov. 8, 2011)); *see also* United States v. Devery, 935 F. Supp. 393, 408 (S.D.N.Y. 1996) ("Just as mundane misconduct may be telling of a witness's character for truthfulness, the loathesomeness of prior misconduct does not necessarily bear on the perpetrator's capacity for truth-telling."), *aff'd sub nom.*, United States v. Torres, 128 F.3d 38 (2d Cir. 1997).

113. *See Credibility by Proxy*, *supra* note 2, at 208–09; *see also Unmasking Demeanor*, *supra* note 2, at 161.

114. *See Credibility by Proxy*, *supra* note 2, at 208–09.

115. *See id.* at 210–11. This does not even get at the related problem of many defendants taking pleas to crimes that have little relationship to their actual conduct. *See, e.g.*, Thea Johnson, *Fictional Pleas*, 94 IND. L.J. 855, 857 (2019) (defining a fictional plea as "a plea bargain agreement in which the defendant pleads guilty to a crime he did not commit, with the consent and knowledge of multiple actors in the criminal justice system—to avoid the profound collateral consequences that would flow from a conviction on his initial charge").

116. I make this claim in more extended form in my article *Credibility by Proxy*, *supra* note 2, at 184, 186, 188–89.

117. Gordon v. United States, 383 F.2d 936, 940 (D.C. Cir. 1967).

118. In other work, I have argued that there are deep historical roots for legal doctrines that suggest crimes of violence are less significant indicators of a witness or defendant's moral failing than crimes such as theft or fraud. Julia Ann Simon–Kerr, *Moral Turpitude*, 2012 UTAH L. REV. 1001, 1017–19 (2012) [hereinafter *Moral Turpitude*] (showing that "[f]raud and oath violations by men were routinely deemed to involve moral turpitude" but violence, because of its association with honor, did not); *Credibility by Proxy*, *supra* note 2, at 170–71 ("Far from being incompatible with 'worthiness of belief,' for a good, honorable, and therefore honest man, violence was at times required to maintain such worthiness.").

appropriate demeanor, meets the performative demands of credibility doctrine.[119]

China's social credit system further draws out this crucial distinction between worthiness-centered credibility and predictive credibility. As the next Section shows, in China, performative, worthiness-centered demands are being both created and enforced in the name of credibility. By contrast, the history of credit-scoring, discussed in Section C, shows one way in which algorithmic tools might be employed to shift the inquiry in the direction of data-driven risk-prediction.

B. Credibility as Worthiness: Chinese Social Credit

China offers a revealing vision of an algorithmic system for credibility assessment. Under the aegis of the central government, some regional governments in China have been investing in what is referred to as "social credit-scoring."[120] As explained by Chinese legal scholar Xin Dai, the word credit in Mandarin is associated with "a host of lofty moral virtues such as trustworthiness, promise-keeping, norm abiding, integrity and general courtesy."[121] The social credit system thus has a broad "moral framing",[122] a framing that roughly tracks many components of credibility doctrine in the United States. It is important to note that the social credit-scoring project in China is massive, multifaceted, and still in its infancy.[123] Through it, however, the Chinese government hopes to achieve a measure of social control and regulation, by harnessing "reputation as an instrumental force of governance."[124] In this Article's parlance, the social credit score is designed to prescribe how to be worthy of belief.

119. Whether it is more or less appropriate to create penalties for prior convictions than it is to penalize demeanor is a topic for another article. The point here is simply that these two credibility markers perform parallel functions in the doctrine, creating behavioral incentives rather than allowing us to make behavioral predictions.

120. *See generally* TRIVIUM REPORT, *supra* note 8. The Trivium Report has been described as "the most comprehensive assessment yet of what China's impending regulatory system means for U.S. corporations and interests." Eamon Barrett, *Blacklist vs. 'Redlist': What to Know About China's New Corporate Social Credit Score*, FORTUNE (Dec. 10, 2020, 6:00 AM), https://fortune.com/2020/12/10/china-corporate-social-credit-system-cscs-blacklist-redlist/. It was produced by a Beijing-based consultancy at the behest of the U.S.-China Economic and Security Review Commission, which is "a government advisory panel established by Congress in 2000." *Id.*

121. Dai, *supra* note 4, at 14.

122. *Id.*

123. *Id.* at 15.

124. *Id.*

There is currently no unified social credit score as such in China.[125] Rather, some municipalities and regions have begun to create social credit scores using varying factors, scales, and sources of data.[126] This Section relies on the available information to outline the Chinese government's reasons for undertaking the social credit project and to canvass some of the ways in which the scores are being implemented. It does not purport to survey the entire spectrum of social credit scoring in China.

Social credit scores have been analogized to United States credit scores, but they differ in notable ways.[127] Most importantly, social credit ratings include information that extends well beyond consumer financial data.[128] The Chinese government maintains centralized databases on its citizens that may include information about anything from their prior legal entanglements and financial history to their charitable contributions and Communist Party membership to whether they return their library books on time.[129] The store of information that could potentially be integrated into the social credit score is vast.

Although the central government is not yet issuing social credit scores, a master database of social credit records called the "National Credit Information Sharing Platform" has already been built.[130] A significant amount of the data this database contains is or will be publicly available.[131] The government has also published a policy paper with guidelines and suggestions for facilitating the social credit system.[132]

125. TRIVIUM REPORT, *supra* note 8, at 5.

126. *Id.*

127. *See, e.g.*, Síthigh & Siems, *supra* note 22, at 1049, 1058–59 (2019) ("The Social Credit System aims to address not only the financial credit-worthiness of individuals and companies but also their sincerity, honesty, and integrity.").

128. TRIVIUM REPORT, *supra* note 8, at 11 fig.1.2-C.

129. *Id.* at 9–11 (describing breadth of Chinese government's existing data collection and a 2016 policy document about datasets to be contributed to the social credit scoring endeavor).

130. *Id.* at 9 (describing platform as "the primary clearinghouse for social credit files on individuals and corporations").

131. *See, e.g.*, *id.* at 16 (describing blacklisting system contingent on public shaming); *see also id.* at 21 (describing availability of corporate social credit records to the public).

132. Guowuyuan Bangong Ting Guanyu Jiakuai Tuijin Shehui Xinyong Tixi Jianshe Goujian Yi Xinyong Wei Jichu de Xinxing Jianguan Jizhi de Zhidao Yijian (国务院办公厅关于加快推进社会信用体系建设构建以信用为基础的新型监管机制的指导意见) [Guiding Opinions of the General Office of the State Council on Accelerating the Construction of the Social Credit System and Building a New Credit-based Supervisory Mechanism] (promulgated by General Office of the State Council of Central People's Government of the People's Republic of China, July 16, 2019, effective July 16, 2019) http://www.gov.cn/zhengce/content/2019-07/16/content_5410120.htm (translation by Tracy Cui on file with author) [hereinafter *Policy Statement*].

These guidelines familiarly conflate credibility with truth, suggesting that "[c]redibility education is needed" in order to "increase awareness of truthfulness in all fields."[133] In the meantime, as part of this larger project, some city and municipal governments are issuing social credit scores to Chinese citizens within their jurisdictions.[134] These localized scoring systems differ along many dimensions, but they all assign a numerical score.[135] In addition to financial factors, the scores may reflect social factors, such as whether or not individuals have received awards for good citizenship, whether they have engaged in plagiarism or other forms of fraud, whether they donate blood, whether they volunteer for charity and whether they are current on child support payments.[136]

In many cities using these scores, app developers are working with local or regional governments to produce apps that will allow people to check their scores and those of others.[137] For example, one nationally-sponsored app allows users to search government-maintained blacklists, or databases of people who fail to fulfill court-ordered obligations.[138] These score-checking apps intentionally create feedback loops so that the scores will affect behavior.[139] Other apps have been developed to

133. *Id.* §1.

134. TRIVIUM REPORT, *supra* note 8, at 24.

135. *See* Nir Kshetri, *China's Social Credit System: Data, Algorithms and Implications*, IT PRO., Mar.–Apr. 2020, at 14, 14–15 (noting that inputs have "varied across municipalities that have implemented" social credit scoring); *see also* TRIVIUM REPORT, *supra* note 8, at 24.

136. *See, e.g.*, TRIVIUM REPORT, *supra* note 8, at 11 fig.1.2-C (listing areas of focus for social credit data collection); *id.* at 15 (describing donating blood, volunteering, or donating to charity as ways to get off government blacklist); *id.* at 31–32 (describing collaboration between Supreme People's Court and a large corporation to offer a service notifying people when they call or are called by someone with debts or delinquent child support payments with message: "Included on the list of dishonest individuals."); *see also* Kshetri, *supra* note 135, at 16 tbl.1.

137. Kendra Schaefer, *The Apps of China's Social Credit System*, TRIVIUM USER BEHAV. (Oct. 14, 2019), http://ub.triviumchina.com/2019/10/long-read-the-apps-of-chinas-social-credit-system/. These blacklists have been analogized to sex offender registries in the United States. TRIVIUM REPORT, *supra* note 8, at 31.

138. Schaefer, *supra* note 137 (stating that individuals with severe and notorious violations of the law or the credit system, which results in the disruption of the society, will be blacklisted).

139. *See* John Lanchester, *Document Number Nine*, LONDON REV. BOOKS (Oct. 10, 2019), https://www.lrb.co.uk/the-paper/v41/n19/john-lanchester/document-number-nine (describing social credit system's goal of making people internalize the state's values and priorities); *see also* Qin Chen, *Court's Punishment of 9-Year-Old Girl Highlights Concern with China's Expansive Social Credit System*, INKSTONE (Dec. 22, 2020), https://www.inkstonenews.com/society/courts-punishment-9-year-old-girl-highlights-concern-chinas-expansive-social-credit-system/article/3114954 (quoting a Chinese attorney's description of the social credit scoring system: "Everyone is like a rat in a cage in such a

incorporate reports of disruptive or illegal behavior on public transportation and incorporate those reports into the social credit records of passengers.[140] One such app allows users to upload reports about other passengers, with, for example, claims that they engaged in behavior such as fighting, boarding without a ticket, or opening emergency exits.[141] These and other violations can result in passengers being banned from future rides.[142] The public transportation app allows for both positive and negative reports and also has a function for users to upload photographic evidence.[143] Reports are purportedly verified and then "logged in the target's social credit file."[144]

The idea for social credit scoring in China originated with a narrower government attempt to create credit and bond rating systems.[145] In its effort to measure "responsibility" or creditworthiness using factors other than formal financial history, the Chinese government is relying on the kind of broad, reputation-based information that populated early credit-reporting in the United States.[146] The system is designed to impress upon the public that good reputation and character—as defined by the government—are essential to creditworthiness and the ability to engage in sanctioned business transactions.[147]

Far from the five factors that go into a modern United States credit score,[148] "public credit information," as it is termed in China, includes "broadly any information and data collected or generated by public agencies that can be used in evaluating how well a subject observes its legal and contractual obligations."[149] Beyond this, the government has embraced the idea that reputational feedback mechanisms can "supplement law and regulations" as a means of creating social compliance with behaviors the government deems desirable.[150] One such

society. By monitoring and simultaneously correcting, many nature [sic] of this rat could be altered.").

140. Schaefer, *supra* note 137.

141. *Id.*

142. *Id.*

143. *Id.*

144. *Id.* It is unclear how widely this app is used, but it has been operating for three years, which suggests it is perceived as efficacious. *Id.*

145. *See* Dai, *supra* note 4, at 10–13, 15.

146. *See* Kshetri, *supra* note 135, at 16 tbl.1 (describing "factors likely to be included" in social credit scoring as including political activity and ideology, engagement in "criminal, illegal, immoral and socially deviant lifestyle and behaviors," "nature of social networks," "role as a productive citizen," and responsible consumer and financial behavior).

147. *See id.* at 17 (noting that the government rationalizes inputs to credit scores by stating that they can promote social harmony and "socialist core values").

148. OLEGARIO, *supra* note 16, at 212 tbl.5.3.

149. Dai, *supra* note 4, at 19.

150. *Id.* at 24.

effort lauded by the Shanghai government involved incorporating "malicious use of bike sharing services" into the social credit model.[151] Another in Guizhou Province aims to give rural residents incentives to uphold "the good moral value of promise keeping."[152] In this system, government rewards are conditioned on "community monitoring and peer review" of individual households.[153] This latter program is a self-conscious attempt to reinvigorate social "village norms" by offering government incentives.[154]

Although social credit in China is still being developed and may never emerge as a robust, centralized system, that may not be the point. Even without broad implementation, the system is one, as Professor Dai puts it, of "social meaning creation."[155] Government entities are instructed to "carry out promotional programs to create the salience of 'honesty and trustworthiness' as a universal norm."[156] Private enterprises then respond by advertising their products as serving the goals of "restoring societal trust and creditworthiness."[157] As one Western critic observes, the ultimate aim of the system can be understood as deeply Foucauldian: "to make people self-censor, self-monitor, self-supervise."[158]

Social credit scoring in China could serve multiple functions. On a practical level, some form of social credit amalgamation is already being used to prevent people with low scores from engaging in certain activities, like traveling on trains or planes.[159] Where they exist, the scores are

151. *Id.*

152. *Id.* at 30 (quoting Wang Fuyu (王富玉), Chengxin Yinlai Jin Fenghuang——Guizhou Sheng Kaizhan Chengxin Nongmin Jianshe de Shijian Yu Sikao (诚信引来金凤凰——贵州省开展诚信农民建设的实践与思考) (*Trust and Honesty Attracts Fortune: Experience of Guizhou's Honest Farmer Project*), Qiushi (求是) [SEEKING TRUTH] (Feb. 17, 2012), http://www.reformdata.org/2012/0216/2393.shtml).

153. *Id.*

154. *Id.* at 30–31.

155. *Id.* at 25.

156. *Id.*

157. *Id.* (describing Sesame Credit's marketing).

158. Lanchester, *supra* note 139. The United States credit-scoring system also has this effect—rewarding those who are able to shape their use of credit in ways beneficial to their credit scores. At the same time, the major credit agencies do not reveal the precise inputs into their credit-scoring algorithms, making the feedback loop between score and behavior much less intentional and more attenuated.

159. Dai, *supra* note 4, at 33 (describing being "prohibited from travelling by air or on high-speed trains, staying in high-end hotels, or playing golf," and enrolling children in expensive private schools as consequences of the Chinese government's blacklist system); TRIVIUM REPORT, *supra* note 8, at 13 (describing judicial blacklist prohibiting high-speed rail travel, air travel, home renovations, staying in high-end hotels, and sending children to expensive schools); *see, e.g.*, Chen, *supra* note 139 (describing uproar when a nine-year-old girl was placed on a blacklist prohibiting her from doing "high level consumption" activities, such as flying or checking into a hotel as a result of debt owed by her father).

often public and work to shape people's day-to-day interactions with strangers as well as to control formal aspects of their lives. While one clear target of the central government is better regulation of the business community, [160] the development of scoring systems thus far has implications for anyone living in a target area.[161] Finally, in its focus on social behavior enforced by panoptic monitoring, the social credit system is meant to reinforce behavior that is seen as socially beneficial.[162] Its ambition is to create a value system and to ascribe new meaning to certain behaviors. A person who donates blood or complies with local customs is deemed credible and credit-worthy, while someone who is remiss on child support or rides a bike aggressively is not. The goal of such a system is not to predict risk to the exclusion of all other concerns, as with the modern United States credit score, but rather to connect credit or worthiness to socially valued behavior as an end in and of itself.[163]

The case of social credit in China offers a living example of how those in power may use social incentives to shape behavior, controlling who is deemed credible and creditworthy and who is not. Because credibility inevitably depends on performance and social perception, it can be used to reward desired behavior and penalize deviance. The example of social credit shows how credibility can be constituted through a score to reflect the conceptions of those who make or apply the rules about what it should mean to be worthy of belief. Rather than predict trustworthiness itself, the government seeks social coherence centered around behaviors it has decided to link to credibility.

It is easy to highlight the dystopian elements of this tool for manipulating credibility, [164] yet the political and legal history of the

160. *See generally* TRIVIUM REPORT, *supra* note 8; *see also Policy Statement*, *supra* note 132 § 1.

161. *See, e.g.*, Drew Donnelly, *An Introduction to the China Social Credit System*, NEW HORIZONS, https://nhglobalpartners.com/chinas-social-credit-system-explained/ (last updated Oct. 26, 2021) (describing social scoring implementation in Chinese cities in which scores are being lowered for offenses like cheating in online videogames or not visiting one's parents often resulting in blacklisting and a loss of the right to buy plane tickets and travel).

162. *See* Dai, *supra* note 4, at 30.

163. TRIVIUM REPORT, *supra* note 8, at 24. The Chinese government does apparently have some "hopes" that social credit data could at some point be used to predict risk. *Id.* at 10.

164. One episode of the television show, *Black Mirror*, which depicts a world in which people score each other from their phones after every interaction, has been described as "eerily similar" to social credit scoring in China. Gabrielle Bruney, *A 'Black Mirror' Episode is Coming to Life in China*, ESQUIRE (Mar. 17, 2018), https://www.esquire.com/news-politics/a19467976/black-mirror-social-credit-china/. In the episode, being superficially nice and appearing well-kept are keys to a high score, which can in turn offer access to social benefits like housing and travel opportunities. *Black Mirror: Nosedive* (Endemol Shine UK

United States (and indeed many nations) is replete with similar, albeit non-algorithmic, examples of the deliberate inculcation of values deemed productive by those in power. In the United States, the law of credibility still reflects values promoted with great intentionality by the nation's founders.[165] It is also focused on a set of attributes or features that suggest worthiness of belief. As with the social credit score, credibility jurisprudence offers credibility as a reward for conforming one's demeanor, one's conduct, or one's reputation to meet a fact-finder's expectations. Formal evidence rules also establish penalties for certain prior convictions, which will presumptively reduce a witness's credibility. Although the feedback loop of credibility jurisprudence is far narrower than that created by Chinese blacklists, when attorneys tell their clients to dress or speak a certain way, or advise them not to testify because they have prior convictions, they reinforce and perpetuate values prioritized by the law of credibility. Finally, as in the case of China's system, credibility jurisprudence can lay no claim to identifying features that in fact uncover a witness's propensity for truth or untruth. Rather, what this body of law portends is social worthiness. In these ways, the common law of credibility looks much like a low-tech version of social credit scoring.

C. Risk-Predictive Credibility: Financial Credit Scoring in America

Even as our credibility jurisprudence is now tailored to witnesses' worthiness of belief, the legal system maintains the fiction that credibility assessments center on the risk that a witness is lying. What would it look like to translate this fiction into reality using algorithmic tools? The following brief history of credit scoring in the United States provides one possible answer.

In the United States, financial creditworthiness and credibility were at one time inextricable.[166] The ability to borrow money depended on some combination of a person's social stature and how others perceived him or her.[167] As Josh Lauer describes in his work on the history of financial identity in America, creditors in the early days of the republic

Oct. 21, 2016). Much of the tension in the episode comes from the main character's attempt to improve her score, which drives her own superficiality into untenable territory. *Id.* As in China's system, the consequences of a low score include being ejected from the airport and then prohibited from traveling on what seems to be the equivalent of a Greyhound bus. *Id.*; Bruney, *supra* note 164.

165. *See Moral Turpitude*, *supra* note 118, at 1007–08.

166. *See generally From Rumor to Written Record*, *supra* note 13, at 311.

167. *From Rumor to Written Record*, *supra* note 13, at 305, 307 (describing "the moral underpinnings" of credit reporting and its dependence on "socially determined markers of trustworthiness and economic legitimacy") (emphasis omitted).

through the mid-1800s would observe borrowers to look for "evidence of integrity or, contrarily, sloth and vice."[168] Benjamin Franklin highlighted the importance of signaling one's virtues to lenders in his autobiography.[169] He wrote that one had to be not only "in reality industrious and frugal," but also to avoid "all appearances of the contrary."[170] Historian Rowena Olegario explains that "[s]ociety was small, and personal reputations . . . were the basis for transactions."[171] If a man encountered financial difficulties, his failure was attributed in part to "defects in character."[172] In making decisions about whether to lend money, creditors wanted to get a sense of the borrower's credibility in the community—how those around him understood his character, and in turn his trustworthiness.

In the early 1800s, the growth of American society led to what Lauer calls "a breakdown of social trust within the commercial sphere."[173] Relying on mutual social networks, personal interaction, or even letters of recommendation to decide whether a borrower could be trusted no longer worked in many situations.[174] In 1843, responding to this changed environment, Lewis Tappan launched the first early credit reporting—or Mercantile—agency.[175] Tappan's agency collected financial information about borrowers, but its early method largely reproduced the system of character-based credit.[176] Lauer explains that Tappan's agency was concerned with merchants' "standing" and "responsibility," which Tappan advertised would help "ascertain whether persons applying for credit are worthy of the same and to what extent."[177] The reporting system consisted of ledgers with "detailed reports on the personal character, financial means, and local reputations of [business] proprietors."[178] These reports, available to lenders on a subscription

168. *Id.* at 307.

169. BENJAMIN FRANKLIN, THE AUTOBIOGRAPHY OF BENJAMIN FRANKLIN (1903).

170. *Id.* at 99 (emphasis omitted).

171. OLEGARIO, *supra* note 16, at 18.

172. *See id.*

173. *From Rumor to Written Record*, *supra* note 13, at 306.

174. *Id.* In one sign of how far evidence law has lagged behind developments in credit-scoring, it took another century for the Supreme Court to observe that "[g]rowth of urban conditions, where one may never know or hear the name of his next-door neighbor, have tended to limit the use" and efficacy of character witnesses. Michelson v. United States, 335 U.S. 469, 480 (1948). Nonetheless, in that 1948 opinion the Court continued to endorse the practice, and it has persisted ever since. *See id.*

175. *From Rumor to Written Record*, *supra* note 13, at 302–03.

176. *See id.* at 309 (describing Tappan's method as "based primarily upon personal knowledge and communal opinion," which were "time-honored and trusted ways of knowing").

177. *Id.* at 303.

178. *Id.*

basis, were initially qualitative, and might include information such as "the political affiliation, moral standing and personal weaknesses (e.g. alcohol) of the applicant."[179]

The courts developed doctrines recognizing that the exchange of credit was founded in and facilitated by reports about the reputations of business proprietors. Chief Justice Marshall laid one such foundation in an early opinion dealing with allegations of fraud that involved a reference given by one merchant to another to recommend a third.[180] When a merchant offers such a reference, Justice Marshall wrote, "he must be presumed to speak from that knowledge only which is given by reputation."[181] In 1851, the Indiana Supreme Court explained that this must be the correct rule because to hold letter-writers responsible for the truth of their references "would be a very dangerous thing, for every such letter would amount to little less than a guaranty."[182] A system that functioned on reputation-based referrals could not survive if those offering the referrals were held accountable for the actual behavior of those they recommended.

Reflecting the deep interconnectedness of early credibility in the law and in the financial world, the system of credit reporting initially relied on notes from local attorneys who endeavored to convey the prospective borrower's local reputation for the benefit of a much wider audience.[183] Their task essentially was to convey their subject's local community standing.[184] If information about a subject's property or financial dealing was scarce, the reporter would add information about his "habits or family connections."[185]

This focus on a borrower's perceived moral character persisted into the twentieth century. In the early twentieth century, for example, department stores issued credit based on the premise that a person's "pride in square dealing" would be handed down within a family,[186] or on information "regarding the character of the position which the customer or her husband holds."[187] As the credit-reporting system grew, reporting agencies struggled for ways to translate local qualitative information into

179. Jeacle & Walsh, *supra* note 14, at 743.

180. Russell v. Clark's Ex'rs, 11 U.S. 69, 94 (1812).

181. *Id.* at 95.

182. State Bank v. Hamilton, 2 Ind. 457, 464 (1851).

183. *From Rumor to Written Record*, *supra* note 13, at 308–09. The reporters were unpaid but might get business referrals in exchange. *Id.* at 308.

184. *Id.*

185. *Id.* at 310.

186. Jeacle & Walsh, *supra* note 14, at 742–43 (quoting PAUL HENRY NYSTROM, RETAIL SELLING AND STORE MANAGEMENT 262–63 (1925)).

187. *Id.* at 743–44 (emphasis omitted) (quoting RUTH LEIGH, ELEMENTS OF RETAILING 344–45 (1923)).

quantitative fact. Lauer writes that the agencies had trouble identifying what information would be relevant because without the social context, it was hard to make sense of the reports.[188]

As they became central to the modern credit economy, credit reports created a new social reality. The scores established a form of social ordering, offering access to financing to those who could achieve the requisite score and denying it or offering more onerous terms if they could not.[189] Despite optimistic views that credit reports might promote "mercantile honor,"[190] they were susceptible to the religious or political prejudices of reporters or even personal grudges.[191] Importantly, this whole system of early credit-reporting largely excluded those who were non-white.[192] As a result, as one proposal to publish a credit rating guide for the African American community in Chicago put it, "[w]e as a race are generally discredited."[193] This exclusion continues to have repercussions today as "certain groups [are denied] favorable access to credit merely because they have been excluded from the credit market in the past."[194]

For those who did fall within the ambit of early credit scoring, because early credit scores were based largely on reputation, as with credibility in the courtroom, more credit was accorded to those who behaved *as if* they were trustworthy and a good bet.[195] Even as written

188. *From Rumor to Written Record*, *supra* note 13, at 313 (observing that when social context was stripped away, it "left individuals looking rather pallid and one-dimensional . . . or hopelessly complex and contradictory").

189. *Id.* at 321. Indeed, Lauer argues that credit-reporting amounted to "disciplinary surveillance" of business behavior. *Id.* Rowena Olegario describes a similar effect. She writes that "credit-reporting agencies institutionalized the informal set of rules—the norms, customs, and habits of thinking—that had served since colonial times as guides for judging who was, and was not, worthy of credit." OLEGARIO, *supra* note 16, at 77. At the same time, Olegario notes a paradox in the early credit economy of the United States. She writes that the "relative egalitarianism" and lack of hierarchy in the country made "credit seem destabilizing because it gave the humbler orders a means to emulate their wealthier neighbors." *Id.* at 28.

190. *From Rumor to Written Record*, *supra* note 13, at 321–22.

191. Josh Lauer writes, for example, that the correspondents who contributed information to early credit reporting agencies "wielded enormous unchecked power." CREDITWORTHY, *supra* note 15, at 40. Private grudges were a concern, and "[i]n rural communities, where divisions along political or religious lines skewed impressions, accusations of prejudice were common." *Id.*

192. *Id.* at 141 (describing how as the credit economy developed it "was largely devoted to the world of white Americans and European immigrants"). When Black Americans were listed in early credit reporting ledgers, their race was indicated unlike whites for whom no race was mentioned. *See id.* at 67.

193. *Id.* at 141.

194. Mikella Hurley & Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 YALE J.L. & TECH. 148, 156 (2016).

195. *See* Jeacle & Walsh, *supra* note 14, at 743.

records displaced oral accounts by merchants, local knowledge was still key to credit reports, and the emphasis was still on "character, morality and other personal characteristics."[196]

The process of converting credit reports to numerical shorthand and then to scores began in the mid 1850's.[197] In 1864, the leading credit agency redesigned the rating system to offer a grading system with a category for "pecuniary strength" that focused on capital rather than character.[198] A second column gave a ranking for "general credit" that "implicitly captured character and capacity."[199] This innovation is widely understood as one that gave rise to the general credit rating. It facilitated a system of comparative credit and credit regulations by eliminating the difficulty of interpreting narrative reports and providing a veneer of objectivity.[200] That it was a veneer is clear, in Bruce Carruthers' words, from the "haphazard set of variably substantiated facts and interpretations" upon which the scores were based.[201] Indeed, even the coding itself "was vague and unspecified."[202] Still, the very fact that the scores grouped borrowers in a single rating scale, proved especially powerful. As Carruthers explains, when compared with other contemporaneous systems of assessing credit, the standardization of the credit score led to its achieving "canonical status."[203]

Much like the predictive claims made for credibility metrics such as prior convictions in today's courtrooms, credit ratings were initially touted without information on how well they might predict default.[204] Carruthers explains that the ratings offered "the appearance of quantitative rigor," without any indication that they could provide it.[205] He goes on to theorize that credit ratings' widespread adoption was a function not of their accuracy but of their ease of use.[206] The ratings provided a "type of procedural rationality" through which banks could streamline their lending decisions by setting thresholds and also cover themselves against accusations of discrimination or favoritism.[207]

196. *Id.* at 744.
197. *From Rumor to Written Record*, *supra* note 13, at 317–18.
198. *Id.*; *see also* Bruce G. Carruthers, *From Uncertainty Toward Risk: The Case of Credit Ratings*, 11 SOCIO-ECON. REV. 525, 532 (2013).
199. *From Rumor to Written Record*, *supra* note 13, at 319.
200. *See id.*; *see also* Carruthers, *supra* note 198, at 533.
201. Carruthers, *supra* note 198, at 532.
202. *Id.* at 533.
203. *Id.* at 529–30 (describing United States credit scores relative to a French system of notaries who matched lenders with borrowers but did not publish their information).
204. *Id.* at 540–41.
205. *Id.* at 540.
206. *Id.* at 544.
207. *Id.*

Rowena Olegario traces the advent of assessing "creditworthiness through statistical analysis of data" to 1958, when Fair, Isaac and Company introduced an early relative of the FICO score.[208] As Olegario explains, by 1991, Fair, Isaac was working with the three biggest credit bureaus.[209] Although they used different proprietary algorithms, this cooperation allowed a unified scoring system where a particular score "indicated the same level of risk" no matter which firm had generated it.[210] This had diverse effects, as lawmakers and others began to use particular scores as credit benchmarks, but one of its effects was to place the focus squarely on quantifying risk.[211]

This focus on quantification changed dramatically the kind of information that was understood as relevant to the question of creditworthiness. In the 1850s, criteria for creditworthiness included a list of thirty-six items, including elements of a borrower's broader credibility, such as "[n]ot very good private character," "[n]ot temperate," and "[h]onesty not fully endorsed."[212] In 2014, by contrast, the FICO score was based on five criteria: payment history, amounts owed, length of credit history, new credit, and types of credit used.[213] Although proprietary algorithms and data are used to come up with a borrower's exact score, the score itself is based on standardized information with none of the "rich details" that had been relied on previously.[214] In a development that has yet to reach the law of credibility, the link between "personal reputation[]" and credit was severed.[215]

Lauer argues that as it developed, the system of credit reporting marked the invention of financial identity itself.[216] Because credit-scoring has become central to modern finance, it may be difficult to understand today how revolutionary the idea of the credit report was. For credit-reporting to function, Lauer explains, the nation had to accept that one's identity with respect to money could be disentangled from one's general social status or character.[217] While that may seem

208. OLEGARIO, *supra* note 16, at 166.
209. *Id.* at 210.
210. *Id.*
211. *Id.* at 210–11. This facilitated subprime lending, among other things. *Id.*
212. *Id.* at 212 tbl.5.3.
213. *Id.*
214. *Id.* at 213, 216.
215. *Id.* at 216.
216. *See generally From Rumor to Written Record*, *supra* note 13, at 302 (arguing that the advent of mercantile agencies marked "the invention of disembodied financial identity").
217. *See id.* at 313–17.

noncontroversial today,[218] it marked a wholesale shift in thinking about identity, and correspondingly, about credibility.

The advent of credit scoring, and particularly scoring using proprietary algorithms, has perpetuated and introduced many problems, among them invidious biases in the scoring algorithms, difficulties in identifying and correcting errors, and the use of credit scores as proxies for character in wholly inappropriate contexts, such as immigration.[219] Still, modern credit scores are fairly accurate at predicting the risk of default on a given type of loan within a certain period.[220] In this way, financial ability and willingness to pay loans has been unbundled from

218. In the twenty-first century, successful presidential candidates have been known to boast of their engagement with the bankruptcy system. *See Legislative Highlights*, AM. BANKR. INST. J., Oct. 2015, at 10 (quoting future President Trump as saying, "I have used the laws of this country . . . the [bankruptcy] chapter laws, to do a great job for my company, for myself, for my employees, [and] for my family," during a Republican primary debate) (alterations in original).

219. Capon, *supra* note 17, at 87; *see also* Hurley & Adebayo, *supra* note 194, at 155–58, 167 (noting the persistence of errors in credit scoring and how that perpetuates biases); STEPHEN L. ROSS & JOHN YINGER, THE COLOR OF CREDIT: MORTGAGE DISCRIMINATION, RESEARCH METHODOLOGY, AND FAIR-LENDING ENFORCEMENT 22–26 (2002) (describing the statistical techniques in credit-scoring and the biases that can be introduced); Andriotis, *supra* note 19 (describing 2019 decision by the Trump administration to make credit scores a factor in immigration decisions). Although credit scores have been seen as an antidote to the bias of individual lenders or businesses, this has not necessarily been the case. In 1974, for example, the Equal Credit Opportunity Act outlawed credit discrimination on the basis of race, religion, national origin, sex or marital status or age. 15 U.S.C. § 1691(a)(l). The Act endorsed the use of credit scores while also providing some constraints on what factors they might consider. *See, e.g.*, *id.* § 1691(b)(3) (allowing use of "any empirically derived credit system which considers age if such system is demonstrably and statistically sound in accord[] with . . . Bureau [regulations]" except that "such system . . . may not . . . assign[] a negative factor or value" to "the age of an elderly applicant"); *see also* Capon, *supra* note 17, at 84. Proponents of the scores argued that credit scores would facilitate objective lending decisions. CREDITWORTHY, *supra* note 15, at 250. Congress's endorsement had led the personal finance market to adopt credit scoring technology more broadly. *See id.* When, in the 1990s, the nation's two largest government sponsored home loan brokers instructed lenders to use credit scores, it led to widespread use of credit scoring in the mortgage industry. *Id.* at 249–50. Neither of these government interventions has been seen as wholly, or even largely beneficial to the communities or industries they were intended to serve. The personal finance market is still rife with inequality, in part due to the use of credit scores. *See, e.g.*, Hurley & Adebayo, *supra* note 194, at 156; ROSS & YINGER, *supra* note 219, at 22–26 (describing the role of credit scoring in lack of access for minority borrowers to prime lenders and consequences of Black and Hispanic borrowers being shunted to subprime lenders). The mortgage industry's meltdown in 2007–08 has likewise been attributed at least in part to the focus on credit scores. *See, e.g.*, Peter Henderson et al., *Credit-Score Panacea Failed to Stop US Mortgage Crisis*, REUTERS (May 10, 2007, 2:22 PM), https://www.reuters.com/article/us-usa-subprime-scores/credit-score-panacea-failed-to-stop-us-mortgage-crisis-idUSN023119182007051O.

220. Dobbie et al., *supra* note 18, at 2393–96 fig.1.

the larger package of "credibility," and assessed using a predictive algorithm.

IV. THE CREDIBILITY OF THE FUTURE

Credibility's future matters. As described in Part II, algorithmic tools for lie detection are already stalking the courtroom. Yet, legal credibility has two distinct meanings that, if embodied algorithmically, would lead to tools as functionally distinct as a financial credit score in the United States and a Chinese social credit score. At present, this dualism contributes to both definitional and functional confusion. If courts are unsure whether there is a distinction between finding that a declaration "is more likely true than false" and evaluating the credibility of the witness,[221] it will be all the more difficult to grapple with tools that purport to automate such judgments.

Indeed, even in a low-tech age, the confusion about credibility already has costs. One is the way in which a worthiness-centered conception privileges the powerful and systematically discredits others—often along lines of race, gender, and class.[222] Another is that credibility findings can be manipulated to circumvent procedural rules. Finally, confusion about whether credibility impeachment predicts lying or assesses worthiness has also caused problems in specific areas of law. In one representative example, applicants applying for disability benefits under the Social Security Act were for many years subject to a "credibility analysis" by the Administrative Law Judge ("ALJ").[223] In conducting that analysis, ALJs were asked to consider whether the claimant's symptoms and their severity and duration are consistent with "the objective medical evidence," among other factors.[224] In practice, however, ALJs also considered evidence they perceived to be relevant to "a plaintiff's general tendencies towards truthfulness."[225] The Social Security Administration eventually "instructed ALJs to no longer call this analysis a credibility analysis" because the word "'credibility' led some ALJs to inquire into impermissible considerations . . . like a claimant's general propensity for

221. Latif v. Obama, 677 F.3d 1175, 1190 (D.C. Cir. 2011).

222. *See Uncovering Credibility*, *supra* note 2, at 598.

223. 20 C.F.R. § 416.929(c)(1), (4) (2017); *see also* STEPHANIE J. TATHAM & MATTHEW LEE WIENER, ADMIN. CONF. OF THE U.S., EVALUATING SUBJECTIVE SYMPTOMS IN DISABILITY CLAIMS 15 (2015) (describing the historical usage of "credibility analysis" and "credibility determinations" in SSA decisions).

224. 20 C.F.R. § 416.929(c)(4) (2017).

225. Dubrawsky v. Berryhill, No. 3:16-cv-00433-AA, 2017 WL 1758054, at *11–12 (D. Or. May 2, 2017).

truthfulness."[226] Using the word credibility also caused "some claimants to believe an adverse determination of credibility was an accusation of untruthfulness."[227] This example is one of many that show the peril in using "credibility" to encompass factors that seem socially or contextually relevant to believing a witness while simultaneously using the term to refer to probabilistic judgment about truthfulness. [228]

Adding big data into this picture has the potential to exacerbate this definitional confusion as well as to reinforce systematic biases. As we anticipate the inevitable expansion of algorithms into credibility assessments, a first step is to consider the credibility models those algorithms might pursue. The examples canvassed in the preceding Part suggest how existing conceptions of credibility might map onto a world of big data. Like the data-driven refinement of credit-scoring, we could choose to unbundle the propensity for truth from the rest of the credibility package and train our algorithms to focus on this predictive question of truthfulness alone. Alternatively, like China, we could use big data to maintain our longstanding approach to credibility, making explicit the proxies that make a witness worthy of belief. Rather than assessing risk, this system would serve to reinforce and reward prioritized behaviors or other external markers by decreeing that they are at the heart of credibility. The remainder of this Part begins to envision what these two paths might look like.

A. The Worthiness Algorithm

What would happen if we tried to replicate existing credibility jurisprudence using something like China's social credit scoring system? A social credit approach to using big data to measure credibility, which I term a "worthiness algorithm," would require information on many facets of our lives, such as our community reputations for being truthful, how we dress, whether we seem to be forthright, our facial features and expressions, which stories we tell and how we tell them, and our prior legal entanglements. As with China's system, this could help reinforce

226. Jines v. Berryhill, No. 18-1234, 2019 WL 4644000, at *4 (W.D. Tenn. Sept. 24, 2019) (citations omitted).

227. *Id.*

228. A similar problem arose in Rule 608 of the Federal Rules of Evidence. FED. R. EVID. 608(b) advisory committee's note to 2003 amendment. In order to address it, in 2003, Rule 608(b) was amended to remove the word "credibility" and replace it with "character for truthfulness." *Id.* The Rules committee explained that "the Rule's use of the *overbroad* term 'credibility' has been read 'to bar extrinsic evidence for bias, competency and contradiction impeachment since they too deal with credibility.'" *Id.* (emphasis added) (quoting DANIEL D. BLINKA ET AL., A.B.A., EMERGING PROBLEMS UNDER THE FEDERAL RULES OF EVIDENCE 161 (David A. Schlueter & Stephen A. Saltzburg eds., 3d ed. 1998)).

the legal system's vision of credibility by offering the reward of being believed to those who comply. Jurors and witnesses, as well as repeat players like attorneys and judges, could disseminate this vision into the broader community. This might even generate public debate about the features that *should* matter to credibility or what performances we should expect from a credible witness. Pursuing this vision of credibility would look very different from an algorithmic approach tailored to a witness's propensity to lie. Rather than using big data to hone in on the risk that a witness would testify falsely, this worthiness algorithm would give shape and meaning to the concept of credibility itself by identifying the behavior and/or attributes that would endow a witness with credibility.

A major concern with the proposed use of an algorithm to replace or inform any legal process is that the algorithm will prove equally or more problematic than the system it replaces.[229] The country has a history of embracing algorithmic tools as seemingly objective mechanisms for eliminating bias, as was the case when Congress first embraced credit-scoring.[230] Yet, as that example illustrates, algorithms can perpetuate and introduce new biases into existing systems. As was already clear at the time Congress first embraced credit scoring, its focus on financial risk means using whatever factors prove predictive of such

229. As Professor Huq notes in the context of algorithmic interventions in criminal justice, "[a]doption of machine learning . . . changes the scale, reach, and operation of state power." Huq, *supra* note 49, at 1065. This has the potential to introduce new forms of discretion or reify old forms of bias. *Id.* at 1066. Although, as Huq says, it may be "premature" to conclude that algorithmic tools will worsen problems like racial bias in the criminal law system, concern over this danger motivates many scholars writing about algorithms in the law. *Id.* at 1077; *see, e.g.*, Stevenson, *supra* note 1, at 341–69 (studying use of pretrial risk assessment algorithm in Kentucky to assess efficiency and impact on racial disparities); *Bias In, Bias Out*, *supra* note 1, at 2268–70 (describing difficulties in measuring and achieving equity along racial lines in application of risk prediction tools); *see also* RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE 3, 4 (2019) ("[C]odes act as narratives" and "operate within powerful systems of meaning that render some things visible, others invisible, and create a vast array of distortions and dangers.").

230. *See, e.g.*, Capon, *supra* note 17, at 86 ("Congress embraced credit scoring systems, believing that their claimed objectivity offered advantages in enforcement of the Equal Credit Opportunity Act . . . ."); OLEGARIO, *supra* note 16, at 147 (describing how "even in the early 1970s, race was still a standard question on credit applications"); Louis Hyman, *Ending Discrimination, Legitimating Debt: The Political Economy of Race, Gender, and Credit Access in the 1960s and 1970s*, 12 ENTER. & SOC'Y 200, 226–29 (2011) (describing in context of the Equal Credit Opportunity Act congressional ignorance of "how statistical models function" and the ease with which "correlated variable[s] . . . acquire the predictive power of the protected category"). Hyman writes that "[i]n passing legislation geared to a world of prejudiced loan officers, Congress made the newer computer-driven credit models actually more discriminatory." *Id.* at 228.

risk.[231] The creators of credit scores who testified at hearings on the Equal Credit Opportunity Act made clear that if it were legal, and race were predictive of the risk of default, race would be used directly in creating the scores.[232] Even without using race itself as a factor, credit scores reflected longstanding patterns of unequal opportunity to build or maintain credit along lines of race and to a lesser extent gender.[233]

This limitation is not unique to credit scoring. As Sandra Mayson has written, "[a]ll prediction functions like a mirror."[234] The whole idea of algorithmic prediction is that we can take things that have already happened and, with the right tailoring, use them to predict the future.[235] In the context of crime prediction, "[i]ndividual traits that correlated with crime commission in the past" are expected to "correlate with crime commission in [the] future."[236] Because we are limited to "project[ing] history forward," our thoughts about future risk will inevitably be channeled through the distorting lenses of race, gender, and class, among other things.[237]

231. *See, e.g.*, Capon, *supra* note 17, at 86–87 (describing data employed by credit scoring algorithms as skirting characteristics, such as race, that are proscribed by Congress, but using other characteristics that "act as surrogates . . . for race," such as zip code); *see also* Raymond H. Brescia, *Subprime Communities: Reverse Redlining, the Fair Housing Act and Emerging Issues in Litigation Regarding the Subprime Mortgage Crisis*, 2 ALB. GOV'T L. REV. 164, 172–79 (2009) (describing disproportionate impact of subprime mortgage crisis on communities of color, due in part to history of discrimination in access to credit); Allen J. Fishbein & Patrick Woodall, *Exotic or Toxic? An Examination of the Non-Traditional Mortgage Market for Consumers and Lenders*, CONSUMER FED'N OF AM. 24 (2006), http://www.consumerfed.org/pdfs/Exotic_Toxic_Mortgage_Report0506.pdf (finding race of borrower affected likelihood of receiving a non-traditional mortgage).

232. *See* Capon, *supra* note 17, at 85–86. Algorithmic tools used in medicine do overtly use race as a factor, resulting in disadvantages for Black patients. Gina Kolata, *Many Medical Decision Tools Disadvantage Black Patients*, N.Y. TIMES (June 17, 2020, 1:31 PM), https://www.nytimes.com/2020/06/17/health/many-medical-decision-tools-disadvantage-black-patients.html.

233. *See* OLEGARIO, *supra* note 16, at 147–48 (describing "barriers" in applying for credit faced by minority communities in the United States leading up to the 1970s). Louis Hyman describes this phenomenon as a form of vicious cycle in which ignoring "the core inequalities of the labor market" and focusing on automating credit means that those who historically had access to credit will continue to do well in the labor market leading to a cycle of more credit and more labor market success. *See* Hyman, *supra* note 230, at 227–30. Those with little access to credit, in contrast, will continue to face difficulty in the labor market leading to less credit and less labor market success. *Id.* at 229–31. All the credit score can do is predict the likelihood of default. It is not designed to rectify the underlying causes of income instability and inequality that is a systemic cause of poor credit. *Id.* at 229.

234. *Bias In, Bias Out*, *supra* note 1, at 2224.

235. *Id.*

236. *Id.*

237. *See id.*

Of course, this is only a problem if one expects an algorithm to be debiasing. A worthiness algorithm would not necessarily aim to reduce bias.[238] Instead, it might intentionally capture current doctrine when choosing the characteristics that will constitute credibility. Under this approach, a witness's outward appearance would be of preeminent importance. Similarly, a witness's prior convictions would also be given great weight. Other factors, such as witness coherence, a poor reputation for truthfulness, or an atypical narrative might also contribute to such an algorithm.

Experience teaches that this type of worthiness algorithm, one trained to account for demeanor and prior convictions, might produce results that vary along lines of race.[239] Such an algorithm might also be expected to reflect superficial distinctions based on gender, level of education, speech patterns, English language fluency, and elements of a witness's appearance, such as whether he or she is dressed formally enough, to name just a few. In response, witnesses might be expected, where possible, to alter their behavior in order to be perceived as worthy of belief. This would be in keeping with the current practice, in which defense attorneys may keep spare clothes for clients in their trunks and police officers are coached on how to seem believable in court.[240]

There are features of current credibility jurisprudence that might be enhanced by a worthiness algorithm. Notably, such an algorithm might aid human decisionmakers in sorting out the body of law that views certain prior convictions as indicative of credibility and others as irrelevant. The common law in this area has carried forward an early distinction between crimes involving deception or theft, which are seen

238. Another way to understand this is to recognize that a worthiness algorithm is not the kind of algorithm that Bernard Harcourt critiques for "mak[ing] possible 'right' and 'wrong' answers" and thereby creating simplistic moral judgments that can "alleviate our scruples." BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 188 (2007). A worthiness algorithm does not produce results that can be "validated, tested, replicated" as is the case for an objective metric. *Id.* Instead, it creates its own right answers.

239. *See* Kolata, *supra* note 232 (describing programming of medical algorithms to set different thresholds for medical intervention depending on race).

240. *See, e.g.*, Bea Bischoff, *When Being a Good Lawyer Means Dressing Your Clients*, RACKED (Jan. 18, 2018, 10:02 AM), https://www.racked.com/2018/1/18/16900864/public-defender-clothing-accused-donations (describing defense attorney's practice of keeping clothes for her clients in her car). Officer training takes various forms, but police officers or departments can take credibility trainings offered by numerous for-profit enterprises. *See, e.g.*, *Courtroom Testimony, Preparation, and Credibility*, L. ENF'T LEARNING, https://lawenforcementlearning.com/course/courtroom-testimony-preparation-and-credibility/ (last visited Oct. 28, 2021) (listing as course objectives: "[e]xplain[ing] how your body language affects your demeanor" and "[l]ist[ing] the five areas that directly affect your appearance as credible").

as bearing on credibility, and crimes of violence, which are not.[241] In jurisdictions that permit impeachment with crimes involving "moral turpitude," the worthiness algorithm might learn from existing precedent that sex crimes are thought of as pertinent to a witness's credibility, meaning that those convicted of sexual assault or prostitution would be more vulnerable to impeachment than those convicted of a different offense like aggravated assault.[242] An algorithm might also facilitate drawing distinctions between prior convictions that have hitherto been seen as too costly. For example, the worthiness algorithm might have the capacity to comb through the records of prior convictions, making it easier to determine if "the conviction required the fact-finder to find, or the defendant to admit, an act of dishonesty or false statement."[243] In this way, FRE 609(a)(2) and its state analogues, which require that courts admit prior convictions involving dishonesty or false statements, might be dramatically expanded.

Of course, once decision-makers see the effects of a worthiness algorithm programmed using data from the common law of credibility, they may wish to alter the formula. Perhaps, as in China's regime, we should take account of behavior deemed socially productive and give a credibility boost to people who behave in ways that are perceived to be beneficial. This would not have to come in the form of credibility bolstering, which is permitted only if a witness's credibility has been attacked.[244] Rather, a worthiness algorithm itself could be programmed to offset an unfavorable demeanor or prior conviction with a witness's volunteerism or donations to charity.

As is the goal of social credit scoring in China, a worthiness algorithm might encourage desirable behaviors in exchange for a credibility boost in the courtroom. The system could also be revised to select new metrics that would lower a person's credibility score and give people incentives to avoid behavior not accounted for by current doctrine. With access to the rich consumer data now in the market, these new metrics could include information as mundane as whether the witness invested in safety gates as the parent of a baby or as complex as how much time the witness spent engaging with conspiracy theories on Facebook.[245] Buying

241. *See Credibility by Proxy*, *supra* note 2, at 192–203.

242. *Moral Turpitude*, *supra* note 118, at 1005, 1011–15; *Unchaste and Incredible*, *supra* note 5, at 1893–97.

243. FED. R. EVID. 609 advisory committee's note to 2006 amendment.

244. *See, e.g.*, FED. R. EVID. 608(a) ("[E]vidence of truthful character is admissible only after the witness's character for truthfulness has been attacked.").

245. Consumer behavior has also been tied to risk-predictive credit-scoring. For example, the purchase of floor protectors is a well-known example from the world of financial credit. Charles Duhigg, *What Does Your Credit-Card Company Know About You?*,

the gates may afford a credibility boost while spending too many hours engaging with Facebook conspiracists might lower it.

Rather than debiasing or predicting truthfulness, this worthiness algorithm approach would harness big data in the service of a system that functions much like today's common law of credibility. It would focus on external markers of credibility that reflect social conceptions of what makes a witness worthy of being believed. Such a system offers credibility as an end in and of itself, one that is available to those who perform it and unavailable to those who will not or cannot conform.

B. The Lying Propensity Algorithm

If instead we look to the credit score's evolution, we find an entirely different model for credibility assessment. Courtroom credibility assessment purports to be a predictive measure of a witness's propensity to lie, but in operation it is more like the early, reputation-based days of credit-scoring and can make no claim of predictive accuracy. Instead, as described above, the courtroom still emphasizes a witness's worthiness of belief as performed in ways that have remained static over the past two hundred years.[246] What would happen if we tried to unbundle "propensity for lying" from the credibility package and focus on that question to the exclusion of others? At the moment, it might mean we give fact-finders less information.[247] In the future, it might mean we use an algorithm to predict a witness's risk of lying on the stand.

If we choose to focus our algorithmic gaze solely on predicting the risk of lying in court, the calculus would be notably different from a worthiness algorithm. This choice to create what I term a "lying propensity" algorithm would reflect the justificatory impulse of policymakers and judges who assert that credibility impeachment in the courtroom with prior convictions is a way to determine a witness's propensity for truthfulness. Although this vision of credibility is more tailored to lying, the notion of a *propensity* for lying still involves a generalization about a trait of character. Unlike lie-detection algorithms,

N.Y. TIMES (May 17, 2009), https://www.nytimes.com/2009/05/17/magazine/17credit-t.html. Using data on purchase history and credit card payments, an insurer concluded that buying felt pads to put under furniture and protect the floor signaled that someone was a good credit risk. *Id.*

246. *Credibility by Proxy*, *supra* note 2, at 221.

247. On this point, I have previously proposed eliminating most impeachment based on prior convictions or past bad acts because such evidence is not predictive of truthfulness or its opposite on the witness stand. *Id.* at 221–25.

it does not determine that a person is actually telling a lie in the moment.[248]

Although current commercial lie-detection tools have a slightly different aim—allegedly detecting lies rather than predicting the risk of lying—critics have pointed out that algorithms like the Avatar and Virtual Policeman, which purport to be able to "read" people, pose significant risks of discrimination and inaccuracy.[249] Indeed, a frequent criticism of these algorithms is that they are not being trained on "data set[s] as diverse as the one[s they] will be evaluating in real life."[250] Similarly, if the judicial system were to rely on a lying propensity algorithm, much would depend on how it was created. Accuracy would be important, and bias would be problematic. Risk prediction algorithms in use in the criminal justice system have been criticized along both dimensions with critics focusing on the data upon which such algorithms are built.[251] Because they make predictions using data generated by the criminal justice system, they reflect inequality and biases within the system that cannot be eradicated simply by shifting from subjective human judgment to an algorithm. Credit-scoring algorithms have been critiqued along similar lines.[252]

Such algorithms may provide the opportunity to "diagnose" the causes of disparities within the system, but they have so far failed to

248. This is in contrast to the memory-detection technology envisioned by Emily Murphy and Jesse Rissman, which would function like a lie-detector. *See* Emily R.D. Murphy & Jesse Rissman, *Evidence of Memory from Brain Data*, J.L. & BIOSCIENCES, Dec. 18, 2020, at 1, 5.

249. Devlin, *supra* note 1. As leading psychology researcher Lisa Feldman Barrett explains, the underlying assumption of such algorithms, that people can universally recognize emotions in facial expressions, is unsupported by science. Lisa Feldman Barrett, *What Faces Can't Tell Us*, N.Y. TIMES (Feb. 28, 2014), https://www.nytimes.com/2014/03/02/opinion/sunday/what-faces-cant-tell-us.html. "So-called emotional expressions," she and her co-authors write, "are more variable and context-dependent than originally assumed," yet despite this, technology companies are "spending millions of research dollars to build devices to read emotions from faces." Lisa Feldman Barrett et al., *Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements*, 20 PSYCH. SCI. PUB. INT. 1, 47–48 (2019). By focusing exclusively on faces, these algorithms also have great potential for bias. Devlin, *supra* note 1. For example, critics of the European Union's experiment with policing lies at the border using artificial intelligence have questioned whether the system might discriminate against travelers on the "basis of their ethnic origin." Gallagher & Jona, *supra* note 1. One crucial question, as with all algorithms, is "[w]ho sets the parameters to establish that a certain subject is lying or not lying?" *Id.* (quoting Giovanni Buttarelli, head of the European Union's data protection watchdog group).

250. Bittle, *supra* note 30.

251. Okidegbe, *supra* note 57, at 332–33.

252. *See supra* note 219.

address them.[253] Would lying propensity algorithms follow a similar path? These algorithms would be similar to existing risk assessment algorithms in the sense that they would aim to offer a likelihood for a future event, in this case lying on the witness stand. It is possible that, like credit-scores, such tools might carry forward the distortions of the past. If, for example, developers were to decide that being convicted of or held liable for fraud indicates a propensity for lying, an algorithm built using data from the criminal or civil justice system would import the biases of those systems.

What seems more likely is that an algorithm with genuine claims to be an evidence-based tool for measuring the risk of lying could not be generated from existing government or private databases. Like facial recognition tools, these algorithms would have to be created in a laboratory setting where researchers could evaluate their accuracy, in this case by assessing what data from a person's past is actually indicative of a risk they will lie while testifying in court.[254] This might be salutary in moving the system away from past inequities, but it comes with its own difficulties, including the possible biases of the programmers, the challenges of recruiting truly heterogenous subjects, and the difficulty in simulating the courtroom setting.[255]

Social factors might also continue to matter. If a lying prediction algorithm shows us that certain proxies for lying vary along racial, gender, or other lines, it might reveal that we do not have the kind of shared understanding of what counts as deceptive that is a predicate to assigning a "propensity" for lying to a given witness. Well-documented difficulties in assessing the plausibility of narratives would also complicate the task of creating an algorithm with claims to accuracy at predicting the propensity for lying.[256] There may simply be disagreement about how reality looks, what is truth, and what is not. Lying propensity

253. *Bias In, Bias Out*, *supra* note 1, at 2284–86.

254. Needless to say, data showing whether a witness was lying cannot be reverse-engineered from trial records except in the rare case when exonerating evidence shows witnesses at a trial to have been lying or mistaken.

255. This assumes, for the sake of argument, that such an algorithm could be created in a trans-substantive form. It might be, however, that programmers would need to create different algorithms for use in civil or criminal trials, or depending on whether the witness was a party, for example.

256. *See, e.g.*, Kim Lane Scheppele, *Foreword: Telling Stories*, 87 MICH. L. REV. 2073, 2083 (1989) (describing how stories rejected as false in the courtroom may in fact "be accurate versions of events that grow from experiences different from the experiences of those who are doing the choosing"); Steven L. Winter, *The Cognitive Dimension of the* Agon *Between Legal Power and Narrative Meaning*, 87 MICH. L. REV. 2225, 2229 (1989) (describing how the human mind uses imagination, and thus narrative, to understand the world and how narrative institutionalizes social meaning in the legal context).

algorithms might also expose a very different set of actors who might be presumed untrustworthy. While the system at present presumes that police are credible,[257] even giving jury instructions to that effect,[258] an algorithmic tool might confirm recent observational data indicating that police testimony should instead be treated with skepticism.[259] Conversely such an algorithmic tool might confirm the assumption that witnesses with prior criminal convictions have a higher propensity for lying.[260]

257. *See* Anna Lvovsky, *The Judicial Presumption of Police Expertise*, 130 HARV. L. REV. 1995, 1998–99 (2017).

258. Vida B. Johnson, *Silenced by Instruction*, 70 EMORY L.J. 309, 317, 333–34 (2020) (arguing that typical jury instructions that tell jurors not to give police officer testimony more or less weight simply because the witness is an officer amount to a credibility boost for the officer).

259. New York has begun taking the problem of police lying on the witness stand so seriously that the court system recently created a program requiring courts to collect data on how often judges "find that police officers provided testimony on the stand that was deemed lacking in credibility." Andrew Denney, *NY Court System Launches Program to Track Evidence Suppression, Police Credibility*, LAW.COM (Jan. 14, 2019, 5:34 PM), https://www.law.com/newyorklawjournal/2019/01/14/ny-court-system-launches-program-to-track-evidence-suppression-police-credibility/?slreturn=20210819204744; *see also* Joseph Goldstein, *Officers Said They Smelled Pot. The Judge Called Them Liars*, N.Y. TIMES (Sept. 12, 2019, 10:07 AM), https://www.nytimes.com/2019/09/12/nyregion/police-searches-smelling-marijuana.html (describing an opinion by a Bronx judge that "accuse[d] police officers of routinely lying to cover up illegal searches"); Anne Schindler, *With Officer's Credibility in Question, Attorneys Say Key Evidence Must be Tossed in Navy Chief Petty Officer Murder Trial*, FIRST COAST NEWS (Feb. 5, 2021, 10:07 PM), https://www.firstcoastnews.com/article/news/crime/defense-for-man-accused-of-killing-ex-fiancee-questions-credibility-of-jso-officers-testimony/77-3625f864-94a8-42a9-a469-55dc62ecc6ab; Ali Rockett, *Officers' Testimony on Thursday Differs from RPD Account of Police Shooting on New Year's Eve*, RICHMOND TIMES-DISPATCH (Feb. 12, 2021), https://richmond.com/news/local/crime-and-courts/officers-testimony-on-thursday-differs-from-rpd-account-of-police-shooting-on-new-year-s/article_ddf010ef-d10d-579d-a3d1-9626c8e36db5.html; Justin Fenton, *Ex-Baltimore Police Officer Sentenced for Lying to the FBI*, WASH. POST (Feb. 14, 2021, 7:02 PM), https://www.washingtonpost.com/local/public-safety/ex-baltimore-police-officer-sentenced-for-lying-to-the-fbi/2021/02/14/b6c42192-6afb-11eb-9f80-3d7646ce1bc0_story.html (noting how a Baltimore police officer's lie to the FBI "stymied" an investigation and how "numerous other officers would have testified to the truthfulness of [the lying officer's] statements").

260. Such a tool would have to be approached with caution when the subject is a criminal defendant. When aimed at a criminal defendant, it would put pressure on a truism that the United States legal system has long ignored, but which the United Kingdom has written into their evidentiary rules—that the question of a criminal defendant's credibility is inextricable from the question of guilt. Campbell v. R [2007] EWCA Crim 1472, 200605014 C4 [30]. England and Wales have discontinued the practice of impeaching criminal defendants with prior convictions. *Id.* at [29]–[32]. The logic behind this decision is that if we accept that in a criminal case the defendant has every incentive to lie if guilty, it is not possible to admit evidence that only goes to the credibility of the defendant. *See id.* at [30]; *see also* MIKE REDMAYNE, CHARACTER IN THE CRIMINAL TRIAL 6–9 (2015). That evidence necessarily also goes to the question of guilt. Campbell v. R [2007] EWCA Crim 1472, 200605014 C4 [30]. Such evidence is also cumulative and unhelpful because the jury should

More broadly, a lying propensity algorithm might bear out or disprove insights from procedural justice research: If one trusts the justice system, might one be less likely to lie in court than someone who for very good reasons has little faith in the system? Would a lying propensity algorithm suggest that people with more negative interactions with police, for example, are more likely to lie in court? Answering these and similar questions, in turn, might be instructive of broader social conditions and of steps we might take to address them.

Apart from these potential hurdles and benefits, any lying propensity algorithm would have to start by answering a basic definitional question: What qualifies as deception? This question has bedeviled philosophers for centuries.[261] No matter the definition, a witness's social or cultural background might contribute to different conceptions of how to tell a story, including what qualifies as a lie or deceptive. Thus, the system could predict a propensity for deception that a different lens would interpret as simply part of the arc of a genuine recounting of events. Even the most thoughtful creators would have great power in creating the epistemic norms about what qualifies as untruthful.

Lying propensity algorithms would also gloss over *why* lies are told. Although witnesses swear to tell the truth, the legal system cannot and does not seek to eliminate every lie. There is a certain tolerance within the system for forms of lying or lies told for particular reasons. For example, perjury is defined narrowly and perjury charges are rarely brought.[262] The system has also tolerated actors who engage in what I have termed "systemic lying" in order to remedy real or perceived injustices. [263] "Testilying," the notorious practice of police inventing testimony at suppression hearings in order to justify a stop or search, is one example.[264] Another is the widespread practice, prior to the advent of no-fault divorce, in which divorcing couples lied about their

determine guilt based on the evidence admissible on that question, rather than by finding that the defendant is probably a liar and is therefore also guilty since only a guilty defendant would lie. *Id.* For this same reason, an algorithm that can predict the risk of lying for a criminal defendant would also provide evidence that is logically inextricable from the defendant's guilt.

261. *See generally* SISSELA BOK, LYING: MORAL CHOICE IN PUBLIC AND PRIVATE LIFE 10–16 (1978).

262. *Id.* ("[P]rosecutions for perjury are rare and, therefore, the perjury sanction would not seem to be a very effective deterrent to criminal defendants lying at trial . . . ."); Stuart P. Green, *Lying, Misleading, and Falsely Denying: How Moral Concepts Inform the Law of Perjury, Fraud, and False Statements*, 53 HASTINGS L.J. 157, 173–75, 177, 180–81 (2001) (arguing that law of perjury and false statement tracks moral intuitions about the relative harm of lying and misleading).

263. Julia Simon-Kerr, *Systemic Lying*, 56 WM. & MARY L. REV. 2175, 2178–79 (2015).

264. *Id.* at 2201, 2205.

relationships to satisfy the terms of divorce statutes that required them to allege fault.[265] A lying propensity algorithm would not go as far towards eliminating this space as an actual lie-detector might, but it might still constrain this space for lying in ways that might be salutary, especially if we believe such spaces often benefit those with more power within the system. On the other hand, making it more costly to lie could also make it harder to remedy injustices that can be produced by the formal constraints of the system.

A lying propensity algorithm, even if it is extremely accurate at judging the risk that a witness will lie, would bring up many of the same issues that arise when probabilistic DNA or complex mathematical evidence is introduced in a criminal trial. For example, how would an expert on lying propensity explain to the judge or jury what the statistics mean? How can we ensure that due process rights are not violated when we introduce possibly inaccurate algorithmically-generated credibility evidence that the defendant is lying?

Even if this difficulty were to be eliminated, say in a world in which lying propensity algorithms reached extreme accuracy, there are other reasons for shying away from this approach. In their article on brain-based memory detection technology, Emily Murphy and Jesse Rissman argue that even if we could detect lies themselves with perfect accuracy, using such technology would not "respect[] the dignity and personhood of the witness as the narrator of their own memory and experience."[266] Murphy and Rissman argue that acknowledging the witness as narrator and as a person is also intertwined with the jury's traditional role as fact-finder.[267] Any algorithmic intervention into credibility assessment, whether by virtue of a memory-detection system as described by Murphy and Rissman, a lying prediction tool, or a worthiness algorithm would upset this relationship, interposing a machine and its creator between the witness and judge or juror. It would also pose a fundamental challenge to the foundational tenet of U.S. law that "the *jury* is the lie detector."[268] Still, in a world in which current credibility jurisprudence itself may often detract from the jury's lie-detection function, keeping algorithmic tools out is not necessarily the best course. Further, by offering what amounts to worthiness points to certain witnesses, the system has already failed at according "respect and dignity" to many who come before it.

265. *Id.* at 2178.

266. Murphy & Rissman, *supra* note 248, at 50.

267. *Id.* at 49–50.

268. United States v. Scheffer, 523 U.S. 303, 313 (1998) (quoting United States v. Barnard, 490 F.2d 907, 912 (9th Cir. 1973)).

Finally, there is one way in which predicting the risk of lying in court using an algorithm could be an unalloyed good. Such an algorithm would reconcile what the system claims it is doing with the actual practice in the courtroom. If we take the assertions that credibility impeachment is meant to offer information about a witness's propensity for truth at face value, then having actual evidence of that propensity would be an improvement over the pretense that unrelated evidence, like prior convictions or demeanor, serve this function.

V. CONCLUSION

Ultimately, the thought experiment in this Article tells us as much about the shortcomings of the current system of credibility impeachment as it does about its algorithmic future.

Problems with inaccuracy, bias, and an inability to break free of the past are not unique to algorithmic risk prediction.[269] The U.S. legal system currently makes claims for credibility impeachment's ability to predict lying that are unjustified and almost certainly unjustifiable. Rather than assist in identifying false testimony, these rules create the very credibility they claim to assess. Being worthy of belief, through demeanor or a clean record, is the essence of legal credibility. Because so much of how we judge credibility is both opaque and subjective, the system tolerates assumptions about credibility that are often racist or sexist, and that are very difficult to challenge.[270]

Algorithmic tools can offer a more "precise mirror" into the system's workings than the opaque workings of credibility doctrine will ever provide.[271] For this reason alone, it is worth considering the potential of either a lying prediction or worthiness algorithm as we contemplate a future in which algorithms play an ever-larger role in our courts. Should we seek to reinforce either of these visions algorithmically? The analogy to China's social credit system may make this seem an easy choice. A program that tallies points for whether your neighbor thinks you lie a lot or in which the clothes you wear to a legal proceeding could cause your credibility score to decrease seems antithetical to central democratic values such as autonomy and privacy.

269. Professor Mayson makes this point as well. She writes that "[a]ctuarial risk assessment . . . has revealed the racial inequality inherent in *all* crime prediction in a racially unequal world." *Bias In, Bias Out*, *supra* note 1, at 2225.

270. *Credibility by Proxy*, *supra* note 2, at 166–67; *Uncovering Credibility*, *supra* note 2, at 595.

271. *Bias In, Bias Out*, *supra* note 1, at 2225.

Yet this is to a large degree how credibility currently manifests in our system of law, and it does so without much reasoned discourse in the law and without any public debate over its contours. Worthiness-centered credibility has occupied a central place in the law since this nation's founding. Indeed, this fact might be taken to suggest that there are reasons to embrace credibility's possibilities as a method of social control or cohesion in the future. Could it be beneficial to have a public algorithm that awards credibility points for certain acts and takes them away for others, particularly if those inputs are debated by and known to the public? Might such an approach offer a mechanism for recreating some shared sense of what it means to be credible, and perhaps of decreasing contestation over truth itself?

Another possibility, of course, is that we would rather try to eliminate from the law that part of credibility that is simply a mirror of and a prescription for social conformity, however that manifests.[272] Instead, the law might prefer algorithms that can identify a "propensity for lying." Though such a tool seems far-fetched, developers are already making claims to have developed a close cousin, deception detecting algorithms. This algorithmic approach also has both costs and benefits. Matching doctrinal claims about the predictive nature of credibility impeachment with reality is one potential benefit, as are potential improvements to truth-seeking. Less salutary, however, might be the degree to which programmers could choose what counts as deception, the inability of lying propensity algorithms to account for the reasons lies might be told, and the potential for lying propensity algorithms to reveal group-based patterns in likely courtroom liars.

Of course, there are still other avenues through which we could approach the question of algorithmic credibility. One of these is to reject algorithmic intervention in this realm altogether. Yet rejecting algorithms should not mean embracing the stasis that has overwhelmed this area of law for more than a century. If the thought of algorithms taking over the function of assessing a witness's worthiness or propensity of lying is distressing, our credibility jurisprudence with its emphasis on external appearance and prior convictions and its false claims to be assessing risk should be equally if not more so. Ultimately, by imagining what might be revealed in the unsparing mirror of an algorithm, we can see that change is needed, algorithmic or not.

272. I have made such an argument in prior work. *See Credibility by Proxy*, *supra* note 2, at 204–05, 207–08.