Stories, Statistics, and the Regulation of Alternative Data

Aniket Kesari[*] *and Mark Verstraete*[**]

ABSTRACT: Financial technology has long relied on data like an applicant's current indebtedness to make decisions about who gets access to new credit, but AI is now enabling credit determinations based on some unusual inputs. This "alternative data"—or data that is not intuitively connected to creditworthiness—includes information like a consumer's online shopping habits, whether they paid their rent and utility bills, and even how many friends they have on social media. The use of this kind of data has exciting potential to expand access to credit but exemplifies a well-known feature of machine learning: It works by finding nonintuitive relationships in large datasets.

Regulators and industry have recognized both the potentials and pitfalls of alternative data. As nascent discussions emerge around appropriate uses of this alternative data, there is little consensus about the ideal shape of regulations for alternative data. This Article offers a path forward for alternative data regulation. In doing so, it surfaces a central tension that regulators must resolve—the contradiction between the desire for intuitive stories and the reality of how these technologies work. For instance, scrolling quickly through an online contract could be indicative of carelessness about legal obligations, which may make someone a risky creditor. Intuitive stories have long formed the basis for social oversight of credit lending. These stories have enabled us to determine that information such as zip codes and medical debt are unfair to include in credit calculations, but that using on-time payment history is acceptable.

Yet the emergence of machine learning models trained on alternative credit data challenges classic assumptions about the connection between data and

[*] Associate Professor, Fordham Law School.

[**] Assistant Professor, Seattle University School of Law. Both Authors contributed equally. We would like to thank Talia Gillis, Katherine Strandburg, Olivier Sylvain, Jane Bambauer, Derek Bambauer, Sonia Katyal, Salome Viljoen, Chinmayi Sharma, Brett Frischmann, Nikita Aggarwal, Elettra Bietti, David Stein, Benjamin Sobel, Katrina Geddes, Stav Zeitouni, Tomer Kenneth, Seth Endo, Maggie Chon, and participants at NYU's Information Law Institute Fellows Workshop for feedback and inspiration. Special thank you to the Carol Evans, Patrice Ficklin, and other CPFB officials who inspired this work at the Cardozo Symposium on Automating Bias. We also would like to thank Christian Bono and Jane Zech for excellent research assistance.

the narratives we tell. For example, intuition is broken if a person "games" data by changing a proxy for creditworthiness without changing the underlying characteristic a system designer intends to measure. By uncovering this potential for breaking causation, this Article surfaces the true normative stakes of regulating alternative credit data. The stakes are not about "connectedness" between data and decisions, as existing regulations imply.

Introduction

Financial technology ("fintech") startups are using machine learning to find correlations to creditworthiness in unusual places. These firms are making credit decisions based on information like how quickly one scrolls through a webpage,[1] what they buy,[2] and even how many friends they have on social media.[3] This "alternative data"—or data that falls outside the scope of traditional credit reports—is emblematic of shifts in credit lending and other industries where businesses are using seemingly unrelated data to make decisions about individuals.[4]

Within credit lending, alternative data varies in how related data appears to creditworthiness. Sometimes this data is intuitively connected to an assessment about a person's likelihood to pay back a loan. For instance, someone who pays their rent or utility bills on time is likely to pay other debts on time as well.[5] Other data is less intuitively related to creditworthiness. In these cases, it is difficult to craft a plausible story about why a variable predicts credit risk. For example, an individual's IP address has little obvious connection to creditworthiness.[6] Nonetheless, machine learning offers businesses the ability to

1. *See* Mikella Hurley & Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 Yale J.L. & Tech. 148, 164–83 (2016) (discussing ZestFinance).

2. Tom Sullivan, *6 Types of Alternative Credit Data for Better Loan Decisions*, Plaid (Feb. 25, 2025), https://plaid.com/resources/lending/alternative-credit-data [https://perma.cc/27T2-4B4A] (discussing transactional data).

3. Insider Intelligence, *Here's How One Alt Lender Is Using Alternative Underwriting Data*, Bus. Insider (Oct. 10, 2017, 8:20 AM), https://www.businessinsider.com/heres-how-one-alt-lender-is-using-alternative-underwriting-data-2017-10 [https://perma.cc/Q26E-USER].

4. Bryan Howard, *Unlocking Hyper-Personalization: 20 Alternative Data Sources for Insurance*, FICO Blog (Oct. 4, 2023), https://www.fico.com/blogs/unlocking-hyper-personalization-20-alternative-data-sources-insurance [https://perma.cc/B6G7-CTQJ] (describing insurers' use of alternative data sources—including IoT, geolocation, social behavior, public records, and more—for hyper-personalized underwriting and risk assessment).

5. *Rent Payment History Offers Greater Predictability into Consumer Credit Performance*, TransUnion (Dec. 7, 2021), https://newsroom.transunion.com/rent-payment-history-offers-greater-predictability-into-consumer-credit-performance [https://perma.cc/7VR2-69QB].

6. *Cf.* Terri Bradford, *"Give Me Some Credit!": Using Alternative Data to Expand Credit Access*, Fed. Rsrv. Bank Kan. City (June 28, 2023), https://www.kansascityfed.org/research/payments-system-research-briefings/give-me-some-credit-using-alternative-data-to-expand-credit-access [htt

uncover relationships between IP address and credit risk despite our collective inability to weave a sensible story that explains the connection.

This tension between empirically predictive relationships and intuitive narratives about legitimacy has important implications for machine learning and law. As Andrew Selbst and Solon Barocas highlight, "What should be clear by now is that intuition is the typical bridge from explanation to normative assessment."[7] Selbst and Barocas also offer the central insight that machine learning excels at finding *nonintuitive* correlations—and these correlations are often more predictive than intuitive ones.[8]

Credit lending provides a uniquely instructive domain to explore this tension because the legal regulation of credit lending has historically relied on intuitive narratives to determine which data is an acceptable basis to make credit decisions. For example, fair lending laws explicitly forbid the use of certain features, such as race or gender, because society deems these bases fundamentally unfair—even though these attributes could be empirically correlated with the likelihood an individual will repay a loan.[9] Modern credit lending laws thus focus heavily on "causal" justifications for features such as prior default history, debt-to-income ratio, and length of credit history.[10] In practice, however, these justifications have often amounted to "strong correlations" accompanied by compelling intuitive stories rather than proven causal relationships in a statistical sense.[11]

The rise of alternative data exposes the fragility of the law's assumption about the stable relationship between intuition and empirical prediction. Recent regulatory actions highlight this dynamic vividly. Consider, for instance,

---

ps://perma.cc/2PQE-DVPC] ("Moreover, the CFPB has noted that some alternative data (such as social circles) may not be related to a person's own financial conduct, and the use of these data could make it more difficult for people to improve their credit standing." (citation omitted)).

7. *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1128 (2018).

8. *Id.*

9. Kenneth Lipartito argues,

> The upshot of the legislation and investigations was to make computer data banks and behavioral or transactional information the answer to concerns about privacy or prejudice. If women or African Americans were denied credit because wittingly or consciously credit evaluators took race, marital status, or personal lifestyle into account, then credit scoring based strictly on patterns of credit use, plus a limited number of individual attributes (such as address or occupation) provided a defense against charges of discrimination.

Kenneth Lipartito, The Narrative and the Algorithm: Genres of Credit Reporting from the Nineteenth Century to Today 32 (Jan. 6, 2010) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1736283 [https://perma.cc/9RCP-SXMU].

10. We use "causal" here the way credit regulators and certain scholars have conceptualized the term. As we explain later, the bar for causality in statistics is much higher and requires specialized frameworks and methods. "Causal" in the credit lending context is closer to "predictive," which is why it is a ripe area for studying how machine learning should be regulated.

11. *See infra* note 146 and accompanying text.

the Consumer Financial Protection Bureau's ("CFPB") recent decision to ban medical debt as an input in credit scoring.[12] The CFPB's decision was arguably not based on accuracy concerns as medical debt appears to be predictive of creditworthiness—someone who is unable to pay medical bills will also be unable to pay an auto loan.[13] Instead, the decision is principally based on the desirability of the causal story that makes medical debt predictive of creditworthiness. That is, a person accrues medical debt when they have a sudden health emergency and are unable to pay the associated costs. And here, the CFPB's decision to explicitly ban the use of this data embodies the principle that it is unfair to deny credit to individuals who have befallen a medical emergency.[14]

This Article offers a roadmap to financial regulators that are attempting to craft viable regulations for alternative data in credit models. Many financial regulators have sought to capture the benefits of alternative data while minimizing the potential risks.[15] In doing so, regulators from across several agencies have issued various public statements that implicitly assume the importance of intuitive stories in credit lending. They have coalesced around acceptable examples of alternative credit data like utility payment

---

12. *CFPB Finalizes Rule to Remove Medical Bills from Credit Reports*, CONSUMER FIN. PROT. BUREAU (Jan. 7, 2025), https://www.consumerfinance.gov/about-us/newsroom/cfpb-finalizes-rule-to-remove-medical-bills-from-credit-reports [https://perma.cc/R6UT-JDMC].

13. Although the CFPB cites a 2014 study suggesting that medical debt is not predictive of credit default, this position has been met with some criticism. A Congressional Research Service report notes:

> To justify excluding medical debt from credit reports, the CFPB relied upon a 2014 CFPB study that stated that medical collections were "not equally predictive of delinquency" as non-medical collections. The CFPB later interpreted this result to subsequently argue that medical debts have "little to no predictive value." This interpretation of the original research has been disputed, with the American Bankers Association pointing out that this difference was driven by consumers with more paid than unpaid bills, while consumers' credit scores with more unpaid bills than paid bills looked facially similar to those without medical debt. Paid medical bills are already being excluded by the credit bureaus' previous voluntary actions. Other industry participants argued that the CFPB study was out of date, limited in scope, and not peer reviewed.

KARL E. SCHNEIDER, CONG. RSCH. SERV., IF12169, AN OVERVIEW OF MEDICAL DEBT: COLLECTION, CREDIT REPORTING, AND RELATED POLICY ISSUES 2 (2025), https://www.congress.gov/crs_external_products/IF/PDF/IF12169/IF12169.5.pdf [https://perma.cc/EL65-L5M2].

14. CFPB Director Rohit Chopra said, "People who get sick shouldn't have their financial future upended." CONSUMER FIN. PROT. BUREAU, *supra* note 12.

15. The CFPB has said, "We're looking into the promises and pitfalls of these innovations and the steps players in this market are taking to harness the benefits while managing the risks." Brian Kreiswirth, Peter Schoenrock & Pavneet Singh, *Using Alternative Data to Evaluate Creditworthiness*, CONSUMER FIN. PROT. BUREAU (Oct. 12, 2023, 4:44 PM), https://www.consumerfinance.gov/about-us/blog/using-alternative-data-evaluate-creditworthiness [https://perma.cc/X4HT-36UK].

history because of the intuitive relationship between proxies and underlying creditworthiness.[16]

In their attempt to capture these ideas, regulators have turned to a familiar, but deeply flawed, standard: nexus tests. Nexus tests abound in various legal domains, including tax, constitutional law, criminal law, and contracts. However, they often are analytical stand-ins for other concepts, and bundle together several different concepts.[17] In fact, nexus tests in one area of law do not usually imply the same test in a different area of law. In examining alternative data, regulators and industry experts have called for alternative data to share some nexus with creditworthiness.[18] However, these same actors have suggested a variety of different nexus tests including "sufficient nexus,"[19] "substantial nexus,"[20] "logical nexus,"[21] "close nexus,"[22] and "significant nexus."[23]

We argue that this confusion is not merely semantic pedantry—it can mislead courts and industry into thinking that regulators really are setting up different tests and standards. Moreover, nexus tests in the law are almost hopelessly vague and, as a result, seemingly inert or outright counterproductive. The move to nexus tests in credit lending potentially creates multiple confusing tests and undermines regulatory goals.

However, regulators have turned to nexus tests to preserve the ex ante relatedness between data and creditworthiness. In other words, fair lending law has historically required that any input data admit an intuitive explanatory story that exists prior to any statistical analysis. Yet machine learning has upended this paradigm by uncovering seemingly unrelated variables that happen to be statistically predictive of credit risk. In these cases, the rationalizations that explain the relationship between the input variable and credit risk are only plausible after the relationship has been discovered.[24]

For example, there may be little obvious ex ante relationship between what someone eats for breakfast and their creditworthiness, but a machine learning model may reveal correlations between this seemingly unrelated

16. *Federal Regulators Issue Joint Statement on the Use of Alternative Data in Credit Underwriting*, CONSUMER FIN. PROT. BUREAU (Mar. 5, 2021, 9:51 AM), https://www.consumerfinance.gov/about-us/newsroom/federal-regulators-issue-joint-statement-use-alternative-data-credit-underwriting [https://perma.cc/R4TJ-XMJB].

17. *See generally* Rapanos v. United States, 547 U.S. 715 (2006) (discussing the nexus between wetlands and navigable waters); Pac. Operators Offshore, LLP v. Valladolid, 565 U.S. 207 (2012) (discussing the nexus between an employee's injury and a company's operations).

18. CONSUMER FIN. PROT. BUREAU, BUILDING A BRIDGE TO CREDIT VISIBILITY 18 (2019) [hereinafter CREDIT VISIBILITY REPORT], https://files.consumerfinance.gov/f/documents/cfpb_building-a-bridge-to-credit-visibility_report.pdf [https://perma.cc/4LLG-E2XH].

19. *See infra* note 29 and accompanying text.

20. *See infra* Section I.B.

21. *See infra* note 32 and accompanying text.

22. *See infra* note 33 and accompanying text.

23. *See infra* note 34 and accompanying text.

24. *See* Selbst & Barocas, *supra* note 7, at 1122–26.

data. Yet once these seemingly unrelated correlations are discovered, it becomes possible to craft plausible explanatory stories, such as breakfast coffee drinkers may be more likely to pay their bills on time because they are more alert.

One perspective is that alternative data does not pose a problem as long as input variables admit a plausible explanatory story after they are discovered (ex post). Although the law has historically emphasized intuitive explanations developed ex ante, compelling ex post narratives can emerge even from opaque machine learning models. These narratives may nevertheless support forms of social oversight.

Still, the legal system must be careful not to overstate the epistemic value of intuition. It is relatively easy to generate stories that sound plausible but have no meaningful empirical grounding.[25] These narratives—whether intuitive from the outset or constructed after the fact—can obscure whether a variable is truly predictive or whether it simply fits a pattern we are socially primed to accept. Although intuitive stories can help structure fairness concerns or guide inquiry, they are an unreliable basis for fully determining whether a model is normatively acceptable.

The central claim of this Article is that machine learning regulation seeks to provide a mechanism for social oversight and ensure that inscrutable algorithms do not make important societal decisions. In order to adequately provide a mechanism of social oversight, regulations must take empirical reality seriously. Intuition has a role to play—it can help frame normative concerns and facilitate public reasoning—but it cannot substitute for careful testing of whether a variable actually performs the function it purports to serve. Predictive accuracy, in particular, must be treated as an empirical question rather than an intuitive one. If regulators fail to distinguish between what feels explanatory and what actually predicts, they risk enshrining familiar but flawed heuristics into law.

This Article uses the literature on the gaming of algorithmic systems to illustrate this point. Gaming of the system is frequently invoked in law and technology debates as a reason to limit transparency or resist reliance on complex models. Lenders similarly cite it as a potential problem they confront.[26] But a closer social scientifically informed examination of the underlying behavioral assumptions reveals that the real normative stakes of gaming are not just that someone "tricks" a model, but that such manipulation results in material

25. *See* DUNCAN J. WATTS, EVERYTHING IS OBVIOUS: ONCE YOU KNOW THE ANSWER 32–35 (2011).

26. For example, a podcast from *Marketplace Tech* discusses strategies consumers might use to "game their finances." *Gaming Your Finances to Get the Perfect Credit Score*, MARKETPLACE (July 11, 2022), https://www.marketplace.org/episode/2022/07/11/gaming-your-finances-to-get-the-perfect-credit-score [https://perma.cc/D579-5H5J]. "Gaming the system" is also generally presented as a potential threat to businesses that rely on algorithmic modeling. *See* Marijn Sax & Hao Wang, *From Threat to Opportunity: Gaming the Algorithmic System as a Service*, INTERNET POL'Y REV. 8 (May 6, 2025), https://policyreview.info/pdf/policyreview-2025-2-2007.pdf [https://perma.cc/Q9VM-JETC].

harm, such as borrowers receiving credit they cannot repay. Seen this way, gaming is not simply a theoretical concern—it is an empirical question: Does gaming occur, and if so, does it systematically harm borrowers? By centering the question this way, regulators are better positioned to design interventions that reflect actual risks. This also opens space for new tools like more rigorous audits of manipulation, not just of accuracy or fairness. Regulators should shift focus to the idea that social oversight is the primary value to preserve even when faced with black-box models, and more flexibly incorporate the scientific and philosophical foundations into that oversight rather than rely on nexus tests to simultaneously prioritize intuition and accuracy at the same time, even when they conflict in reality.

This Article proceeds as follows. Part I outlines "nexus tests" in law and why credit regulators have converged on nexus tests despite their serious flaws. Part II provides a brief overview of the history and origins of algorithmic credit lending and specifically surfaces the "causal turn" in credit lending. Part III surfaces the philosophical, social science, and computer science foundations of the role that intuition plays in credit lending and other algorithmic systems. Part IV addresses the conceptual tension of relying on intuition in machine learning contexts and suggests that regulators instead look to prioritize the empirical reality of algorithmic decisions. Part V offers a case study on algorithmic gaming in credit lending and how an empirical approach untangles the regulatory issues in that space.

I. THE REGULATORY "CONSENSUS" AROUND NEXUS TESTS

Alternative credit scoring systems are in their infancy, with limited existing regulatory interventions specifically designed to cover these systems. Despite the novelty of these systems, different regulators have begun using shockingly similar language to describe the proper relationship between alternative data and creditworthiness.[27] Although these conversations have not yet risen to formal rulemaking, fintech companies do seem to be looking to these musings as suggestions for how they might build their products.

More specifically, many different agencies and trade groups have focused on the "nexus" between alternative data and creditworthiness. These different groups argue that alternative data must bear a specific type of nexus relationship to creditworthiness. However, the type and scope of the nexus between alternative data and creditworthiness varies subtly across these groups.

Nexus tests are a poor fit for the alternative data context because they miss what is important about the relationship between data and creditworthiness. Regulators have rushed to them in part because they provide a way to scrutinize

27. *See infra* notes 28–34 and accompanying text.

the inputs to new fintech products. However, this focus misses the fact that AI enables many nonobvious insights.[28]

This Part examines regulators' discussions of nexus relationships between alternative data and credit scoring. In doing so, it offers several points of intervention. First, this Part catalogs the regulatory turn toward nexus analyses across different federal agencies. Next, this Part examines legal doctrines around distinct nexus tests. This discussion reveals that although regulators diverge only slightly on the appropriate nexus between alternative data and credit, federal courts are likely to view these subtle differences as material and constitutive of different tests. Finally, this examination teases out the causal roots of nexus tests as they have developed in the common law and demonstrates that regulators are chiefly concerned with causal relationships between underlying data and creditworthiness.

A. Regulatory Agencies and Nexus Talk

Regulatory agencies have begun to issue guidance about how firms should implement credit scoring based on alternative data while also remaining within the guidelines of existing fair lending regulations. These discussions often force regulatory agencies to offer clarifications about how alternative data should be linked to creditworthiness. While there are many possible conceptual relationships between alternative data and creditworthiness, regulators have coalesced around the idea that alternative data and creditworthiness must share some "nexus."

While these agencies have seemingly aligned on a nexus requirement, they depart company in their vision for what type of "nexus" alternative data must share with creditworthiness. Each agency seems to have adopted some version of a nexus test for the relationship between alternative data and creditworthiness, yet there is little agreement on the ideal nexus that links alternative data to creditworthiness. Oddly enough, there are seemingly as many nexus tests as there are regulatory agencies. Some agencies, such as the Office of the Comptroller of the Currency ("OCC"), advise developers of algorithmic credit scoring systems to use data with a "concrete nexus" to creditworthiness in determining an individual's credit score.[29] In contrast, the Vice Chair for Supervision of the Federal Reserve ("the Fed") claimed that

28. *See* Selbst & Barocas, *supra* note 7, at 1094–96 (discussing the ability of machine learning models to uncover relationships within data that humans may not obviously recognize).

29. OFF. OF THE COMPTROLLER OF THE CURRENCY, COMPTROLLER'S HANDBOOK: FAIR LENDING 83 (2023) ("[A] credit scoring system that uses a small number of attributes, simple decision rules, or data with a concrete nexus to creditworthiness may be lower risk than one that uses a larger number of attributes, modeling methods involving equations, algorithms, or complex decision rules, and data that are novel, proxied, or do not have a concrete nexus to creditworthiness. Highly complex scoring systems, such as those based on machine learning or data not widely used for credit decisions, generally have the highest inherent fair lending risk.").

alternative data should have a "sufficient nexus" to creditworthiness.[30] Another Fed official opined that alternative data should instead have an "obvious nexus" to creditworthiness.[31]

Alas, more nexus relationships persist. In the CFPB's report on reducing credit invisibility, a Federal Trade Commission ("FTC") official stated that alternative data should have a "logical nexus" to creditworthiness.[32] Moreover, the Bank Policy Institute—a nonpartisan research and advocacy group representing banks—claimed that credit scoring data should have a "close nexus" to creditworthiness.[33] And finally, a guidance document produced by the law firm Goodwin Procter suggests that there needs to be a "significant nexus" between alternative data and a credit decision.[34]

As this brief tour demonstrates, each agency seems to have adopted its own version of what type of nexus should link alternative data to creditworthiness. Yet the modifiers before "nexus" are not just synonyms. Going

---

30. More specifically, Barr claims that data could potentially lead to undesirable results if it is "correlated with a protected class and lack[s] a *sufficient nexus* to creditworthiness." Michael S. Barr, Vice Chair for Supervision, Bd. of Governors of the Fed. Rsrv. Sys., Speech at the "Fair Housing at 55—Advancing a Blueprint for Equity" National Fair Housing Alliance 2023 National Conference: Furthering the Vision of the Fair Housing Act 4 (July 18, 2023) (emphasis added). Interestingly, it seems that these remarks indicate that the nexus to creditworthiness is a site of concern even in situations where the machine learning algorithm is not using data that is correlated to protected characteristics. *Id.*

31. Carol A. Evans, *Keeping FinTech Fair: Thinking About Fair Lending and UDAP Risks*, CONSUMER COMPLIANCE OUTLOOK (2017), https://www.consumercomplianceoutlook.org/2017/second-issue/keeping-fintech-fair-thinking-about-fair-lending-and-udap-risks [https://perma.cc/7NLD-SSTJ].

32. CREDIT VISIBILITY REPORT, *supra* note 18, at 18. The CFPB's report notes that some consumers may be denied opportunities based on activities that lack a logical nexus to creditworthiness.

33. For example, banks review variables to ensure that models do not consider prohibited bases or close proxies for prohibited bases to mitigate disparate treatment risk. Banks also consider whether each variable has a close nexus to creditworthiness and, if not, whether the variable might result in additional fair lending risk. Banks also conduct statistical testing of model outcomes to assess whether facially neutral models pose disparate impact risk and whether model changes would produce less impact on a protected class without undermining model performance.

E-mail from Stephanie Wake, Vice President, Bank Pol'y Inst., to Chief Couns.'s Off., Off. of the Comptroller of the Currency et al. 24 (June 25, 2021), https://www.fdic.gov/system/files/2024-06/2021-rfi-financial-institutions-ai-3064-za24-c-015.pdf [https://perma.cc/7AVL-GENZ] (responding to a Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence).

34. Using alternative credit data with a significant nexus to creditworthiness is a good practice that can create highly consistent credit scoring models. However, before using alternative credit data, fintechs should consider the nexus between the data collected and creditworthiness. Not all forms of alternative credit data are as beneficial in determining a borrower's creditworthiness as others.

Danielle Reyes & Nico Ramos, *Traps for the Unwary: Using Alternative Credit Data to Expand Credit Access to LMI Individuals and Underrepresented Communities*, GOODWIN FINTECH FLASH (Mar. 26, 2024), https://www.goodwinlaw.com/en/insights/publications/2024/03/insights-finance-ftec-traps-for-the-unwary-using-alternative [https://perma.cc/9UAY-3CYG].

through just the examples presented here, regulators and commenters invoked different concepts for these different nexus tests. Is nexus about simplicity ("concrete"), intuition ("obvious"), disparate impact ("sufficient" and "close"), inferences based on group membership ("logical"), consistency ("significant"), or inaccuracy ("speculative")? Or is it all the above?

Despite invoking different concepts with their nexus tests, regulators and industry groups appear to have common ideas about alternative data. The CFPB issued a call for information about uses of alternative data and distinguished between data directly related to a consumer's finances (such as rent or cable TV payments) and data less related to a consumer's finances (such as education, occupation, or social media use).[35] However, this distinction is not as much of a bright-line rule as the CFPB would hope. For instance, data about a person's occupation is often highly correlated with a person's finances.[36]

Similarly, the CFPB has conducted studies which suggest that cash flow data can help lenders extend credit.[37] In a joint statement by several regulatory bodies (including the Fed, CFPB, the Federal Deposit Insurance Corporation, and others), regulators claimed that "using alternative data, such as cash flow data, that are *directly related* to consumer's finances and how consumers manage their financial commitments may present lower risks."[38] Likewise, Goodwin Procter claims that using cash flow is acceptable, yet education status is unacceptable despite the fact that lenders often use this information in other lending contexts.[39]

Taken together, these statements form a tangled morass. This putative guidance from regulators does little to instruct firms about how alternative data should be related to creditworthiness. The next Section examines federal common law doctrine in an attempt to make sense of nexus tests across several legal fields.

---

35. *See* Kreiswirth et al., *supra* note 15.

36. The Bureau of Labor Statistics collects information about different occupations and average salary ranges. *See Overview of BLS Wage Data by Area and Occupation*, U.S. BUREAU LAB. STAT. (Aug. 1, 2024), https://www.bls.gov/bls/blswage.htm [https://perma.cc/5GBA-CPQA].

37. *See* Alexei Alexandrov, Alyssa Brown & Samyak Jain, *Looking at Credit Scores Only Tells Part of the Story – Cashflow Data May Tell Another Part*, CONSUMER FIN. PROT. BUREAU (Aug 22, 2024, 11:30 AM), https://www.consumerfinance.gov/about-us/blog/credit-scores-only-tells-part-of-the-story-cashflow-data [https://perma.cc/36YL-CPPL].

38. BD. OF GOVERNORS OF THE FED. RSRV. SYS., CONSUMER FIN. PROT. BUREAU, FED. DEPOSIT INS. CORP., NAT'L CREDIT UNION ADMIN. & OFF. OF THE COMPTROLLER OF THE CURRENCY, INTERAGENCY STATEMENT ON THE USE OF ALTERNATIVE DATA IN CREDIT UNDERWRITING (2019), https://files.consumerfinance.gov/f/documents/cfpb_interagency-statement_alternative-data.pdf [https://perma.cc/Z9B3-NTDR] (emphasis added).

39. *See* Reyes & Ramos, *supra* note 34; Jennifer Brozic, *Can I Buy a Car with No Credit?*, INTUIT CREDIT KARMA (Nov. 11, 2024), https://www.creditkarma.com/auto/i/buy-car-no-credit [https://perma.cc/Y4BG-72P9] ("If you're a student or recent graduate, you may be able to qualify for special financing. Some lenders and automakers have programs specifically designed for students and graduates with little or no credit history.").

B. What the Common Law Can Teach About Nexus Tests

Nexus tests are pervasive in American law. Many distinct areas of the legal system seek to understand the "nexus" between two things. A constitutionally valid search warrant must "establish a 'minimally sufficient nexus'" between the place to be searched and the items to be seized.[40] To prove that a gun was "possessed 'in furtherance' of a drug trafficking crime, the government must show 'some nexus between the firearm and drug selling operation.'"[41] Due process requires that there "be a sufficient nexus between the defendant and the United States."[42] To satisfy the Commerce Clause, a state must show that any tax is applied only "to an activity with a substantial nexus with the taxing state."[43]

The Supreme Court has recently invoked nexus tests in a variety of areas. Returning to tax, the Supreme Court recently said that an entity's substantial nexus to a state does not require a physical presence.[44] In environmental law, the Court has said that the Environmental Protection Agency ("EPA") cannot regulate wetlands when they lack a "significant nexus" to a navigable body of water.[45] And in property law, the Supreme Court held that "[real estate] permit[ting] conditions must have an 'essential nexus' to the government's land-use interest."[46]

Federal courts have long been puzzled by nexus tests. This is largely because nexus tests are seemingly difficult to administer and are notoriously unworkable. Though nexus tests are conceptually puzzling, they operate across nearly the whole spectrum of areas of law. Credit lending law is no exception. However, it is also not immune to the conceptual confusion surrounding the appropriate application of nexus tests.

1. Distinction Between Nexus Tests

Analyzing the reasoning of federal judicial opinions reveals that courts implicitly reject general theorizing about nexus tests. This rejection operates on two distinct levels.

40. United States v. Neal, 106 F.4th 568, 572 (6th Cir. 2024) (quoting United States v, Carpenter, 360 F.3d 591, 596 (6th Cir. 2004)).

41. United States v. Molina, 443 F.3d 824, 829 (11th Cir. 2006) (quoting United States v. Timmons, 283 F.3d 1246, 1253 (11th Cir. 2002)).

42. United States v. Al Kassar, 660 F.3d 108, 118 (2d Cir. 2011) ("In order to apply extraterritorially a federal criminal statute to a defendant consistently with due process, there must be a sufficient nexus between the defendant and the United States, so that such application would not be arbitrary or fundamentally unfair." (quoting United States v. Yousef, 327 F.3d 56, 111 (2d Cir. 2003))).

43. St. Tammany Par. Tax Collector v. Barnesandnoble.com, 481 F. Supp. 2d 575, 577 (E.D. La. 2007).

44. South Dakota v. Wayfair, Inc., 585 U.S. 162, 178 (2018).

45. Sackett v. EPA, 598 U.S. 651, 680 (2023).

46. Sheetz v. County of El Dorado, 601 U.S. 267, 275 (2024) (quoting Nollan v. Cal. Coastal Comm'n, 483 U.S. 825, 837 (1987)).

First, the content of a specific nexus test may depend on the legal context in which it is invoked. For example, in *Pacific Operators Offshore, LLP v. Valladolid*, the Supreme Court considered whether a worker's fatal injury that occurred at an onshore facility qualified for workers' compensation under a statute that extended coverage to injuries that occurred in federal waters.[47] The Court concluded that there needed to be a "substantial nexus" between the injury and operations on the water.[48]

Although Justice Scalia mostly agreed with the majority, in his usual candor, he maligned "substantial nexus" as "an indeterminate phrase that lacks all pedigree."[49] He articulated the idea that the same nexus tests used in different legal contexts are, in fact, different tests.[50] The key problem was that the Court's previous "substantial nexus" jurisprudence focused on taxation, which did not imply any requirement for showing proximate cause.[51] Conversely, the workers' compensation scheme implied something more like "substantial *causal* nexus" because it did embed a proximate cause requirement.[52]

Second, the Supreme Court has implicitly recognized that subtly different nexus tests articulate different standards. In dissecting the proposed meaning of "substantial nexus," Justice Scalia looked only to other instances where the Supreme Court invoked "substantial" nexus as a term of art.[53] In other words, "sufficient nexus" or similar nexus tests do not shed much, if any, light on what a "substantial nexus" requires.

---

47. Pac. Operators Offshore, LLP v. Valladolid, 565 U.S. 207, 210–13 (2012). More specifically, the Supreme Court sought to determine whether a fatal injury that occurred at an onshore oil and gas processing facility was covered under Longshore and Harbor Workers' Compensation Act ("LHWCA"), pursuant to the Outer Continental Shelf Lands Act ("OCSLA"). *Id.* at 212–13. The OCSLA extended LHWCA coverage to injuries "occurring as the result of operations conducted on the [OCS]" for the purpose of extracting natural resources from the shelf. *Id.* In determining whether the fatal injury was covered as "the result of operations conducted" on the outer continental shelf, even though the injury occurred at an onshore processing facility, the majority opinion concluded that there must be a "substantial nexus" between the injury and the extractive operations on the shelf. *Id.* at 222.

48. *Id.*

49. *Id.* at 225 (Scalia, J., concurring in part and concurring in judgment).

50. *Id.*

51. *Id.* at 223–25.

52. Specifically, he says:

> That clarification—and any further clarification in the Commerce Clause context—will not be remotely helpful to lower courts attempting to apply the substantial-nexus test in the very different legal context of workers' compensation under §1333(b). In this latter context, I assume the Court means by "substantial nexus" a substantial *causal* nexus—since §1333(b)'s "as the result of" language "plainly suggests causation," . . . . Like the word "nexus" itself, the definition of "substantial nexus" in our state-tax cases does not require any *causal* relationship whatsoever. The proximate-cause test, by comparison, represents a much more natural interpretation of a statute that turns on causation.

Id. at 225 (citation omitted).

53. *Id.*

For instance Justice Kennedy, in his *Rapanos v. United States* concurrence, argued that "navigable waters" for the Clean Water Act ("CWA") extends to any waters with a "significant nexus" to navigable waters.[54] Moreover, the Army Corps of Engineers adopted the "significant nexus" test in their rulemaking about the scope of "waters of the United States."[55] Importantly, however, Scalia's concurrence in *Pacific Operators Offshore* did not look to the Court's writing on "significant nexus" to help determine the scope and content of "substantial nexus,"[56] thus lending significant credibility to the idea that "substantial nexus" and "significant nexus" refer to different tests.

Federal courts often interpret nexus tests through a lens that emphasizes subtle distinctions in language, leading judges to treat regulators' slightly different phrasing as signaling entirely separate analyses. While terms like "concrete," "logical," "sufficient," and "significant" nexus may aim to express the same principle, judges are likely to see them as conceptually distinct. As a result, it makes sense to express the principle that regulators are attempting to articulate more clearly: Data should demonstrate an intuitive link to credit-worthiness determinations, not merely the abstract connectedness suggested by traditional nexus tests.

2. Nexus Tests as Analytical Stand-Ins

Nexus tests are often used as analytical stand-ins for some other consideration. In credit lending, considerations around accuracy, fairness, transparency, and causation are all embedded within nexus talk.[57] Credit lending is not unique in this respect though, as other legal areas have similarly bundled multiple concepts into nexus.

Sometimes, nexus tests are stand-ins for determining the limits of regulatory power. For example, significant nexus has been invoked in environmental law to determine how connected a particular waterway is to a body of water that the EPA is permitted to regulate.[58] In tax law, a substantial nexus determines upon which activities a state may exercise its taxing authority.[59]

Other times, nexus tests are a stand-in for the extraterritorial reach of American courts. The sufficient nexus test for constitutional due process "ensures that a United States court will assert jurisdiction only over a defendant

54. Rapanos v. United States, 547 U.S. 715, 759 (2006) (Kennedy, J., concurring).

55. Despite attempting to apply the nexus test, the Court later overturned the Army Corps of Engineers definition and said that in this case a "significant nexus" requires a "continuous surface connection" to lakes and rivers involved in interstate commerce. *See* Sackett v. EPA, 598 U.S. 651, 678–79 (2023). It is difficult to see how this principle might be broadly applicable to other areas that invoke substantial nexus.

56. *See Pac. Operators Offshore*, 565 U.S. at 223–27 (Scalia, J., concurring) (discussing the scope and meaning of "substantial nexus").

57. *See, e.g.*, Evans, *supra* note 31.

58. *See Sackett*, 598 U.S. at 667–69.

59. Quill Corp. v. North Dakota *ex rel.* Heitkamp, 504 U.S. 298, 311 (1992), *overruled on other grounds by* South Dakota v. Wayfair, Inc., 585 U.S. 162 (2018).

who should reasonably anticipate being hauled into court in [the United States]."[60] Because of the vagueness of the "reasonable anticipations" of a potential defendant, courts turn to a nexus test in an attempt to make the issue more tractable and potentially expand the justice system's reach.[61] Rather than examining whether a reasonable person would anticipate being subject to potential criminal liability based on their conduct abroad, courts have instead considered whether there is a substantial nexus between the criminal activity and "American interests."[62]

In yet another legal context, nexus is a stand-in for determining sentencing enhancements. In criminal law, the nexus between a firearm and drug trafficking activity is used to determine whether the gun was used "in furtherance" of drug trafficking. Arguably because of the difficulty of proving (or even determining) how and whether possession of a gun advances a drug trafficking scheme, courts have turned (yet again) to a nexus test.[63] However, there are many possible cases where there is a nexus between a gun and a drug trafficking scheme, but the gun possession does not further the scheme. For example, the Eighth Circuit claims that "[a] nexus can exist when a firearm is in proximity to items identified as relating to drug trafficking."[64] While proximity alone does not guarantee that gun possession is "in furtherance" of drug trafficking, it could provide sufficient basis for the jury to draw that conclusion. In other words, the nexus between a firearm and instrumentalities of trafficking is used as a proxy for determining when a firearm furthers a trafficking scheme. The nexus between gun possession and trafficking may be underinclusive. For instance, a person could keep weapons in a remote location, away from all the instrumentalities of trafficking, but the possibility of accessing it could embolden him to begin his operation. In this case, gun possession is clearly "in furtherance" of drug trafficking but would likely fail the nexus test.

Adding to the confusion, nexus is used as a stand-in for insurance claims sharing facts and circumstances. Insurance contracts sometimes include a

---

60. United States v. Skinner, 536 F. Supp. 3d 23, 42 (E.D. Va. 2021) (quoting United States v. Mohammad-Omar, 323 F. App'x 259, 261 (4th Cir. 2009)).

61. *See id.*

62. Indeed, the substantial nexus test in the due process analysis has led to an expansive reach of American courts' jurisdiction over foreign nationals. In *Brehm,* which has become the controlling nexus test for due process, the Fourth Circuit found that a South African citizen could be subject to criminal liability in the United States after stabbing a British citizen on a NATO-controlled airfield in Afghanistan. United States v. Brehm, 691 F.3d 547, 549 (4th Cir. 2012). The Fourt Circuit found that although Brehm did not intend to target or affect American soil or American commerce, Brehm's criminal actions affected significant American interests, including maintaining order on the base, the maintenance of military discipline, and the reallocation of Department of Defense resources to capture the defendant and investigate the crime. *Id.* at 552. Importantly, however, the *Brehm* decision highlighted how haling Brehm into an American court was not substantially unfair because he should have anticipated being prosecuted "somewhere" for his conduct. *Id.*

63. United States v. Druger, 920 F.3d 567, 570 (8th Cir. 2019).

64. *Id.*

"common nexus" clause that allows the insurer to deny claims when an insured party makes multiple claims that share "related" facts.[65] Again, this language clarifies little, as courts differ on whether this language is clear or unambiguous, and whether "common nexus" as covering the same exact facts,[66] or those that are "causally connected" to a wrongful act in question.

With regard to alternative data, regulators seem to be bundling several different concepts together with their use of nexus tests. At a minimum, regulators have referred to the accuracy of algorithmic credit lending models,[67] the minimization of disparate impact,[68] and the explainability of algorithmic decisions[69] in their invocation of nexus tests. Yet attempting to optimize each of these concepts will likely lead to conflicts—a more accurate algorithm may create more disparate impact, for example.[70] With alternative data specifically, the root of the problem is a concern about causality.

3. The Causal Roots of Nexus Tests

Regulators' convergence on nexus tests is understandable in the sense that nexus tests have roots in causality. Causality underlies statutory requirements for explanations in adverse credit decisions, which makes nexus analyses seemingly even more attractive.[71] As a result, nexus tests become tempting vehicles for legal analysis of causality. Echoing this idea, Black's Law Dictionary defines nexus as "[a] point of causal intersection, link, relation, [and]

---

65. Eric Jesse & Catherine Serafin, *When One Insurance Claim Is Better than Many*, CFO (Apr. 3, 2017), https://www.cfo.com/news/when-one-insurance-claim-is-better-than-many/660518/ [https://perma.cc/F7L9-8DAC].

66. For example, a Pennsylvania court ruled that "'common nexus' was ambiguous." Bryce Friedman, *Pennsylvania Court Rules that Two Lawsuits Do Not Allege Interrelated Wrongful Acts*, SIMPSON THACHER (Apr. 29, 2016), https://www.stblaw.com/about-us/publications/view/2016/04/29/pennsylvania-court-rules-that-two-lawsuits-do-not-allege-interrelated-wrongful-acts [https://perma.cc/Y6T8-5PPG]. However, the Second Circuit said that this kind of language was plain and unambiguous. *Second Circuit Rejects "Factual Nexus" Test Where Plain and Unambiguous Language of Insurance Policy Can Apply to Determine Relatedness of Claims*, BRESSLER, AMERY & ROSS, P.C. (Oct. 27, 2015), https://www.bressler.com/publication-second-circuit-rejects-factual-nexus-test-where-plain-and-unambiguous-language-of-insurance-policy-can-apply-to-determine-relatedness-of-claims [https://perma.cc/8EPE-5GWM].

67. *See supra* notes 29–34 and accompanying text.

68. *See supra* notes 30–33.

69. *See supra* notes 12–16 and accompanying text.

70. This "accuracy-fairness" trade-off has been explored in the fairness in machine learning literature. *See, e.g.*, Aditya Krishna Menon & Robert C. Williamson, *The Cost of Fairness in Binary Classification*, 81 PROC. MACH. LEARNING RSCH. 9–10 (2018), https://proceedings.mlr.press/v81/menon18a.html [https://perma.cc/U4AQ-E7AC]; Aniket Kesari, *The Privacy-Fairness-Accuracy Frontier: A Computational Law & Economics Toolkit for Making Algorithmic Tradeoffs*, ASS'N FOR COMPUTING MACH.: SYMP. ON COMPUT. SCI. & L. 77 (Nov. 2022), https://dl.acm.org/doi/pdf/10.1145/3511265.3550437 [https://perma.cc/43GY-6E7M].

71. *See, e.g.*, Home Mortgage Disclosure Act of 1975, 12 U.S.C. §§ 2801–2811 (2018).

connection."[72] Similarly, several federal judges have found that claims are related if they "arise from a nexus of logically or causally related facts."[73]

Legal scholars have similarly investigated the causal roots of nexus. In refugee and asylum law, Michelle Foster analyzes the causal nexus requirements implied by the "for reasons of" language in the Refugee Convention.[74] In environmental law, Daniel Farber analyzes the "geographic nexus" between a government interest and an environmental problem that has a causal element in analyzing how much environmental damage in one location needs to affect an individual in another location.[75] Within insurance law, Peter Swisher examines causal requirements for substantial or sufficient nexus in auto insurance claims.[76] In property law, Jan G. Laitos and Teresa Helms Abel examine the Takings Clause's requirement that in order to prove sufficient nexus, the plaintiff shows that the government was the cause of property harm.[77]

Alongside the widespread scholarly treatment of nexus tests as causal tests,[78] judges often consider nexus tests to be principally causal. On this front, Justice Scalia offers a ringing judicial endorsement of nexus tests as causal relationships. Returning to *Pacific Operators Offshore*, Justice Scalia took serious issue with the impracticability of the substantial nexus test and claimed it was a "novel [invention of] legalese."[79] However, Scalia claimed that, rather than substantial nexus, the underlying text of the statute governing workers' compensation supported a proximate cause test. Specifically, he claimed:

> Unlike the substantial-nexus test, proximate cause provides a "vocabulary" for answering questions like the one raised by the facts of this case. It may be productive, for example, to consider whether the injury was "within the scope of the risk" created by [Outer Continental Shelf] operations, or whether some "superseding or intervening cause" exists.[80]

---

72. *Nexus*, LAW DICTIONARY, https://thelawdictionary.org/nexus [https://perma.cc/Z7XC-AGK9].

73. *See, e.g.*, Papalia v. Arch Ins. Co., No. 2:15-cv-02856, 2017 WL 3288113, at *11 (D.N.J. Aug. 1, 2017) (quoting Columbus Life Ins. Co. v. Arch Ins. Co., No. 3:14-CV-01659, 2016 WL 2865952, at *8 (N.D. Ind. May 17, 2016)).

74. Michelle Foster, *Causation in Context: Interpreting the Nexus Clause in Refugee Convention*, 23 MICH. J. INT'L L. 265, 266 (2002).

75. Daniel A. Farber, *Stretching the Margins: The Geographic Nexus in Environmental Law*, 48 STAN. L. REV. 1247, 1248 (1996).

76. Peter Nash Swisher, *Causation Requirements in Tort and Insurance Law Practice: Demystifying Some Legal Causation "Riddles,"* 43 TORT TRIAL & INS. PRAC. L.J. 1, 31–33 (2007).

77. Jan. G. Laitos & Teresa Helms Abel, *The Role of Causation When Determining the Proper Defendant in a Takings Lawsuit*, 20 WM. & MARY BILL RTS. J. 1181, 1194–99 (2012).

78. Scholarly consensus has developed around nexus tests describing causal relationships, but some take an issue with this.

79. Pac. Operators Offshore, LLP v. Valladolid, 565 U.S. 207, 223 (2012) (Scalia, J., concurring).

80. *Id.* at 224–25 (citation omitted).

As we can see from Scalia's statement, substantial nexus tests mark similar territory to proximate cause. In fact, there are strong arguments that proximate cause could act as a reasonable—and more administrable—substitute for a substantial nexus test.

This brief tour of nexus tests shows that while it makes sense for regulators to gravitate to nexus tests because of their roots in causation, these tests offer little clarity for handling new situations. Sometimes language like "substantial nexus" might imply "substantial *causal* nexus," and sometimes it has no element of causation at all.[81] Nexus tests also carry the burden of encompassing other concepts such as physical proximity, activity levels, and factual similarity. Thus, if credit lending regulators really are concerned about causal stories in credit lending, it is easy to see why nexus is attractive, but imperfect.

* * *

Although nexus tests abound in various areas of law, they seem to create more administrability problems than they solve. There are at least half a dozen different invocations of nexus tests that are seemingly similar. However, each test implies different requirements depending on the legal context. In fact, even the *same* nexus test has different standards in two different areas of law (e.g., "substantial nexus" in tax versus workers' compensation).[82] Causation is a fundamental concept in credit lending law, and regulators likely gravitated toward nexus because of its causal roots.

II. ALGORITHMIC CREDIT SCORING

A. *THE ADVENT OF THE CREDIT SCORE*

In this Section, we focus on developments related to the definition and rise of "alternative" data and distinguish alternative data from "traditional" data.

1. Tracing the "Causal Turn" in Credit Lending

One key theme emerging from the history of U.S. credit lending is how modern credit lending emerged as a response to the widespread use of alternative-like data. In modern discourse, alternative data refers to information that falls outside factors normally considered in a credit score.[83] Some examples include cash in checking accounts and strength of social ties on social media.[84] Although these data are framed as "alternative" now because they do not enter

81. *Id.* at 223–25.

82. *See, e.g.*, *id.*

83. *See* Sullivan, *supra* note 2.

84. *See* Nizan Geslevich Packin & Yafit Lev-Aretz, *On Social Credit and the Right to Be Unnetworked*, 2016 COLUM. BUS. L. REV. 339, 361; *Federal Regulators Issue Joint Statement on the Use of Alternative Data in Credit Underwriting*, CONSUMER FIN. PROT. BUREAU (Mar. 5, 2021, 9:51 AM), https://www.consumerfinance.gov/about-us/newsroom/federal-regulators-issue-joint-statement-use-alternative-data-credit-underwriting [https://perma.cc/5H7B-2VZZ] ("Alternative data include cash flow data derived from consumers' bank account records.").

traditional credit score calculations, they actually have roots in how credit lending worked for much of American history.

For the early part of U.S. history, consumer credit was usually extended on the basis of personal relationships.[85] Moreover, credit was usually extended directly from an individual seller to an individual buyer.[86] Kenneth Lipartito traces much of this history, discussing how consumers often had informal credit mechanisms such as "book credit" at local bookshops, or would carry a tab at local bars, restaurants, or grocery stores.[87]

Even in this early stage of credit lending, lenders would use factors like the length of a relationship and on-time payments as proxies for how much credit to extend.[88] The main feature of this time period was that credit was largely based on social ties and relied on the fact that neither businesses nor consumers tended to move frequently.[89]

As consumers started making larger purchases such as appliances and automobiles, installment plans were sometimes offered directly by the seller.[90] But this period also saw the emergence of intermediaries.[91] Finance companies resembling modern credit bureaus emerged to offer installment financing, and they employed agents who would assemble credit files of individual consumers and acted as both information sources and debt enforcers for sellers of these big-ticket items.[92] Some of the information contained in these

85. *See* Lipartito, *supra* note 9, at 1.

86. *See id.* at 21.

87. *Id.*

88. *See id.*

89. DONNCHA MARRON, CONSUMER CREDIT IN THE UNITED STATES: A SOCIOLOGICAL PERSPECTIVE FROM THE 19TH CENTURY TO THE PRESENT 100 (2009). Lenders at this time also made an interesting distinction between "business" and "consumer" debt. Business debt was generally seen as more morally acceptable, whereas consumer debt was discouraged. *See* Rowena Olegario, *The History of Credit in America*, OXFORD RSCH. ENCYCLOPEDIAS (May 23, 2019), https://oxfordre.com/americanhistory/display/10.1093/acrefore/9780199329175.001.0001/acrefore-9780199329175-e-625 (on file with the *Iowa Law Review*). Narratives about these various kinds of debt reflected these differences in attitudes. *See* LENDOL CALDER, FINANCING THE AMERICAN DREAM 211–12 (1999). People taking out consumer debt faced significant social stigma, and narratives at the time prioritized frugality and living within one's means. Echoes of these narratives can be seen in modern discourse around creditworthiness. In the late nineteenth and early twentieth centuries, the American economy grew and became more nationalized, leading to changes in credit lending and revealing the origins of the tension between quantitative and qualitative credit rating. American consumers were buying more big-ticket items such as sewing machines and automobiles on installment plans. *See id.* at 203.

90. Lipartito, *supra* note 9, at 21–22.

91. *See* Josh Lauer, *Credit Reporting and the History of Commercial Surveillance in America*, OXFORD RSCH. ENCYCLOPEDIAS (Nov. 22, 2022), https://oxfordre.com/americanhistory/display/10.1093/acrefore/9780199329175.001.0001/acrefore-9780199329175-e-988 (on file with the *Iowa Law Review*).

92. At this time, these lenders were still local—thousands proliferated across the United States and their efficacy largely depended on their ability to assemble detailed dossiers of borrowers and maintain good relationships with sellers. *See id.*

credit files was highly quantified, including the number of existing accounts, number of on-time payments, and dates of purchases. Other aspects of these files were more qualitative or narrative—investigators would collect and verify statements about creditors' business and personal relationships. This brief history shows that quantified data was not always the default data and indeed might have been considered alternative data during the first half of the twentieth century.[93]

The shift from narrative files to quantified data reports occurred between the 1970s and early 1990s with the advent and growth of credit cards. Credit cards marked a shift from consumers primarily borrowing from the seller to borrowing from a financial institution. Here, we see the origins of the "causal turn" in credit lending. While features such as race, income, and gender could be *predictive* of creditworthiness, lawmakers did not feel these proxies were fair. Actual consumer behavior was seen as having a stronger *causal* basis in making credit decisions.[94]

This emphasis on causality ultimately led to the emergence of the FICO credit score. FICO scores promise to boil down all of the complexities in the narrative dossiers that preceded them into an easy to interpret three-digit number.[95] Any lender could infer credit risk from the credit score and make decisions accordingly.[96] Importantly, the FICO score uses direct measures of

93. Lipartito argues,

> As the monitoring of consumer credit expanded to a national scale, credit scoring and reductive quantitative scoring actually decreased for a time in favor of richer contextual data. . . . Looking to the mixed experience of ratings in trade, consumer credit managers argued that rating books were old fashioned and went out of date too quickly. Instead, the bureaus emphasized the thickness and thoroughness of their files.

Lipartito *supra* note 9, at 26. In this era, quantified information was seen as potentially unreliable, or at least non-predictive, and qualitative information as the better basis for a credit decision. The history that comes after this point reveals the underlying logic of alternative data—is there information that can be used to extend credit to those who lack the traditional markers of creditworthiness? Importantly though, this question did not turn on quantification, but rather finding new proxies in the absence of traditional ones. Specifically, in the post-World War Two era, as suburbanization drastically increased and lenders had less information about new arrivals to a locale, credit lenders started looking at factors such as income and race to proxy for creditworthiness. *See generally* Derek S. Hoff, *The Original Housing Crisis: Suburbanization, Segregation, and the State in Postwar America*, 36 REVS. AM. HIST. 259 (2008) (reviewing DAVID M.P. FREUND, COLORED PROPERTY (2007)); *see* RICHARD ROTHSTEIN, THE COLOR OF LAW 59–76 (2017) (arguing that government policies enabled discriminatory lending by private lenders for mortgages and contributed to historic redlining); *see* Becky Nicolaides & Andrew Wiese, *Suburbanization in the United States After 1945*, OXFORD RSCH. ENCYCLOPEDIAS (Apr. 26, 2017), https://oxfordre.com/americanhistory/display/10.1093/acrefore/9780199329175.001.0001/acrefore-9780199329175-e-64 (on file with the *Iowa Law Review*) (discussing the suburbanization of the United States).

94. Lipartito, *supra* note 9, at 34.

95. *What's in My FICO Scores?*, MYFICO (2025), https://www.myfico.com/credit-education/whats-in-your-credit-score [https://perma.cc/2EKN-X2SU].

96. *See* Lauer, *supra* note 91.

consumer behavior such as payment history, total debt, length of credit history, types of existing credit, and recent credit applications.[97] While each of these may correlate with characteristics such as age, race, and gender, they ultimately are related to an individual's choices, thus satisfying a definition of fairness that prioritizes a causal, or intuitive, relationship between an individual's actual behavior and their creditworthiness.

Another interesting lesson from the history of credit is how quantification may have led to current conceptual problems around gameability and nexus tests. While biased in many ways, the types of narratives historically used were difficult to game. While an individual could try to hide certain pieces of unfavorable information, it was difficult to game negative information coming out of independent interviews with members of the local community.

2. How Do Credit Scores Work?

The move from narrative to quantified data ultimately led to the birth of the modern credit score in the United States. The features used to determine creditworthiness are standard, and the credit score essentially averages all of this information into a single three-digit number, ranging from 300 to 850.[98] The most widely recognized credit score in the United States is the FICO score, which was first developed in the late 1980s.[99] The core components of a FICO score and their respective contributions are as follows: payment history (thirty-five percent), amounts owed (thirty percent), length of credit history (fifteen percent), credit mix (ten percent), and new credit inquiries (ten percent).[100] These five components reflect FICO's fundamental philosophy that individual creditworthiness is best assessed by examining behaviors that directly relate to financial responsibility.[101] It is also important to note that these components and their respective weightings have evolved over time, with adjustments made to reflect changes in economic trends and shifts in consumer behavior.[102]

---

97. *What's in My FICO Scores?*, *supra* note 95.

98. *See Credit Scores*, FED. TRADE COMM'N (Sept. 2024), https://consumer.ftc.gov/articles/credit-scores [https://perma.cc/2XRS-CU6Y].

99. Lipartito, *supra* note 9, at 31.

100. *What's in My FICO Scores?*, *supra* note 95.

101. *See* Therese Henry, *Making Decisions Based on Customer Behavior*, FICO CMTY. BLOG (Apr. 1, 2019), https://community.fico.com/s/blog-post/a5Q2E000001cDKUAY/fico1669 [https://perma.cc/7JM3-HG9Y].

102. For instance, when the FICO score was first introduced in 1989, payment history accounted for approximately forty percent of the total score, but by 2024, this weight had been adjusted to thirty-five percent to reflect a more balanced view of credit risk. Similarly, the weight for amounts owed increased from twenty-five percent in 1989 to thirty percent in 2024, acknowledging the importance of credit utilization in predicting financial stability. Additionally, the significance of new credit inquiries has fluctuated, with its weight decreasing from fifteen percent in earlier versions to ten percent in recent models, as the industry moved towards minimizing the penalization of individuals seeking new credit.

Traditionally, certain types of easily surveilled information enter into credit score calculations, whereas others that are harder to report remain invisible. For example, traditional mortgage payments, auto loans, and student loans are consistently reported to credit bureaus, directly contributing to an individual's credit profile and boosting their score if payments are made on time.[103]

Moreover, the types of credit extended substantially affect a credit score calculation. Revolving credit such as credit card debt tends to affect scores negatively, whereas stationary or installment credit such as student loans tends to affect scores positively.[104] For example, a medical school graduate with hundreds of thousands of dollars in student loans, a mortgage, and an auto loan regularly reports a lot of information to credit rating agencies. Much of this debt is seen as "good" debt, and regular payments will dramatically increase a credit score.[105]

Conversely, a high school graduate who rents an apartment and purchases a used vehicle looks relatively credit invisible—while they are making recurring payments on similar goods, none of these show up in a credit report. Of course, there are good reasons for why certain types of payments have historically been reported to credit rating agencies and some were not. At a basic level, from a creditor's perspective, while both of these hypothetical individuals are regularly making payments on similar goods, the first individual is paying off debt while the second is making recurring consumption purchases. The medical student is therefore a better credit risk under the traditional model.

Therefore, traditional credit reporting does not necessarily capture all of the potentially useful information in determining an individual's creditworthiness. While it makes some sense that only debt-paying activities should count toward a credit score, this essentially is a strong assumption on credit lenders' part. Other types of recurring payments could provide similar information about potential creditworthiness as well.[106]

One major lesson is about creativity around what is a useful predictor of creditworthiness. Although intuitively, linking debt payments to creditworthiness makes sense, this arguably leaves a gap in the market for financial services. Indeed, the emergence of fintech companies that use alternative data for credit lending arguably proves this gap. Realizing that the traditional focus on the credit–debt link overlooked individuals who were financially responsible

103. *What Does Credit Mix Mean?*, MYFICO https://www.myfico.com/credit-education/credit-scores/credit-mix [https://perma.cc/D95D-2Q75].

104. *Installment vs. Revolving Credit – Key Differences*, EQUIFAX, https://www.equifax.com/personal/education/credit/score/articles/-/learn/revolving-credit-vs-installment-credit [https://perma.cc/7F69-MG5R].

105. *Physician Mortgage Loans*, WHITE COAT INV., https://www.whitecoatinvestor.com/personal-finance/the-doctor-mortgage-loan [https://perma.cc/6G5J-3NBF].

106. *See supra* Section II.C.

but struggled to get into the financial system, companies have started using data such as rent payments and cash on hand to extend credit to these individuals.

That being said, this move toward finding new consumers is not without risk to those consumers. As Abbye Atkinson argues, provisioning credit to certain individuals can actually make them *worse off* than if they were denied credit.[107] Credit can be an important tool to help individuals smooth consumption over time by borrowing against their future wealthier selves. For individuals who will *not* be better off in the future, expanding credit access essentially pushes them further into poverty and redistributes wealth toward wealthier individuals.

B. *Legal Definitions of Fair Lending*

In this Section, we outline the relevant legal definitions for fair lending law. Collectively, these laws represent the backbone of fair lending in the United States. Other scholars have provided extensive overviews of these laws. We focus on the implications for alternative data specifically, highlighting how the laws shape the way credit data is collected and used.

1. Fair Credit Reporting Act

The Fair Credit Reporting Act ("FCRA") was originally passed in 1970 and intended to provide consumers with a process for challenging erroneous information in their credit reports.[108] Among its most famous provisions are those that allow consumers to access free credit reports (after a 2003 amendment),[109] require disclosures when employers and credit lenders use credit reports,[110] and give consumers procedures for correcting erroneous information.[111]

Most relevantly, FCRA gives consumers control over certain negative information about themselves.[112] This requirement was operationalized as

107. Abbye Atkinson, *Rethinking Credit as Social Provision*, 71 STAN. L. REV. 1093, 1147–53 (2019).

108. Fair Credit Reporting Act, 15 U.S.C. §§ 1681–1681w.

109. *Id.* § 1681b(a).

110. *Id.* § 1681b(b)(2).

111. *Id.* § 1681i.

112. Senator William Proxmire, one of the drafters of the FCRA, noted that:

> As I drafted this bill originally, I provided that whenever adverse information is included in any file, the consumer would have to be notified, and I was strongly for that. This was discussed by the committee. It was discussed in the hearings at some length. We were finally convinced that this would involve so much expense and so much difficulty for the credit agencies that they had a legitimate complaint about it.

115 Cong. Rec. 33411 (1969).

The FCRA also explicitly includes certain negative events such as certain bankruptcy events and hard credit checks. It explicitly excludes certain civil judgments, bankruptcies, and tax liens after a certain time period. Senator Proxmire was concerned with regulating "arbitrary, erroneous, and malicious credit information." *Id.* at 33408. Although the definitions of usable credit reporting bureaus needing to provide notice of adverse information in monthly billing or other statements, rather than each and every time it is reported. Legislators were balancing consumer control over adverse information and costs to credit bureaus. The legislative report noted that the law would "prevent consumers from being unjustly damaged because of inaccurate or arbitrary information in a credit report."[113] Much of the current discourse on alternative credit data echoes similar concerns about adverse information.[114]

2. Equal Credit Opportunity Act

The Equal Credit Opportunity Act ("ECOA"), passed in 1974, was explicitly aimed at banning the use of protected characteristics in making credit decisions.[115] It prohibits the use of factors such as race, national origin, and sex in determining credit lending, and provides other protections against using creditor information such as a consumer exercising their rights under other consumer credit legislation.[116] ECOA also requires certain disclosures when credit is denied and levies civil fines for violations of its provisions.[117] One of the consequences of ECOA is that it encouraged the move from narrative credit files to the FICO credit score.[118]

Because of the substantial liability for using protected characteristics in a credit decision, the appeal of computed credit scores that do not *explicitly* use these factors in their calculations is that they sidestep these possible ECOA concerns.

ECOA's prohibitions raise a natural question: What standard should be used to evaluate when discrimination occurs? Common practices at the time of the law's passage, such as denying unmarried women loans, were clear examples of disparate treatment.[119] For the purposes of alternative credit

---

data did not explicitly address this concern, other parts of the FCRA do. For example, the FCRA defines investigative consumer reports as those that contain "information on a consumer's character, general reputation, personal characteristics, or mode of living." 15 U.S.C. § 1681a(e). Any such reports require clear and accurate disclosures to consumers. Although, as discussed earlier, these types of reports became less common as simpler tools like credit scores became more commonplace, the effort does reflect an attempt to give consumers more information about how things resembling alternative credit data are being used in credit decisions. *See id.*

113. S. REP. No. 91-517, at 1 (1969).

114. *See supra* Section I.A.

115. 12 U.S.C. § 1691(a); *see also Equal Credit Opportunity Act (Regulation B)*, NAT'L CREDIT UNION ADMIN. (Jan. 12, 2023), https://ncua.gov/regulation-supervision/manuals-guides/federal-consumer-financial-protection-guide/compliance-management/lending-regulations/equal-credit-opportunity-act-regulation-b [https://perma.cc/Y24D-3W43].

116. *The Equal Credit Opportunity Act*, C.R. DIV., U.S. DEP'T JUST. (Jan. 2, 2025), https://www.justice.gov/crt/equal-credit-opportunity-act-3 [https://perma.cc/3ANQ-QTNE].

117. 12 U.S.C. § 1691b(e); *id.* § 1691e.

118. *See* Lipartito, *supra* note 9, at 3–6.

119. Savannah Peters, *50 Years Ago, It Was Legal to Deny a Woman Credit Without a Male Co-Signer*, MARKETPLACE (Oct. 21, 2024), https://www.marketplace.org/story/2024/10/21/ecoa-e

data, disparate impact claims are potentially more relevant. ECOA incorporates both theories, but disparate impact is notoriously more difficult to prove. Disparate impact analysis requires that a facially neutral policy has an adverse impact on a group with a protected characteristic, unless that impact has a legitimate business need.[120]

With alternative credit data, a number of questions may arise. For instance, does using checking account balance negatively impact male borrowers relative to female borrowers? Do on-time rent payments benefit white borrowers over Black borrowers? These sorts of empirical questions are the heart of disparate impact analysis, but there is relatively little information about the likely effects of proposed alternative credit data on these questions. Although ECOA does create broad antidiscrimination principles in credit lending, and likely helped spur the quantification of credit scores, it is less clear how well it can adapt to alternative credit data measures.

3. Home Mortgage Disclosure Act

The Home Mortgage Disclosure Act ("HMDA"), enacted in 1975, aims to assist with auditing the mortgage lending industry for unfair lending practices.[121] The HMDA requires financial institutions to publicly report data on mortgage applications, including details about the applicants' demographics, loan characteristics, and whether applications were approved or denied.

In the context of alternative credit data, HMDA provides lessons in helping to define what counts as traditional and what counts as alternative data. HMDA specifies what types of data must be reported and focuses on many of the traditional markers of creditworthiness such as debt-to-income ratios, available collateral, and income.[122] These features, reported alongside demographic information, can help regulators uncover patterns of disparate lending. Requiring these specific disclosures has an obvious advantage in that lenders cannot easily game these measures. However, they could also potentially miss the use of alternative data that is used to extend credit to individuals who lack the traditional markers the HMDA focuses on.

4. Dodd–Frank

For alternative credit data, the Dodd–Frank Wall Street Reform and Consumer Protection Act's ("Dodd–Frank") provisions on transparency, fairness, and accountability are particularly relevant. The CFPB has the authority to scrutinize the use of alternative data to ensure it complies with

qual-credit-opportunity-act-legacy-impact-women-credit-lending-mortgage [https://perma.cc/Z38D-Z6DW].

120. *See Section VII- Proving Discrimination- Disparate Impact*, C.R. DIV., U.S. DEPT. JUST., https://www.justice.gov/crt/fcs/T6Manual7 [https://perma.cc/Q3UJ-YHZX].

121. *See* 12 U.S.C. §§ 2801–2811.

122. *Mortgage Data (HDMA)*, CONSUMER FIN. PROT. BUREAU (Dec. 13, 2024, 11:00 AM), https://www.consumerfinance.gov/data-research/hmda [https://perma.cc/USQ2-XATF].

existing consumer protection laws, such as ECOA and HMDA.[123] For instance, the CFPB has the power to examine whether the inclusion of alternative data like rental payment histories or social media data aligns with fair lending principles and does not result in disparate impacts on protected groups.

Additionally, Dodd–Frank empowers the CFPB to enforce corrective measures if alternative data practices are found to be unfair or deceptive.[124] The oversight function provided by the CFPB under Dodd–Frank is therefore essential in monitoring how alternative data is employed. Importantly for moving forward, any likely legal and doctrinal definitions of sufficient nexus tests for alternative data are likely to require the CFPB's approval, and indeed the agency has been active in soliciting information and providing guidance for acceptable uses of alternative credit data.

C. Algorithmic Credit Scoring and the Introduction of Alternative Data

1. Machine Learning and Credit Models

Widespread advances in computation enabled the use of machine learning models in many domains, including credit scoring. At a high level, machine learning differs from older statistical models primarily in that it uses data-adaptive procedures to make the best model.[125] Essentially, the choice of what features to include in a model and how much to weigh each one is taken out of the analyst's hands and instead "learned" by the computer. The major advantage of using machine learning over simple credit models is that with sufficient amounts of data, the data-adaptive machine learning models can uncover patterns that would not be obvious from using traditional models.[126]

Machine learning has made credit scoring models more sophisticated primarily because it facilitates working with "big data." One aspect of this is that machine learning models can more flexibly incorporate hundreds or even thousands of features.[127] The other major advantage of using machine learning is that these models can uncover nonintuitive or complex patterns in large datasets. Andrew D. Selbst and Solon Barocas specifically address this issue.[128] By focusing on legal interventions that tackle the inscrutability of black-box machine learning models, they argue that this effort is ultimately trying to force out information about intuitive relationships between features used in a machine learning model and an ultimate decision about an individual. However, they argue this focus on intuitive explanations is misguided

123. *See* 12 U.S.C. § 5533.

124. *Id.* § 5531 (defining unfair, deceptive, or abusive practices).

125. *See generally* GARETH JAMES, DANIELA WITTEN, TREVOR HASTIE & ROBERT TIBSHIRANI, AN INTRODUCTION TO STATISTICAL LEARNING (2d ed. 2021).

126. *Id.*

127. *Id.*

128. Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 677 (2016).

because machine learning works in part by uncovering nonintuitive relationships and improving predictions through these relationships. Instead of placing the focus on the inputs to models, the authors suggest focusing regulatory attention on transparency and impact analysis concerns.[129]

2. Industry Uses of Alternative Data

Alternative data is already proliferating in various credit lending contexts.[130] Mikella Hurley and Julius Adebayo detail many of the different sources of information that lenders use besides a traditional credit score.[131] For example, ZestFinance uses information such as an individual's technology usage or how often they scroll through terms of service.[132]

Another popular alternative data service is rent reporting. Although information such as mortgage loan payments or revolving debt payments typically gets reflected in a credit score, on-time rental payments ordinarily do not, as these are generally paid by cash or check.[133] Seeing this problem, several fintech solutions emerged and marketed themselves toward renters, particularly younger ones in high-income industries. For example, RentPlus charges a monthly fee to handle a tenant's rent payments and report them to credit bureaus.[134] Mastercard launched the Bilt credit card and partnered with corporate apartment buildings to allow consumers to earn credit card points on rent and report a consumer's rental history to creditors and reflect on-

---

129. *Id.* at 712.

130. Although the primary focus of this piece is the use of alternative data in credit lending, it also has important applications in other highly regulated areas as well. In actuarial assessments, alternative data is increasingly used to price products like homeowners' insurance in the face of climate risks, for example. Some of these alternative data sources include the use of imagery captured by aerial drones to assess an individual's home after natural disasters. *See* Albert Fox Cahn, *Through the Roof: My Journey into the Surreal, Infuriating Future of Homeowners Insurance*, BUS. INSIDER (Aug. 7, 2024, 4:54 AM), https://www.businessinsider.com/homeowners-insurance-nightmare-cancellation-surveillance-drone-ai-future-2024-8 [https://perma.cc/S4KD-AWX9]. In the car insurance industry, vehicle telematics data such as individual driving behavior, vehicle diagnostics, and location data are used to price insurance products. *See* Howard, *supra* note 4. Even pet insurance is not immune—the ASPCA Pet Insurance, for example, uses data from health monitors and GPS trackers to relay information about a pet's behavior and health to veterinarians and insurance companies. *See* Mary Beth Leininger, *Pets and Wearable Tech*, ASPCA PET HEALTH INS., https://www.aspcapetinsurance.com/resources/pets-and-wearable-tech [https://perma.cc/6SZA-8BMK].

131. Hurley & Adebayo, *supra* note 1, at 166.

132. *Id.*

133. *See* DAVID HAO ZHANG, HOW DO PEOPLE PAY RENT? 7 (2016), https://www.bostonfed.org/publications/research-data-report/2016/how-do-people-pay-rent.aspx [https://perma.cc/GW6J-TJU7] (noting that cash is the second most common way of paying rent at twenty-two percent).

134. *Unlock Your Financial Future with RentPlus*, RENTPLUS (2024), https://www.rentplus.com [https://perma.cc/C8N7-WW9B].

time payments in a credit report.[135] These services largely are geared toward well-off renters in high-cost-of-living areas, a group likely seen as low risk but not served by traditional financial products like mortgages.

Other fintech players target lower-income or "credit invisible" portions of the market. Plaid, a digital finance platform, suggests the problem lenders face is that they "struggle to gain a holistic picture of loan applicants due to the amount of manual work and time it takes to collect accurate data from multiple sources" and offers to solve this issue by doing automatic asset verification of consumers by accessing their bank accounts, income, and employment information.[136] Some of these include early access to paychecks, high yield savings accounts, and overdraft protection.

* * *

Algorithmic credit scoring ultimately emerged as a natural result of the quantification of creditworthiness. The history of credit scoring shows that at different times what counted as "data" shifted. The legal landscape of the second half of the twentieth century and early twenty-first century provided the framework for distinguishing between "traditional" data that entered standardized credit reports and "alternative" data that was everything else. But one theme is present in the history of credit scoring, the legal definitions, and the growth of fintech products: the role of intuitive narratives about what makes a person creditworthy.

III. Intuition and Algorithms

Thus far, this Article has made two main points. First, that nexus tests are a poor analytical stand-in for some other concept or multiple concepts. And second, that the delineation of "traditional" and "alternative" data arose out of a causal turn in credit lending that emphasized highly quantified measures with clear causal stories between creditworthiness and data. Thus, this Article's contention is that when regulators are deploying nexus speak around alternative data, what they actually care about is legal causation. But causation is notoriously difficult, so what are its conceptual boundaries in this context? This Part outlines the idea that for credit lending and algorithmic decision-making more broadly, the real problem is deteriorating social oversight.

---

135. AnnaMaria Andriotis, *The Unusual Economics of the Bilt Credit Card*, WSJ PODCASTS: THE JOURNAL. (June 25, 2024, 4:57 PM), https://www.wsj.com/podcasts/the-journal/the-unusual-economics-of-the-bilt-credit-card/96a42e60-e530-4765-a477-5a35a56e3d7b (on file with the *Iowa Law Review).*

136. *See Build a Complete Picture of a Borrower - on One Platform*, PLAID, https://plaid.com/use-cases/lending [https://perma.cc/SAN2-PKBW]. Other examples include MoneyLion, which offers services such as no- and low-interest payday loans and personal loans without hard credit checks, along with a promise to report favorable information to help increase borrowers' credit scores. *Make Money Easy*, MONEYLION, https://www.moneylion.com [https://perma.cc/U62N-5VV6]. A prominent example is Chime, a fintech that targets unbanked and low-income individuals. *See Everyday, Fee-Free Banking*, CHIME, https://www.chime.com [https://perma.cc/4UKB-CET4].

This Part shows why the algorithmic turn in credit lending strains the traditional focus on using intuitive narratives to enable social oversight. First, this Part situates the problem of modeling creditworthiness in social science literature discussing "construct validity." Specifically, it argues that credit data is always just a proxy for some underlying, unmeasurable character trait. Second, this Part complicates the normal focus on "intuition" in nexus tests by bringing forth insights from science and philosophy. It incorporates causal theories from various disciplines to make the case that causation is about winnowing numerous stories down to plausible ones.

A. "Measuring" Creditworthiness

Credit lending is an interesting example of a broader problem in law and the social sciences: How do we measure unmeasurable social concepts? Social science and statistics are filled with examples of attempts to measure some underlying social characteristic that is not obviously quantifiable. How good of an employee is a prospective hire likely to be? Which student is most likely to succeed and deserves to be awarded a scholarship? Which scientific team shows the most promise and should be awarded a grant? These examples show the core of the problem: What is a "good" employee, or student, or scientist? These examples are all exercises in trying to measure some underlying *character trait* that is fundamentally unmeasurable.

So instead, decision-makers rely on various proxies. Perhaps previous work experience means someone will be a good employee because an employer thinks that the employee learned time management, critical thinking, and organization skills. The college admissions officer trying to allocate scholarships may also have a difficult time identifying good students and assumes that high grades and lots of leadership in extracurricular activities proxy for a student's hardworking attitude, ambition, and intellectual curiosity. The officer at the National Institutes of Health might look at how many patents a scientist has as a proxy for productivity and innovativeness. In each of these examples, there is a clear story that can be told about why the proxy is a good indicator of the underlying character trait.

This general idea is a well-known problem in social sciences and statistics. Social scientists generally talk about the notion of *construct validity*, or the idea that the quantity being measured is a valid reflection of the concept it is meant to represent.[137] For example, does a survey asking about participant attitudes toward their friends and family accurately measure underlying psychological traits such as conscientiousness or openness to new experiences? Much of social science research is devoted to interrogating whether constructs are truly

137. *See* Lee J. Cronbach & Paul E. Meehl, *Construct Validity in Psychological Tests*, 52 PSYCH. BULL. 281, 281 (1955) (first coining the term "construct validity").

measuring what they purport to be measuring and how to improve measurements of complex underlying social phenomena.[138]

Credit lending is simply another example of this overall problem. Credit lending is about the underlying character trait of "creditworthiness," but what does it mean to be worthy of credit? There are several theories that could be advanced. One could imagine that creditworthiness is about character traits like responsibility, discipline, and conscientiousness.[139] Alternatively, maybe creditworthiness is simply about how wealthy an individual's family is.[140] Or creditworthiness may be the opposite of both of these and instead lenders are looking for individuals who are likely to accrue lots of interest and carry debt balances.[141]

Why does construct validity matter for credit lending, and legal applications more generally? It matters because it speaks to the requirement credit lending—and the legal system—has for *causal* links between the underlying character trait and the proxy being measured. In each of the examples given, there is always a plausible causal story that could be told.

So too with credit lending. For example, credit history can be a positive indicator because it shows that someone has a history of being responsible with their credit previously.[142] Stationary credit, such as a student loan, can be positive because it indicates that someone is investing in their future, and this can either indicate that they are taking responsibility for their future wealth, or even very basically suggest they can borrow against a likely higher future income stream.[143]

---

138. *See* Jessica K. Flake, Jolynn Pek & Eric Hehman, *Construct Validation in Social and Personality Research: Current Practice and Recommendations*, 8 SOC. PSYCH. & PERSONALITY SCI. 370, 370 (2017); *see also* Jennifer H. Arlen & Eric L. Talley, *Experimental Law and Economics*, at xvi–xxvi (N.Y.U. Ctr. for L. & Econ., Working Paper No. 08-30, 2008), https://scholarship.law.columbia.edu/faculty_scholarship/1539 [https://perma.cc/PT5T-9FUZ].

139. *See* Claire Greene, Oz Shy & Joanna Stavins, *Personality Traits and Financial Outcomes* 2–3 (Fed. Rsrv. Bank Bos., Working Paper No. 23-4, 2023), https://www.bostonfed.org/publications/research-department-working-paper/2023/personality-traits-and-financial-outcomes.aspx [https://perma.cc/A5RP-F4B9]; *see also* Emily Rosamond, *"All Data Is Credit Data": Reputation, Regulation and Character in the Entrepreneurial Imaginary*, 25 PARAGRANA 112, 113 (2016), https://www.degruyter.com/document/doi/10.1515/para-2016-0032/html [https://perma.cc/3KXN-MS7E].

140. *See* Matteo Benetton, Marianna Kudlyak & John Mondragon, *Dynastic Home Equity* 44–45 (Fed. Rsrv. Bank S.F., Working Paper 2022-13, 2022), https://www.frbsf.org/wp-content/uploads/sites/4/wp2022-13.pdf [https://perma.cc/Z8H3-9SP9].

141. *See* Sandra E. Black & Donald P. Morgan, *Meet the New Borrowers*, 5 CURRENT ISSUES ECON. FINS., Feb. 1999, at 5. In practice, credit lending seems to reflect all these theories in various ways. Different credit lenders may have different character traits in which are interested. In turn, they may use different constructs to measure these different character traits. *Id.* at 2–3.

142. Bev O'Shea, *How Length of Credit History Affects Your Credit Scores*, NERDWALLET (Apr. 22, 2024), https://www.nerdwallet.com/article/finance/credit-age-length-of-credit-history [https://perma.cc/7AUF-ZL5F].

143. Ben Luthi, *Do Student Loans Help Build Credit?*, EXPERIAN (Apr. 15, 2024), https://www.experian.com/blogs/ask-experian/do-student-loans-help-build-credit [https://perma.cc/35WY-8EJ8].

Likewise, we can tell stories about negative character traits. A history of delinquent payments can suggest someone is irresponsible and inattentive. Revolving credit such as credit card debt can indicate that someone spends recklessly and is therefore an undisciplined individual.[144] Statutes and regulations have emphasized these causal stories in credit lending because they sound plausible and ultimately create an intuitive link between the measurement and the underlying character trait. In short, the legitimacy of credit decisions rests on the idea that there are narratives that society is willing to accept as valid reasons to extend or deny credit.

In Part V, we draw on these basic observations about measuring credit-worthiness in the context of algorithmic gaming. As a conceptual matter, scrutinizing gameability is a helpful way to understand what nexus tests are aiming to do, and why it is a clearer way of understanding what is at stake with credit lending and alternative data. Gaming severs the causal link between the underlying character trait and its construct. Someone who takes out a stationary loan to pay off a revolving one so that they can continue to spend lavishly on luxury goods does not change their underlying character trait of being fiscally irresponsible. Instead, they changed the constructs used to measure fiscal responsibility and therefore changed outside assessment of that character trait.

Once a construct becomes gameable, it no longer has a causal link to the underlying thing it is measuring. Despite what much of the confused discussion of nexus tests suggest, the normative stakes of alternative data in credit lending are not wholly about good or bad uses of data. Instead, the stakes are about how gameable constructs are—and therefore whether they provide causal stories that society can evaluate.

B. *Complicating Causation*

Thus far, we have treated causality as the ability to tell a plausible story. As long as there is an intuitive narrative connecting a piece of data and underlying creditworthiness, regulators and financial institutions treat this relationship as "causal." Yet in other fields, causation and intuition are separate concepts. Delving into both scientific and philosophical approaches around these questions of causation and intuition can help illuminate what is at stake for the law.

1. Statistical and Scientific Approaches

While legal interpretations of causality often rely on intuitive reasoning, statistical and scientific approaches address the question differently. Intuition does play an important role in hypothesis generation, but oftentimes empiricists test *counterintuitive* relationships. Further, statistics is generally concerned

---

144. Erin El Issa, *Why Do People Call Credit Card Debt 'Bad' Debt?*, NERDWALLET (July 21, 2023, 12:25 AM), https://www.nerdwallet.com/article/credit-cards/credit-card-debt-bad-debt [https://perma.cc/5XHJ-N4MM].

with disentangling causation from correlation.[145] Even if a correlational relationship between two quantities makes sense intuitively, it would not be considered *causal* unless it is tested under certain assumptions and criteria.

One major lesson is that "selection"—or an individual opting into some benefit—breaks causation. When people can choose to enter or exit the benefit, we can never be sure whether any observed effects are because of that benefit or something else. This concept is central to the "Neyman-Rubin causal inference framework," also known as the potential outcomes framework.[146] When "units" (people, cities, states, etc.) select into treatment, they essentially game the mechanism such that we do not know if any effects are because of the treatment itself or because of underlying characteristics that correlate with selection into treatment *and* the outcome. For example, does drinking red wine every day improve heart health, or are the type of people who regularly drink red wine also the sort who are wealthier and have more time to eat well and exercise? The kind of person who selected into treatment (drinking red wine every day) is very different than the kind who does not, making it difficult to disentangle the effect of red wine from any other correlated behaviors.

Another potential stumbling block for proving causation is the existence of "backdoors" that break the causal link between two variables. A backdoor path is one that correlates with both an outcome (e.g., creditworthiness) and a treatment (e.g., on-time payment history).[147] The Pearl framework helps visualize these potential "gaming" opportunities and allows analysts to construct their statistical models in a way that "blocks" these backdoor paths.[148]

The bar for causation in statistics is higher than in law in that statistical approaches require ruling out *all* other explanations, whereas law generally requires that there exists *a* plausible story.[149] That said, legal causation does not imply an endless sea of possible acceptable stories. Much of causation jurisprudence is about winnowing down the set of plausible stories.[150] Although credit lending law does not require that all potential other explanations are discarded, the lessons about selection effects and backdoor paths show how intuition can differ from empirical reality.

145. *See* JOSHUA D. ANGRIST & JÖRN-STEFFEN PISCHKE, MASTERING 'METRICS: THE PATH FROM CAUSE TO EFFECT, at xiii (2014).

146. A full explanation of the Neyman-Rubin framework is outside the scope of this Article, but for an introduction to the framework for empirical legal studies audiences, see Daniel E. Ho & Donald B. Rubin, *Credible Causal Inference for Empirical Legal Studies*, 7 ANN. REV. L. & SOC. SCI. 17, 21–22 (2011).

147. *See generally* JUDEA PEARL & DANA MACKENZIE, THE BOOK OF WHY: THE NEW SCIENCE OF CAUSE AND EFFECT (2018).

148. The Pearl causal model, introduced by Judea Pearl, offers a more structural approach to understanding and testing causal relationships by using graphs to illuminate assumed associations between variables. *See id.*

149. *See, e.g.*, Kingston v. Chicago & N.W. Ry. Co., 211 N.W. 913, 915 (Wis. 1927).

150. *See, e.g.*, *id.* at 915 (illustrating that when multiple causal stories exist—here, two converging fires—courts must determine which narrative establishes liability).

2. Overview of Intuitive Theories

Intuitive theories play an important role in how people process and understand information. Cognitive scientists have long recognized that people organize basic knowledge about the world in ways that are seemingly analogous to scientific theories. This organizational structure is essential for how people understand causal relationships.

In turn, this intuitive theory of causal relationships provides the cognitive framework for providing explanations for things in the world and predicting how future events will unfold.[151] More specifically, intuitive theories provide the mental scaffolding to provide support for acquiring new causal knowledge, determining people's judgments about possible causal relationships, and guiding people's interpretation of data they observe.[152]

The presence, content, and effect of intuitive theories are corroborated by extensive experimental evidence. One foundational study demonstrated that people have a strong tendency to understand interactions between moving shapes on a screen in terms of rational agents with intentions, beliefs, and desires.[153] To that end, study participants attributed specific mental states to shapes interacting with each other. Similarly, other researchers crafted a study where participants observed images of discs moving on a screen. When one disc bumped a second disc and the second one began to move immediately, study participants stated that the first disc caused the second to move; thus, demonstrating a willingness to import intuitive notions of cause and effect.[154]

Many different intuitive systems have been uncovered through experimental evidence and make up a body of working knowledge that is often referenced

151. *See* Joshua B. Tenenbaum, Thomas L. Griffiths & Sourabh Niyogi, *Intuitive Theories as Grammars for Causal Inference*, *in* CAUSAL LEARNING: PSYCHOLOGY, PHILOSOPHY, AND COMPUTATION 301, 301–02 (Alison Gopnik & Laura Schulz eds., 2007).

152. *See id.*; *see also* Michael J. Pazzani, *Inducing Causal and Social Theories: A Prerequisite for Explanation-Based Learning*, *in* PROCEEDINGS OF THE FOURTH INTERNATIONAL WORKSHOP ON MACHINE LEARNING 230, 230–31 (1987).

153. For instance, in a video displaying shapes bouncing into each other, study participants attributed mental states to these shapes to make sense of their interactions. Fritz Heider & Marianne Simmel, *An Experimental Study of Apparent Behavior*, 57 AM. J. PSYCH. 243, 246 (1944).

154. By contrast, when the second disc began to move one fifth of a second after contact with the first, study participants did not induce a causal relationship even where there was perfect correlation. A. MICHOTTE, THE PERCEPTION OF CAUSALITY 92–93 (T.R. Miles & Elaine Miles trans., 1963).

as "folk" knowledge, such as "folk psychology,"[155] "folk sociology," or "folk economics."[156]

Intuitive theories, like scientific theories, are based on systems of causal laws. However, these theories often posit unobservable entities ("forces," "germs," "beliefs," or "traits") that causally operate to produce observable phenomena.[157] Moreover, these causal entities are used to explain various phenomena. Taking "traits" as causal entities allows people to understand and predict behavior in reference to certain traits that a person possesses.

Data analysts have uncovered a trove of information that can be used to predict how risky it is to extend credit to any individual. J.P. Martin—who was described by the New York Times as "a math-loving" Canadian credit executive[158]—was one of the pioneers of using alternative data to forecast credit risk. Martin's data demonstrated a litany of relationships between behavior and credit risk. For example, people who purchased "cheap, generic automotive oil" were more likely to miss credit card payments than someone who purchased more expensive oil.[159] People who bought carbon monoxide monitors or furniture sliders to protect their floor from scratching made good on their repayments almost invariably.[160]

Often the causal story that underlies the relationship between alternative data and creditworthiness is told through referencing character traits. Returning to the automobile oil example, commentators suggest different intuitive stories for why people who purchase expensive oil are more likely to pay back their loans. For some, the choice to purchase expensive oil demonstrates

155. *See* Johannes B. Mahr & Gergely Csibra, *Why Do We Remember? The Communicative Function of Episodic Memory*, BEHAV. BRAIN SCI. 2 (Jan. 19, 2017), https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/why-do-we-remember-the-communicative-function-of-episodic-memory/CC092B430A7213B4C9BF2C4B48415B64 [https://perma.cc/2LLR-GBQN] ("Overall, our strategy is to reason from form to function: From the design features of the episodic memory system identified at the outset, we will infer the cognitive tasks this system has likely been selected to solve.").

156. *See* Paul H. Rubin, *Folk Economics*, 70 S. ECON. J. 157, 157 (2003) ("Although psychologists have studied 'folk psychology,' 'folk physics,' and 'folk biology,' among others, there has been less attention paid to 'folk economics'—that is, the economic notions that naïve (untrained) individuals have and the perceptions of such individuals about the economy." (footnote omitted)).

157. Johannes B. Mahr & Gergely Csibra, *A Short History of Theories of Intuitive Theories*, *in* A LIFE IN COGNITION 219, 220 (Judit Gervain, Gergely Csibra & Kristóf Kovács eds., 2022); *see also* Tobias Gerstenberg & Joshua B. Tenenbaum, *Intuitive Theories*, *in* THE OXFORD HANDBOOK OF CAUSAL REASONING 515, 516–17 (Michael R. Waldmann ed., 2017).

158. Charles Duhigg, *What Does Your Credit-Card Company Know About You?*, N.Y. TIMES (May 12, 2009), https://www.nytimes.com/2009/05/17/magazine/17credit-t.html (on file with the *Iowa Law Review*).

159. *Id.*

160. *Id.*

foresight and conscientiousness, which is indicative of character traits that make a person more likely to pay back a loan.[161]

The oil example portrays how intuitive theories are underdetermined. That is, a certain behavior is compatible with several underlying causal phenomena that are all consistent with the theory.[162] In other words, different people can incorporate the data about people who purchase expensive oil being more likely to pay back a loan and create different causal stories that are all consistent with an intuitive theory of folk psychology. For some, the choice to use expensive oil indicates some conscientiousness, and conscientious people are more likely to pay back loans. An entirely different possibility is that rich families have a propensity to drive expensive cars that need this oil, and they pass along this propensity to their children.

Taken together, the choice of purchasing expensive motor oil is potentially indicative of many different underlying character traits that could each make it more likely that a person pays back a loan. Of course, there are some background assumptions at play in the theory as well, such as the belief that people typically act in accordance with character traits that they demonstrate. There would be little predictive benefit of discerning relevant character traits in a person's past behavior if people were unlikely to act in accordance with these traits in the future.[163]

3. Intuitive Theories in the Law of Machine Learning

Andrew Selbst and Solon Barocas argue that calls for algorithmic explainability are grounded less in a desire for technical transparency but rather in a deeper need for normative legitimacy. Their central claim is that explainability serves as a bridge between empirical observation and normative evaluation, and that this bridge often relies on intuition.[164] On this front, Selbst and Barocas distinguish between two distinct challenges posed by machine learning: inscrutability and nonintuitiveness.[165] Inscrutability occurs when technical complexity resists direct inspection, and nonintuitiveness occurs when algorithmic reasoning fails to cohere with human expectations. While legal and technical solutions to inscrutability have received considerable attention, they argue that the problem of nonintuitiveness is more difficult and socially consequential.

---

161. *See* Jane Bambauer, Tal Zarsky & Jonathan Mayer, *When a Small Change Makes a Big Difference: Algorithmic Fairness Among Similar Individuals*, 55 U.C. DAVIS L. REV. 2337, 2352 (2022); *see also* Dennis Hirsch, *Predictive Analytics Law and Policy: A New Field Emerges*, 14 I/S 1, 2 (2017).

162. *See* Gerstenberg & Tenenbaum, *supra* note 157, at 516.

163. What role do intuitive theories play, truth tracking or social function? This is important for determining whether credit scoring models should comply with intuitive theories. S*ee* Mahr & Csibra, *supra* note 157, at 219.

164. *See* Selbst & Barocas, s*upra* note 7, at 1128.

165. *Id.* at 1089.

In their discussion of nonintuitiveness, Selbst and Barocas demonstrate that intuitive explanations play an essential role in enabling public accountability. Yet they also warn that intuition is fraught with risk. People tend to apply intuitive reasoning inconsistently, often demanding explanations only when outcomes seem unjust or surprising.[166] Moreover, stakeholders may accept post hoc narratives that feel satisfying without meaningful empirical grounding. This asymmetry allows systems to avoid scrutiny when outcomes align with expectations and obscures potential harms that lack an intuitive storyline. In this way, the intuitive appeal of an explanation may obscure rather than reveal the normative foundations of algorithmic decision-making.[167]

To address this problem, Selbst and Barocas advocate a shifting away from intuitive explanation and toward institutional mechanisms that promote meaningful evaluation. Their proposed solution centers on documentation and deliberation. Rather than requiring that models themselves be rendered explainable, they argue that regulators should focus on understanding the design process, including the tradeoffs, assumptions, and constraints that shape model development. This process-oriented approach does not guarantee that any given model will be interpretable, but it creates a paper trail of accountability that can be evaluated in light of legal and ethical standards. In doing so, it reframes explainability not as a purely technical property, but as a governance challenge that requires organizational transparency, public reasoning, and institutional design.[168]

This Article takes this central insight from Selbst and Barocas as a starting point and examines how it unfolds in an emerging field: alternative credit data. As discussed earlier, the "causal turn" in credit scoring during the late twentieth century embraced inputs that were both predictive and made intuitive sense. Legal and regulatory actors gravitated toward variables that seemed to track traits like responsibility and stability—paying rent on time, maintaining a long credit history—even when the statistical justifications were often correlational rather than truly causal. In this context, intuition played a dual role: It helped policymakers and courts justify the use of certain data points, and it enabled public oversight by allowing observers to understand (or at least believe they understood) why a system made the decisions it did.

Beyond Selbst's and Barocas's treatment of the issue, the role of intuitive theories in the law has been examined across a variety of legal domains. Within discussions of the appropriate governance systems for algorithmic

166. *Id.* at 1096–99.

167. Likewise, they caution against the common assumption that intuitive explanations necessarily reflect valid reasoning. *Id.* at 1129. They note that machine learning's greatest value lies in uncovering patterns that defy human expectations, and that this very feature clashes with our desire for models that "make sense." *Id.* This disconnect leads to a regulatory paradox: Models that are accurate may be normatively suspect because they are nonintuitive, while models that are intuitive may offer false comfort despite poor predictive performance. *Id.*

168. *Id.* at 1129–38.

decision-making, commentators principally focus on how intuitive theories provide a schema for providing explanations for algorithmic systems. When some variable produces a change to some output, administrators of these technical tools rely on intuitive theories to craft some causal story about how the variable is linked to output.

Consider the Fourth Amendment's probable cause requirement for a warrant. As Kiel Brennan-Marquez explains, there is a widely held intuition that pure statistical probability is insufficient to satisfy the probable cause standard.[169]

And moreover, the uneasiness that individuals have with purely statistical grants of probable cause seems to be borne out in the courts as well.[170] In other words, even if an algorithmic system that was vetted to be perfectly reliable and accurate indicated that a particular residence had an eighty percent chance of containing drugs, there are strong reasons to think that this is insufficient for probable case.[171]

Instead, the Fourth Amendment's probable cause standard implies some level of plausibility; that is, an officer swearing an affidavit for a search warrant must be able to weave some sort of causal story about how the evidence makes it probable that the items listed in the search warrant are likely to be found.[172] By requiring plausibility in the probable cause analysis, there is an implicit requirement for an explanation that provides an additional layer of accountability to law enforcement.[173]

Moreover, the plausibility requirement allows a magistrate to assess whether law enforcement's causal story is more plausible than a more innocent explanation about the relationship between the evidence and the determination.[174] For instance, it is up to a magistrate to determine if increased power usage at a residence is better explained by that residence housing a marijuana growing operation that requires power for greenhouse lamps or some other innocent or innocuous causal story.[175]

Other commentators have gravitated toward intuitiveness as a check on machine learning models. James Grimmelmann and Daniel Westreich appeal to intuitiveness in order to determine if a model is relying on improper data.[176] Grimmelmann and Westreich imagine a scenario where a machine learning model relies on musical taste to determine job performance. However,

---

169. Kiel Brennan-Marquez, *"Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249, 1252 (2017).

170. *See* Emily Berman, *Individualized Suspicion in the Age of Big Data*, 105 IOWA L. REV. 463, 465 (2020); Jane Bambauer, *Hassle*, 113 MICH. L. REV. 461, 462 (2015).

171. *See* Brennan-Marquez, *supra* note 169, at 1253–54.

172. *Id.* at 1256.

173. *Id.* at 1292–94.

174. *Id.* at 1253.

175. *Id.*

176. *See generally* James Grimmelmann & Daniel Westreich, *Incomprehensible Discrimination*, 7 CALIF. L. REV. ONLINE 164 (2017).

they further stipulate that musical taste is correlated with protected class membership.[177] The problem, for Grimmelmann and Westreich, is that we may not be able to tell whether the model is relying on musical taste or protected class membership.[178] Because of this ambiguity, they suggest a default rule that if one variable is nonintuitive then the model is actually relying on the inappropriate data.[179] In other words, if we cannot weave an intelligible story about how musical taste is connected to job performance, then the model is impermissibly using protected class membership for its determinations.[180]

One complication is that Grimmelmann and Westreich do not provide a metric for edge cases where there is some intuitive story about a variable's connection to output, but it is not particularly compelling. One puzzle with intuitive explanations is that they are hard to compare and defy evaluation. As cognitive scientists show, many people refer to intuitive explanations in a way that resists the need for additional evidence. When pushed further, people often claim that an intuitive explanation is just obvious in a way that shows there is no need for additional support.

Other commentators have argued for using intuitive narratives as a check on algorithmic systems. For instance, Pauline Kim points out that various data indicate that employees who installed web browsers that did not come with their work computers stayed longer on the job.[181] Further, Kim claims that if we cannot craft a story about why installing a new browser is evidence for staying longer at a job and the model has a disparate impact, then it should be illegal.[182]

However, intuitive theories may be imperfect at parsing improper use of models. What this test would do well is provide solid determinations when no story could be told. However, it is imperfect because there may be situations where we can weave a story to explain the connection between data and conclusion, yet this may not be the relationship that the model identifies. Taking Kim's example, it could be that people who plan to stay longer on the job customize their work environments—they decorate their offices, and they customize their work computers by downloading their preferred web browser.

One of the upshots of their argument is that model complexity—that is, machine learning models that use hundreds or thousands of features—is the

177. *Id.* at 173.

178. *Id.*

179. *Id.* at 175–76.

180. *Id.* One complication is that Grimmelmann and Westreich do not provide a metric for edge cases where there is some intuitive story about a variable's connection to output, but it is not particularly compelling. One puzzle with intuitive explanations is that they are hard to compare and defy evaluation. As cognitive scientists show, many people refer to intuitive explanations in a way the resists the need for additional evidence. When pushed further, people often claim that an intuitive explanation is just obvious in a way the shows there is no need for additional support.

181. Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 922 (2017).

182. *Id.* at 922–25.

source of this tension. Advances in computing allow for sophisticated models that can uncover these nonintuitive relationships from high-dimensional feature spaces, rendering strong predictions using feature combinations that would make little sense to a human being. When thinking about alternative data, one lesson here might be that certain types of fintech can more easily pass a sufficient nexus test than others. A simple product that uses a handful of features like rental payment history and checking account balance would not be overly complex and would be scrutable for a human being. On the other hand, a financial product that uses thousands of small user interactions with various websites such as time spent scrolling, propensity to click on certain colored buttons, cookies, etc. could fail a sufficient nexus test because any prediction would be a complex combination of these thousands of variables.

* * *

Causation as intuition is a useful framework because of its roots in law, game theory, statistics, philosophy, and machine learning. While each of these fields thinks about causation in different ways, and causation even within law is a notoriously difficult concept, this interplay provides a way to address causal requirements in credit lending. Some basic tensions emerge from this discussion: Measurement requires assumptions about the link between data and character traits, rigorous theories of causation require getting the "right" answer while the law requires a "plausible" answer, and machine learning often enables *counterintuitive* inferences that are nonetheless empirically correct. Having identified these tensions, the next Part turns to how to manage them in algorithmic decision-making contexts.

IV. From Nexus Tests to Social Oversight

Nexus tests are a stand-in for ensuring the legitimacy of a decision by requiring that input data is sufficiently related to creditworthiness. However, nexus tests fail to provide meaningful guidance about the shape of this relationship. Worse still, nexus tests fail to track the normative concerns that motivate regulators to use these tests in the first instance. At least one justification for the legitimacy of credit lending decisions is the idea that data about an individual is causally linked to their underlying "creditworthiness."

Because of these failures, this Part argues that regulators should focus attention on what actually motivates them: a desire to enable social oversight of acceptable algorithmic decision-making processes. First, it discusses the idea that the legitimacy of credit data has largely stemmed from the ability to tell plausible intuitive or "causal" stories. Second, it looks at how machine learning has complicated this approach because it enables a plethora of plausible ex ante *and* ex post stories. Third, it resolves this tension by suggesting that regulators expand their toolkit by focusing on empirical reality.

A. THE LEGITIMACY OF CREDIT DATA

Legitimizing credit data starts with being able to tell a causal story linking data to an underlying characteristic. The "causal turn" in credit lending was an important development because it was seen as necessary for establishing the legitimacy of credit scoring. Even with alternative data, regulators have fixated on mechanisms like cash flow because of the easy causal stories one can tell between that piece of data and underlying creditworthiness.

Although the causal story is the start of the legitimation process, it is not the end, as society imposes other constraints as well. In other words, even when some data is predictive of creditworthiness, it is not necessarily normatively appropriate to use this data for credit determinations. Most obviously, some data may correlate to creditworthiness, but the only causal story that explains why this data correlates hinges on the data's connection to protected characteristics (such as race or gender).

However, even when data does not correlate to protected classifications, legitimation is still subject to scrutiny. In some cases, the causal story that connects alternative data with creditworthiness is socially undesirable, even if it does not implicate a protected characteristic. For example, marital status is not used in determining creditworthiness. Even if marital status was predictive of creditworthiness, there may be reasons to limit its use in credit decisions.

The causal story that underlies the predictive power of marital status data could potentially turn on the fact that unmarried people have less familial support with repaying a loan. However, whether it is fair and legitimate to distribute credit based on family support is subject to social oversight. The principal concern with alternative data is whether a particular piece (or cluster) of data admits of a plausible causal story that can be evaluated for its desirability. Distributing credit based on familial support may be acceptable or not. Yet, this determination should be made according to public consideration.

This social oversight process operates in practice to alter which features of a model are considered accepted. To that end, in 2024 the CFPB issued a rule that restricted the use of medical debt in credit reports.[183] Then-Vice President Kamala Harris praised the CFPB's decision by saying that medical emergencies should not bar economic opportunities.[184] Evaluating the normative valence of using medical debt data to determine credit scores is only possible because medical debt is amenable to plausible causal stories about why it is predictive of credit score. And further, these plausible causal stories are subject to social evaluation which, in turn, legitimates the use of some piece

183. Stefanie Jackman, Kristen Eastman, Jonathan Floyd, David N. Anthony & Ethan G. Ostroff, *CFPB Intensifies Scrutiny on Medical Debt Collection Practices*, TROUTMAN PEPPER LOCKE (Oct. 3, 2024), https://www.consumerfinancialserviceslawmonitor.com/2024/10/cfpb-intensifies-scrutiny-on-medical-debt-collection-practices [https://perma.cc/N4L9-CKM4].

184. Darreonna Davis, *New Biden Administration Rule Would Ban Medical Debt from Credit Reports*, 19TH (June 11, 2024, 2:56 PM), https://19thnews.org/2024/06/biden-medical-debt-harris-rule [https://perma.cc/C4F8-JHVZ].

of data. While medical debt may be predictive of ability to repay a loan, there is widely shared moral concern that economic prospects should not be limited by medical hardship.

Any system of credit scoring that uses alternative data should preserve society's ability to evaluate whether the causal stories connecting the data to creditworthiness are acceptable. Of course, these determinations may shift over time. For instance, society may accept some causal story at one time but later reject it. The next two Sections examine how the social machinery that legitimates alternative data may short-circuit. They also offer regulatory responses that preserve the role of the public in overseeing alternative data.

B. *Ex Ante and Ex Post Intuition in Crafting Causal Stories*

Machine learning programs are often able to ascertain previously unknown correlations. As a result, data scientists can uncover patterns that were previously unrecognized. The possibility of finding statistical relationships in the vast firmament of data has led commentators to assess the stakes of this transformation. Scholars have recognized the role that intuition (or plausibility) performs in facilitating democratic oversight and ensuring value pluralism.[185] However, machine learning potentially upends intuitive relationships, leading to an era of "impossible-to-understand reason."[186] In such a system, advertisers may soon market products to individuals based on seemingly unrelated data such as a person's breakfast choices.

The era of "impossible-to-understand reason" is predominately a product of the nonintuitiveness of machine learning models.[187] Machine learning models are nonintuitive when it is difficult (or impossible) to construct a plausible story that explains the statistical relationships in a model.[188] Constructing these stories is difficult in certain machine learning settings because the models may be "black boxes" where learned statistical relationships cannot be scrutinized,[189] or because certain methods such as clustering find correlations that would not occur to humans.[190]

Returning to alternative credit data, credit scoring models could certainly use data that is nonintuitive. For instance, sophisticated machine learning models may uncover a statistical relationship between whether a person decorates their house with Christmas lights or a simple Christmas wreath and

185. *See* Brennan-Marquez, *supra* note 169, at 1256–58.

186. Paul Ohm, *The Fourth Amendment in a World Without Privacy*, 81 MISS. L.J. 1309, 1318 (2012).

187. *Id.*

188. Selbst & Barocas, *supra* note 7, at 1097.

189. FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 6–9 (2015).

190. For example, outside of social contexts, this technology is used to find new protein structures by looking for correlations that would be impossible for humans to find because of scale and complexity. *See, e.g.*, John Jumper et al., *Highly Accurate Protein Structure Prediction with AlphaFold*, 596 NATURE 583, 583–84 (2021).

the statistical likelihood that they will repay a loan. The choice of Christmas decoration, then, would be nonintuitive alternative data because it is exceedingly difficult to craft a plausible story about how this choice bears some relationship to creditworthiness.

The role of intuition is more complicated than much of the literature recognizes. In the turn from traditional credit data to alternative data, intuition operates at two different moments in time. Credit lending law has traditionally focused on ex ante considerations: whether the use of a particular piece of data is legitimate because it is causally linked to creditworthiness. If a piece of data is causally related (e.g., length of credit account, credit-to-debt ratio etc.), it is permissible to use it for credit decisions. Conversely, if ex ante there is not an obvious story (e.g., scroll speed on a website, social media connections), this would usually be seen as illegitimate.

Alternative data complicates this picture because it permits certain ex post rationalizations of causal stories. Returning to the Christmas decorations example, despite the difficulty of coming up with a plausible ex ante story, there are potential ex post explanations. Perhaps a big-box retail store analyzes customer data and finds that customers who buy wreaths and not lights also tend to pay by credit card. Although buying lights is not causally connected to using credit cards, it could be mediated by a third variable: the consumer's income. The store may then offer these customers the opportunity to apply for a store-branded credit card and offer financing on certain items so that the store may benefit from the customer's interest payments. This example is not purely hypothetical—many retailers analyze consumer data to offer this kind of "embedded lending."[191]

Historically, credit lending law focused on these ex ante stories because of fairness concerns. However, alternative data opens up the possibility of plausible ex post rationalizations. Thus, while intuition may be necessary for creating the conditions of social oversight, the long-held assumption that it must be ex ante intuition likely should be revisited with the advent of machine learning in credit lending.

C. Empirical Reality as a Basis for Social Oversight

The central challenge regulators face with alternative credit data is that machine learning breaks the intuitive and causal narratives that traditionally enabled social oversight. Credit determinations are not merely technical outputs; they are social decisions that must be legitimate in the eyes of the public. For decades, legitimacy in credit lending rested on the ability to tell an intuitive story. These stories formed the basis for public trust and regulatory

191. Louis Thompsett, *Embedded Lending & Credit: Enhancing Consumer Experiences*, FINTECH MAG. (May 1, 2024), https://fintechmagazine.com/articles/embedded-lending-credit-enhancing-consumer-experiences [https://perma.cc/TGQ2-E89E].

oversight. But as discussed earlier, these intuitive stories did not always have to be "right," just plausible.

Machine learning models complicate this problem further because they are designed precisely to uncover relationships that elude intuition. They do not begin with a hypothesis about what should predict creditworthiness. Rather, they optimize for prediction, often surfacing correlations that defy human understanding. This makes intuitive or causal explanation, whether ex ante or ex post, less feasible.

The appeal of the intuitive approach is clear: Intuitive stories help regulators and the public evaluate whether data is being used in a way that aligns with social values. However, once machine learning enters the picture, the reliability of intuitive storytelling as a regulatory tool diminishes. Sometimes a plausible story can be crafted ex post, but it may give consumers little comfort when it is always possible to craft a story even if it is not the actual basis for a decision.

This tension underscores the need for a regulatory pivot. Instead of relying on whether an input feels causally connected to creditworthiness, regulators should focus on whether a model's predictions are empirically grounded. In other words, oversight should shift from intuition to evidence. This means subjecting algorithmic outputs to rigorous testing, examining not only their predictive validity but also the social consequences they produce—particularly regarding fairness, transparency, and susceptibility to gaming.

This empirically grounded approach offers several advantages. First, it re-centers the discussion on what credit scores are supposed to do: accurately predict repayment risk. Second, it invites a richer tool kit for regulation, including empirical audits, counterfactual testing, and causal inference methods. Third, and most importantly, it maintains space for social oversight not by insisting on causal stories that machine learning can no longer reliably support, but by requiring that the outcomes of algorithmic decisions be observable, testable, and justifiable in terms of their real-world effects.

Regulators should therefore reorient their oversight frameworks around the empirical performance of credit models, not the plausibility of the stories they tell. In doing so, they preserve the core aim of causality and intuition in lending law: to ensure that credit determinations reflect meaningful and fair assessments of individuals. But rather than using intuition as a proxy for fairness, they should treat fairness and legitimacy as questions answerable through empirical inquiry.

Such an empirical focus already has a strong basis in the legal literature. Talia Gillis argues that input scrutiny will never achieve disparate impact's goals, and instead bias should be evaluated on the outcomes of a predictive exercise.[192] To put a fine point on this, she specifically argues that protected characteristics such as race and gender should be included in model training

192. Talia B. Gillis, *The Input Fallacy*, 106 MINN. L. REV. 1175, 1243–51 (2022).

so that outcomes can be properly evaluated.[193] Nikita Aggarwal argues that ex ante regulation of data flows makes it difficult for regulators to regulate ex post uses of data. She argues that the ultimate use of data is the proper place for regulatory activity rather than the inputs or flows of data.[194]

Focusing on empirical outcomes rather than intuitive justifications also aligns with broader shifts in regulatory philosophy. As machine learning models grow in complexity and opacity, legitimacy can no longer rest solely on whether decisions can be explained in ways that feel satisfying. Instead, legitimacy must come from whether those decisions work in practice. This pivot does not mean wholesale abandoning intuition and existing legal frameworks, but rather updating the tools used to realize them in an era of statistical prediction. Social oversight, then, must be grounded not only in whether we can tell a good story, but also whether we can evaluate and contest what the story does.

* * *

Nexus tests fail because they conflate different normative concerns—intuition, fairness, causation, and accuracy—into a single, unworkable concept. This Part showed that a better path forward lies in reclaiming the core regulatory function of enabling social oversight. This means asking not whether a piece of data feels connected to creditworthiness, but whether it functions in ways that serve justifiable policy goals. To do so, regulators must embrace empirical methods and treat legitimacy not as a rhetorical exercise, but as a measurable, testable, and contestable standard. In the next Part, we show how this might work in practice with a core problem in the algorithmic decision-making literature: gaming.

V. CASE STUDY: ALGORITHMIC GAMING

How would a focus on empirical reality operate in practice? To address this question, we turn to a recent debate in the law and technology literature: algorithmic gaming. Algorithmic decision-making systems rely on troves of data to make their predictions, but these predictions can falter when the data is bad.[195] Bad data can come in the form of unrepresentative samples, missing data, or insufficient quantity.[196] Gaming presents a different problem: What if

---

193. Talia B. Gillis, *Orthogonalizing Inputs*, ASS'N FOR COMPUTING MACH.: SYMP. ON COMPUT. SCI. & L. 2 (Mar. 2024), https://dl.acm.org/doi/pdf/10.1145/3614407.3643698 [https://perma.cc/Z9K5-9LC6].

194. Nikita Aggarwal, *Locating Consumer Financial Regulation*, 46 CARDOZO L. REV. 927, 929–35 (2025).

195. The machine learning literature generally uses the phrase "Garbage In, Garbage Out" to describe the general problem of statistical modeling being unable to overcome the problem of bad data. *See* Ron Ozminkowski, *Garbage In, Garbage Out*, TOWARDS DATA SCI. (Nov. 13, 2021), https://towardsdatascience.com/garbage-in-garbage-out-721b5b299bc1 [https://perma.cc/7SQZ-D2JT].

196. *See generally* Selbst & Barocas, *supra* note 7.

people strategically manipulate data about themselves to receive a more favorable prediction?

A. *What Is Gaming?*

Much of the literature on algorithmic scoring assumes that individuals will simply accept the scores that they receive from algorithmic systems.[197] This static vision of the relationship between individuals and algorithmic scoring is an oversimplification. Individuals often manipulate their behavior to receive preferential (and arguably unjustified) treatment from algorithmic systems. System designers are aware of these attempts and engage in their own system of countermoves to increase the resiliency of their system to an individual's gaming strategies. The result of these gaming moves and countermoves is a dynamic choreography where individuals modify their behavior to improve their position and system designers tweak their systems to limit this activity.

Evaluations are a part of life. Often the characteristics that people want to evaluate are unknown, or even unknowable. In these cases, we defer to proxies for these qualities to make determinations. Both human evaluations and algorithmic evaluations rely on proxies. That is, rather than measuring some characteristic directly, evaluators are forced to measure a different characteristic that is related to the key characteristic.

Individuals often modify their behavior to change proxies and, ultimately, influence how they are perceived.[198] For example, a person may choose to evaluate others according to whether they are wealthy or not. In response, a person who wishes to appear wealthy might modify their behavior so that they exhibit proxies for wealth, such as wearing an expensive watch or driving a flashy car. When a person rents a sports car for a night to drive to their high school reunion, they are manipulating a proxy to change how people evaluate a characteristic about them (here, wealth).

Credit determinations are no different. In the case of credit, the key characteristic that lenders purport to want to discover is the likelihood that credit will be repaid. However, there is no way to directly measure the likelihood of future repayment risk. The choice to extend credit has always been based on some proxy for creditworthiness. At times these decisions were based on local knowledge. A person in a small town who came from a "good family," who was not seen at the local bar too often, and exhibited a genial disposition might be extended credit. However, another person with different community standing may be denied credit.

---

197. *See* Jane Bambauer & Tal Zarsky, *The Algorithm Game*, 94 NOTRE DAME L. REV. 1, 3 (2018) ("The policy literature has largely assumed that algorithmic systems dictate a score, and individuals accept the results.").

198. This claim is one frequently made in the machine learning literature, but not without critique. *See* Ignacio N. Cofone & Katherine J. Strandburg, *Strategic Games and Algorithmic Secrecy*, 64 MCGILL L.J. 623, 661 (2019) (arguing that claims about the ease of gameability are largely exaggerated and outlining the conditions for gameability).

The ability of individuals to game scoring systems depends on the gap between proxy variables and the key characteristic. Take FICO credit scoring. In calculating these scores, FICO uses length of credit history as a proxy for likelihood to repay loans. Credit issuers are not interested in the amount of time that a person has had credit. Instead, they are interested in a person's credit history to the extent that it can be used to forecast their future behavior.

Financial advice columns offer an array of tips for improving credit scores.[199] Many of these tips do not make an individual more creditworthy but, instead, simply change their proxy data to make them appear that way. In other words, tips are simply gaming strategies for individuals to raise their credit scores. For instance, some financial outlets recommend that parents add their children as "authorized users" to their credit cards to establish a lengthier credit history for their children.[200] Other articles recommend maintaining a credit balance rather than paying off outstanding debt in full.[201]

Some proxies are more closely tied to creditworthiness, such as loan repayment history. Most credit scores rely on data about a person's previous credit repayment history. Repayment history tracks creditworthiness quite closely, yet previous payment history is still a proxy for creditworthiness. It is an imperfect measure for future behavior and relies on the assumption that a person's past payment history is indicative of future behavior.

Moreover, there could be cases where previous repayment history fails to track creditworthiness. For instance, a financially reckless nephew could have his loans repaid by a rich uncle who has since vowed to never bail him out again. In this case, loan repayment history provides less predictive power about the nephew's likelihood to repay future loans. In other words, it fails to track the usual causal story that previously paying a loan on time is indicative of a financially diligent person who is likely to repay credit in the future.

However, maybe the uncle has issued an empty threat and will bail the nephew out again. If that is the case, then loan repayment history may not tell

---

199. *See, e.g.*, *How Do I Get and Keep a Good Credit Score?*, CONSUMER FIN. PROT. BUREAU (Dec. 12, 2024), https://www.consumerfinance.gov/ask-cfpb/how-do-i-get-and-keep-a-good-credit-score-en-318 [https://perma.cc/4WED-5Q4F].

200. *See* Ben Luthi, *Should You Add Your Child to Your Credit Card as an Authorized User?*, EXPERIAN (May 29, 2022), https://www.experian.com/blogs/ask-experian/should-you-add-child-as-authorized-user-credit-card [https://perma.cc/M563-36HU].

201. This distinction may break down because it may just change the causal story. People with parents who establish a credit history for them are likely to be taught the importance of maintaining good credit scores and will be likely to repay a loan. So, maybe, gaming versus not gaming is not as neat of a distinction as originally thought. Alternatively, if placing furniture pads on furniture makes you a lower credit risk, is it gaming if you place furniture pads on your furniture? Does behaving diligently improperly raise your credit score or does the *practice* of doing these things actually make you more likely to repay the loan? I suggest that gaming turns on whether the act of placing the furniture pad makes you more likely to repay. If it does, then doing so is not gaming. Gaming assumes some objective probability that is improperly manipulated. Maybe future behavior is a difficult case for gaming.

us much about the nephew's financial diligence, but instead offers some modest forecasting about whether the loan will be repaid.

B. *Gaming Strategies, Empirical Reality, and Social Oversight*

The ability of individuals to pursue gaming strategies is largely determined by the type of data that is used to calculate scores. Data must be mutable for individuals to game it. That is, there must be some capacity to change the data without also affecting the attribute that serves as the basis for the score. For example, insurance companies offer discounts for walking ten thousand steps per day. Committing to this behavior is *not* gaming because it will probably make the person healthier. However, attaching a tracker to their dog would be gaming because it alters the reported data without altering the underlying characteristic (human health).[202]

Input data differs in its mutability. Some data is practically immutable. For instance, information about an adult's height is (almost always) immutable because it is unlikely to change over time. Other data has large barriers to mutability. An algorithmic scoring system may consider whether a person has an advanced degree as part of its input data. This feature is quite difficult to change and typically requires several years of consistent effort to change it. Other data is highly mutable. It is both possible and cheap to alter it. For instance, the choice of web browser could be related to a score determination. Yet individuals could easily alter this characteristic by simply downloading and using a different web browser.

Gaming strategies are grounded in strategic rationality. Attempting to game along some input data may not make strategic sense. For example, it would not make sense to spend the time and money associated with earning an advanced degree simply to modestly improve one's credit score. Gaming strategies, then, are most likely to be pursued on variables that are highly and cheaply mutable.

What are the incentives for gaming data such that it breaks causal links? Jane Bambauer and Tal Zarsky identify four basic types of gaming strategies[203]: avoidance,[204] altered conduct,[205] altered input,[206] and obfuscation.[207] Katherine Strandburg and Ignacio Cofone add to this typology by arguing that social

202. *See generally* Bambauer & Zarsky, *supra* note 197.

203. *Id.* at 12–13.

204. *Id.* at 12 ("[A] process by which a person avoids being the subject of an algorithm's model at all."); *see also* FINN BRUNTON & HELEN NISSENBAUM, OBFUSCATION: A USER'S GUIDE FOR PRIVACY AND PROTEST 47 (2015).

205. Bambauer & Zarsky, *supra* note 197, at 12 ("Altered conduct involves changing behavior with the hope that the new behavior will change the proxies (inputs) that a model is using and recording, thus resulting in a changed estimate.").

206. *Id.* ("[A]ltered inputs involve manipulating or falsely reporting an input rather than changing conduct in order to manipulate a correctly reported input.").

207. BRUNTON & NISSENBAUM, *supra* note 204, at 1 ("[T]he deliberate addition of ambiguous, confusing, or misleading information to interfere with surveillance and data collection.").

desirability bias incentivizes gaming of altered conduct.[208] For alternative credit data, Baumbauer and Zarsky identify several potential strategies, including that debtors "might *alter their conduct* by switching their devices off (or leaving them at home) when visiting high-risk locations like casinos or discount and liquor stores," or "[t]hey might *alter the inputs* by affixing their Fitbit (or other wearable) to their hyper dogs to get credit for steps while sitting around watching TV."[209]

There are also examples of altered inputs, particularly false inputs, when savvy users realize the potential for gaming certain fintech platforms' technology. For instance, one high-profile example involved fintech companies that advertised products allowing consumers to quickly borrow money for home-improvement projects. Some companies had deficient identification verification systems and became victims of synthetic identity fraud, where an individual pretended to be someone else and received cash.[210]

Moreover, there is a cottage industry surrounding giving gaming advice to potential borrowers. For example, the *Marketplace Tech* podcast discusses several gaming strategies to boost credit scores, including adding an authorized user to a credit account.[211] LendingTree offers strategies for quickly boosting a credit score in thirty days, including temporarily lowering credit utilization rates by asking for a credit limit increase and adding alternative data like utility payments.[212]

Indeed, several fintech startups are almost explicitly geared toward gaming traditional credit scoring models. To illustrate, "credit builder" cards have become a popular product for helping people build their credit scores. The key feature of these cards is that they do not actually lend any money. Instead, they allow people to deposit money into an account and then pay bills from that account while reporting this activity to traditional credit scorers like FICO or VantageScore.[213]

These examples illustrate the overall problem that gaming poses for legitimacy: They frustrate attempts to provide social oversight of acceptable and unacceptable uses of credit data. The examples provided by Bambauer and Zarsky illustrate how selective reporting of conduct or manipulation of

---

208. Cofone & Strandburg, *supra* note 198, at 662–63.

209. Bambauer & Zarsky, *supra* note 197, at 19–20.

210. Alana Semuels, *How Easy Lending Can Lead to Fraud*, TIME (May 29, 2024, 10:21 AM), https://time.com/6982203/solar-lending-fraud [https://perma.cc/BC7C-QVK9].

211. *Gaming Your Finances to Get the Perfect Credit Score*, MARKETPLACE (July 11, 2022), https://www.marketplace.org/episode/2022/07/11/gaming-your-finances-to-get-the-perfect-credit-score [https://perma.cc/2UGF-LWXY].

212. Kristen Grau, *Can You Raise Your Credit Score by 100 Points in 30 Days?*, LENDINGTREE (Aug. 28, 2024), https://www.lendingtree.com/credit-cards/articles/raise-credit-score-30-days [https://perma.cc/8G2Q-W3RC].

213. Gina Heeb, *Credit Scores Without Debt? Fintech Cards Baffle Credit Industry*, WALL ST. J. (Aug. 28, 2024, 5:30 AM), https://www.wsj.com/finance/credit-score-building-cards-fintech-ea8e61d8 (on file with the *Iowa Law Review*).

data make it impossible to assess whether the variables in questions are socially legitimate.[214] In the case of identity theft, the illegitimacy of committing fraud crowds out an assessment of the underlying product. Even fintech that aims to disrupt traditional scores leaves traditional financial institutions flummoxed and in disagreement about how to proceed. Essentially, in these situations, the public is cut out from evaluating the credit lending activity on its own merits.

C. Policy Proposals to Facilitate Oversight of Gaming

The general problem we have identified with alternative data and creditworthiness is a special case of Goodhart's Law. In simple terms, Goodhart's Law states that once a measure becomes publicly known, it is no longer useful because people will manipulate their behavior to fit the measure rather than the underlying trait the measure was trying to capture.[215]

Regulators can take the implications of Goodhart's Law and outline requirements for testing gameability of algorithmic systems in credit lending systems. This effort means validating measures of creditworthiness not just at one point in time, but across time. Having established that the legitimacy of credit lending hinges on gaming not breaking causal stories between data and credit decisions, we offer several policy proposals for taking this idea seriously.[216]

One policy proposal with a strong foundation in existing law is updating current requirements for data disclosure by credit lenders. The HMDA already has disclosure requirements for mortgage data, and this data is regularly analyzed by the Fed and CFPB.[217] The CFPB has issued inquiries requiring auto lenders to disclose granular data about their lending activities and issues similar inquiries for lenders in other industries.[218] Because these disclosure requirements already exist, they could easily be extended to cover alternative credit data.

For dealing with nonintuitive variables, we argue that an overly rigid focus on ex ante justification of inputs should be revisited. The benefit of machine learning is that it can uncover relationships that are not ex ante intuitive.

---

214. Bambauer & Zarsky, *supra* note 197, at 25–28.

215. Christopher Mattson, Reamer L. Bushardt & Anthony R. Artino, Jr., "*When a Measure Becomes a Target, It Ceases to Be a Good Measure*," J. GRADUATE MED. EDUC. 3 (Feb. 2021), https://meridian.allenpress.com/jgme/article-pdf/13/1/2/3268367/i1949-8357-13-1-2.pdf [https://perma.cc/27HR-GNSC].

216. We focus primarily on alternative data, but adopting this principle would also help reinvigorate the regulation of traditional credit data as well. Although some traditional data likely isn't gameable, some has become so well-known and understood that individuals almost certainly game particular parts of their credit files. Validating the components of credit scores periodically could help resolve some of these problems.

217. *See Mortgage Data (HMDA)*, *supra* note 122. *See generally* 12 U.S.C. § 2803 (outlining HMDA disclosure requirements).

218. Chris Kukla, Richard Landau & Ashwin Vasan, *Our Auto Finance Data Pilot*, CONSUMER FIN. PROT. BUREAU (Feb. 23, 2023), https://www.consumerfinance.gov/about-us/blog/our-auto-finance-data-pilot [https://perma.cc/KCV4-Z6DQ].

More importantly, this ex ante nonintuitiveness may actually be desirable insofar as it makes gaming more difficult. In terms of a policy shift, allowing for ex post explanations about credit determinations could preserve the protection that alternative data provides against gaming, while also satisfying the need to allow for social oversight of the decision-making process to build legitimacy.

Actualizing this in policy could look something like adding regulatory requirements to encourage or force lenders to test for the gameability of their data. Cofone and Strandburg already offer a general theory about how one might assess how incentivized someone would be to game a particular variable.[219] Specifically, they argue that gaming is less of a problem than is commonly assumed. Their argument provides a strong basis for evaluating these problems in an empirical fashion: If gaming is not really occurring at a scale that matters, this suggests that regulation should not be focused on deterring consumers.

Mathematical and statistical approaches have also been explored in the technical literature. For example, Daniel Björkegren and his coauthors suggest explicitly modeling the cost of manipulation (how easy it would be for an individual to change some measured data) within machine learning systems designed at provisioning social services.[220] Moritz Hardt and coauthors propose a class of "sequential learning" algorithms that simulate the gaming problem to produce machine learning classifiers that are robust to manipulation.[221]

Within law, there have been several proposed interventions as well. Cary Coglianese has discussed the problem of gaming in machine learning systems deployed by the federal government.[222] Rebecca Crootof, Margot Kaminski, and Nicholson Price suggest using a "human-in-the-loop" approach to solve the gaming issue, among other ethical issues.[223] One of us has previously suggested validating algorithmic audits with holdout sets subjected to randomization.[224]

Again, such an intervention has a strong basis in existing law. For example, the Truth in Lending Act's Regulation Z requires that lenders perform periodic audits to ensure that they have made a good faith effort to ensure that

---

219. Cofone & Strandburg, *supra* note 198, at 662–63.

220. *See generally* Daniel Björkegren, Joshua E. Blumenstock & Samsun Knight, *Manipulation-Proof Machine Learning* (Ctr. for Effective Glob. Action, Working Paper No. 186, 2021), https://escholarship.org/content/qt0w44v8pb/qt0w44v8pb.pdf [https://perma.cc/SCQ3-M4HT].

221. *See generally* Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou & Mary Wootters, *Strategic Classification*, ARXIV (Nov. 24, 2015), https://arxiv.org/pdf/1506.06980 [https://perma.cc/3E3H-8X3J].

222. CARY COGLIANESE, A FRAMEWORK FOR GOVERNMENTAL USE OF MACHINE LEARNING 43 (2020), https://www.acus.gov/sites/default/files/documents/Coglianese%20ACUS%20Final%20Report%20w%20Cover%20Page.pdf [https://perma.cc/26GQ-QERF].

223. Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 VAND. L. REV. 429, 487–88 (2023). Human-in-the-loop approaches broadly mean that algorithmic decisions should always be overseen or vetted by actual people. *Id.* at 446.

224. *See* Aniket Kesari, *Predicting Cybersecurity Incidents with Machine Learning and Mandatory Disclosure Regulation*, 2022 U. ILL. J.L. TECH. & POL'Y 57, 94–95.

borrowers can repay their debt.[225] ECOA's Regulation B prohibits discrimination on the basis of protected characteristics, and credit lenders are advised to regularly conduct bias and fairness audits to ensure compliance with disparate impact law.[226] Apart from credit lending, New York City passed a law requiring that businesses using automated employment decisions tools submit annual bias and fairness audit reports.[227]

Requiring that lenders conduct internal gameability audits—in addition to existing accuracy and fairness audits—can help address concerns about the legitimacy of alternative credit data. While not a full panacea, as these disclosure regimes have their own criticisms, it is still a powerful tool for creating legal standards when full oversight is impractical.

* * *

This gaming case study ultimately reveals that it is not just the cleverness of consumers or the limits of current models, but conceptual confusion that still surrounds how to regulate alternative credit data. Is the problem deception, causation, relevance, or fairness? By returning to the principle of empirical grounding, regulators can cut through this confusion. Even when intuition breaks down, regulators can still tackle the problem by developing tools that enable social oversight. Audits, stress tests, counterfactual analysis, and adversarial simulations all become part of a coherent regulatory strategy aimed at sustaining the real-world reliability and legitimacy of credit models.

CONCLUSION

While alternative credit data and AI-driven credit lending systems seem to present complex and possibly intractable challenges, ultimately the solution is simple: Focus on empirical reality. Nexus talk has obscured the real normative stakes of the alternative credit data debate. By moving beyond nexus and instead talking about the empirical relationship between data and credit decisions, conceptual focus can be placed on appropriate social oversight. Clarifying these stakes can help empower regulators and also provide clear guidance to industry and consumers as they navigate a world increasingly mediated by algorithms.

225. *See* 12 C.F.R. § 1026.1 (2024); *What Is the Ability-to-Repay Rule?,* CONSUMER FIN. PROT. BUREAU (Apr. 26, 2024, 12:51 PM), https://www.consumerfinance.gov/ask-cfpb/what-is-the-ability-to-repay-rule-en-1787 [https://perma.cc/YS8N-SGBL].

226. *Banking, Checklist - Best Practices for Avoiding AI Bias in Lending Decisions*, BLOOMBERG L., https://www.bloomberglaw.com/external/document/X9690B3G000000/banking-checklist-best-practices-for-avoiding-ai-bias-in-lending (on file with the *Iowa Law Review*).

227. *See Automated Employment Decision Tools (AEDT)*, N.Y.C. CONSUMER & WORKER PROT., https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page [https://perma.cc/88CB-7PTJ].