Northern Ireland
Legal Quarterly
Winter Vol. 72 No. 4 (2021) 771–798
Article DOI: 10.53386/nilq.v72i4.931

Law, information, and contemporary finance in the United States: a sociological perspective

Bruce G Carruthers

Northwestern University, Illinois
Correspondence email: b-carruthers@northwestern.edu

ABSTRACT

The global financial system has become remarkably complex as it combines high transaction volumes with growing speed. Financial transactions depend critically on information to mitigate uncertainty and vulnerability, and such transactions are therefore affected by recent developments in information technology, driven by fintech firms and commonly termed 'big data'. The volume, velocity and variety of information is unprecedented and poses new challenges for governance. Legal rules for data ownership, privacy, security and usage are becoming obsolete and ineffective in the context of algorithmic information processing and decision-making. Yet some types of information remain embedded in financial contracts and regulations, relatively unaffected by these developments. This variation challenges simplistic claims about big data and underscores how some of the historical particularities of the United States have gained global significance.

Keywords: information; 'big data'; financial markets; regulation.

INTRODUCTION

Financial markets depend on information.[1] Over the last three centuries, use of price lists, financial newspapers, carrier pigeons, telegraphs, stock tickers, telephones, computers, Bloomberg terminals, fibre-optic cables, and now cloud computing and the internet all reflect the enormous appetite that financial market participants have for information, via whatever technology is currently available. The reason is straightforward: credit transactions and investment decisions face the twin problems of vulnerability and uncertainty. Vulnerability means that the financial interests of the lender/investor are at the mercy of someone else's future actions, depending on the size and maturity of the loan or investment. A debtor who does not repay harms the lender,

[1] Malcolm Campbell-Verduyn, Marcel Goguen and Tony Porter, 'Finding fault lines in long chains of financial information' (2019) 26 Review of International Political Economy 911, 913.

and the bigger the loan, the worse the harm. Lenders and investors also face uncertainty. They are ignorant of the borrower's future actions, not knowing who will repay, or if an investment will be profitable. Of necessity, they make predictions.

Lenders and investors manage their uncertainty by collecting information about the borrower/investee's willingness and ability to repay, both *ex ante* and *ex post*. Knowledge is the solution to ignorance. Vulnerability is frequently addressed using collateral, which requires that the lender be able to identify and track a specific asset subject to a security interest. Here registration and traceability are critical, and both depend on a tracking system or registration infrastructure. Vulnerability can also be mitigated via diversification into multiple loans or investments, which requires the investor to identify and pursue independent financial alternatives. Additionally, it can be addressed by *ex post* constraints imposed on borrower behaviour.

Hardware and software developments in information technology are now reshaping how participants in financial markets address problems of vulnerability and uncertainty. The term 'big data' signals a substantial increase in the volume, velocity and variety of information, and 'fintech' firms wed such data to financial decision-making.[2] Advances in information technology make it possible to manage and process ever more data. Many of these developments are captured by Shoshana Zuboff's idea of 'surveillance capitalism', although there is less discontinuity with the past than her dramatic formulation suggests.[3] This: 'new form of information capitalism aims to predict and modify human behaviour as a means to produce revenue and market control'.[4] It utilises the exceptional amounts of information that are now available about the actions of billions of individuals, harvesting, aggregating, analysing, and monetising digital traces of online activity, and operates: 'through unprecedented asymmetries in knowledge and the power that accrues to knowledge'.[5] The goal is to predict and influence human activity, in pursuit of profit. Social media offers a vivid example of what Zuboff has in mind: Facebook tracks in detail the behaviour of billions of users and earns billions of dollars in profits.

With their historical dependence on information, financial markets provide a good opportunity to assess Zuboff's claims against a broader

[2] Federal Trade Commission, *Big Data: A Tool for Inclusion or Exclusion?* (Washington DC 2016) 1.

[3] Shoshana Zuboff, *The Age of Surveillance Capitalism* (Public Affairs 2019).

[4] Shoshana Zuboff, 'Big other: surveillance capitalism and the prospects of an information civilization' (2015) 30 Journal of Information Technology 75.

[5] Zuboff (n 3 above) 11.

backdrop.[6] In the United States (US), creditors and investors have always sought to predict and influence what debtors will do, so these aspirations are not new. Neither is the concern for earnings. Nor is the appetite for information, for some for-profit companies have long specialised in the production of large volumes of information about the creditworthiness of both firms and individuals. As detailed below, for borrowers that issue debt securities, US bond rating agencies have been arbiters of creditworthiness, gathering information and rating bonds since the early twentieth century. Their rating systems are now legible around the world. Agencies like S&P tout the predictive value of their ratings by documenting the association between default rates and ratings (higher ratings mean lower default rates). Information also drives the allocation of trade credit among small firms, and in the US they have had their financial status assessed since the mid-nineteenth century by Dun and Bradstreet and its predecessors. And individual consumers have for many decades been tracked by rating agencies like TransUnion, Experian and Equifax, which in the US maintain credit files on hundreds of millions of persons and calculate FICO scores to measure their creditworthiness. All this information guides investors and lenders in their financial decisions: where to invest? To whom to lend? How to make more profitable predictions about the future? It may not exactly be the surveillance that concerns Zuboff, but the creation and use of information set precedents and posed problems that can put ‘surveillance capitalism’ in proper perspective.

Financially relevant information comes from many sources, both public and private. Some information is provided as a public good: the US Commerce Department publishes an enormous amount of economic data, and each release is closely monitored by financial market participants. Almost every country generates national income statistics that describe the state of their economy, its size, growth, and trade with the rest of the world. Other information is privately created and distributed. Financial exchanges provide price and transaction data, while maintaining strict ownership over such information as a type of intellectual property. Likewise, for-profit bond rating agencies evaluate fixed income securities and distribute their ratings as a key piece of credit information. Some private information is publicly mandated, and so has a hybrid provenance. For example, publicly traded US corporations regularly disclose standardised and independently audited financial information about their performance. They are required to do so as a matter of federal securities law.

It is not only private lenders and investors who demand financial information to reduce uncertainty. Some information has been used in

[6] See also James W Cortada, ‘A history of information in the United States since 1870’ (2017) 52 Information and Culture 64.

the legal and regulatory apparatus that governs financial markets. Early in the twentieth century, US judges used credit ratings as indicators of business practice standards in order to gauge 'prudent' decisions.[7] When the Great Depression hit, bond ratings were employed by federal bank examiners as a way to bolster banks by changing how their bond portfolios were valued.[8] Instead of applying a simple mark-to-market rule that used current market prices, bonds rated 'above investment grade' could be valued at their historical cost, ie at their original purchase price. Since many bonds lost value during the Depression, this new rule, institutionalised by the Comptroller of the Currency in 1931, allowed banks to inflate the value of their bonds and strengthen their balance sheets, with the knowing cooperation of bank examiners. Other state and federal regulators followed suit and incorporated private bond ratings into their own prudential regulations, preventing pension funds and insurance companies from investing in assets that were too risky.

Lenders and investors have long sought information as a way to solve the problems of uncertainty and vulnerability. Financial markets had a tremendous interest in information and obtained it from many different sources. Recent changes in information technology offered new opportunities to generate and manage information, but does this mark a qualitatively new stage of capitalism? I argue that claims about the novelty of 'surveillance capitalism' are overstated, and that large amounts of information have already been generated and utilised in the past, particularly in finance. Furthermore, previous attempts to regulate information, but also to use it in regulation, give some purchase about how public policy might respond to the challenges posed by big data. Recognition of how much information has increased must be put into a broader context that also considers the quality of information, its form, content, structure, distribution, and usage. My approach draws on organisational sociology and work in the sociology of quantification to offer a richer appreciation of information and its role in finance, and I address a number of aspects: who is the subject of information? Who creates it? Who uses the information?

Terms like 'data', 'information', 'knowledge', 'variables', and 'measures' have been applied loosely in discussions of big data. For simplicity, I will use the term 'information' to denote some kind of

[7] Marc Flandreau and Joanna Kinga Sławatyniec, 'Understanding rating addiction: US courts and the origins of rating agencies' regulatory license (1900–1940)' (2013) 20 Financial History Review 237.

[8] Bruce G Carruthers, 'Financial decommodification: risk and the politics of valuation in US Banks' in Edward J Balleisen et al (eds), *Policy Shocks: Recalibrating Risk and Regulation after Oil Spills, Nuclear Accidents, and Financial Crashes* (Cambridge University Press 2017).

symbolic representation of an object, person, outcome, or process. I will particularly focus on electronic information, but obviously it can assume multiple forms. 'Data' refers to a corpus of information, typically but not necessarily organised as a set of variables that measure or reflect specific features of objects, texts, persons, outcomes or processes. Depending on the level of measurement, variables can be nominal (categorical), ordinal (ordered categories), or cardinal (with some kind of numerical value). And variables can be combined and processed to create new variables, or analysed to unearth relations between them (eg correlations). Such manipulations enable raw data to be turned into counts, ratios, indices, and other higher-order information that can serve interpretive and predictive purposes.

BIG DATA AND FINANCE

As more social and economic activity moves online, people are encouraged to create rich, granular data streams that are being continuously harvested, analysed, shared, and monetised. It is no longer that offline behaviour is recorded using separate measuring instruments, but rather that digital behaviour creates traces that are continuously and automatically stored. For example, even before someone makes a purchase, a consumer's interest in a particular commodity can be closely tracked via their online search behaviour, by the websites they visit, by their 'clicks' and 'likes', and the amount of time they spend viewing particular screens using their mobile and desktop devices. The software that enables people to move between websites also automatically records their movement. Firms now have a variety of techniques for tracking users, including 'cookies', device fingerprinting, and 'history sniffing', and they can even link users' behaviour across multiple devices, including smartphones, desktop computers, tablets, laptop computers, and other inanimate objects.[9] Additionally, their personal interests and activities can be linked to those of their friends and family, to whom they are connected via social media platforms, and to their own past behaviour. Electronic payment systems allow the platform to gather continuous data on a consumer's financial activity and status, and predictive models of consumer attention, engagement, and purchasing enable platform firms to sell advertising and otherwise steer the consumers to preferred websites and sponsored options. Online experiments with full randomisation and huge sample sizes allow the host website to determine which 'nudges' and 'decision architecture' work best in moving users toward a preferred alternative. Software is updated frequently, and 'black box' models

[9] Federal Trade Commission (n 2 above).

exploit the latest in machine learning (ML) algorithms and artificial intelligence (AI) to develop statistical models that can involve millions of variables. Such complexity makes them virtually incomprehensible to ordinary humans. In the offline world, mobile phones now function like tracking devices, allowing apps to follow people wherever they go. Proximity to other mobile phone users can be confirmed if anyone posts a selfie online, as facial recognition software can identify each person in the image. Such data is collected by monopolistic 'big tech' firms, and only recently have authorities in Europe and the US begun to activate competition policy or impose privacy rules in response. Yet information seldom remains solely in the possession of the original repository. Instead, an ecosystem of data aggregators and brokers pull together information from multiple sources and sell it to multiple users, almost always without the knowledge of the data subjects. As these activities flourish, it is clear that they have gotten well ahead of public understanding and policy.

Many of the new opportunities to exploit information are being pursued by 'fintech' firms. These 'disrupters' of the financial system emerged from the tech industry, exemplified by Silicon Valley software, social media and platform firms.[10] Their core expertise is in information management rather than finance. Although small in size, they are relatively unconstrained by limited organisational capacity and as non-depository institutions they escape many of the strictures that regulators impose on banks.[11] Thanks to the development of cloud computing, fintech firms have less need to maintain their own information hardware. Instead, they can turn to vendors and easily scale up or down as needed.[12] New lending platforms match debtors with creditors, and the platform itself harvests information as well as fees.[13] Whether fintech will destabilise finance remains to be seen, but in one familiar scenario market incumbents acquire the startups and absorb them, or form alliances with them, before any major disruption occurs. A number of the largest banks have already acquired expertise in cybercurrencies not because they believe bitcoin to be the money of the future, or because they have embraced crypto-anarchist philosophies, but because the underlying distributed ledger technology offers useful

[10] Xavier Vives, 'Digital disruption in banking' (2019) 11 Annual Review of Financial Economics 243.

[11] William Magnuson, 'Financial regulation in the bitcoin era' (2018) 23 Stanford Journal of Law, Business and Finance 162.

[12] Franklin Allen, Xian Gu and Julapa Jagtiani, 'A survey of fintech research and policy discussion' (2020) WP 20-21 Federal Reserve Bank of Philadelphia Working Paper Series 35.

[13] Chris Clarke, 'Platform lending and the politics of financial infrastructure' (2019) 26 Review of International Political Economy 863.

capabilities that they wish to possess. Of course, disruption may occur if the big tech firms that currently manage online searches, purchases, or social activity decide to provide financial services to their large user-bases. If they do, these firms (eg Facebook, Google, Amazon, etc) are large enough to challenge even the biggest banks. Facebook, for example, recently announced its intention to develop its own currency, initially called 'libra' and now named 'diem'.

The significance of information partly stems from its incompleteness and uneven distribution. No one knows with certainty what the future will bring or what consequences may follow from a particular course of action. The severity of cognitive limits has been recognised in behavioural economics and its antecedents in organisational sociology by notions of 'bounded rationality' and 'uncertainty absorption'.[14] Such limits were stressed by Austrian School economists to argue for the superiority of decentralised markets over centralised planned economies.[15] Their unevenness (termed 'asymmetries of information') prompted the new economics of information, building on models of 'markets for lemons' and illustrating the significance of systematic differences in regard to who knows what.[16] And the particularities of how information is organised have been well appreciated by sociologists of quantification, accounting, and categories,[17] particularly when that information is inscribed in organisational practices. Within financial economics, however, the still highly influential 'efficient markets' approach proposes that, in efficient markets, prices fully reflect all available information.[18] According to this approach, market prices provide the single best summary of all that is known by everyone.

Information has often been treated as something of which there is simply more or less. With more information, a decision-maker

[14] James G March and Herbert A Simon, *Organizations* (John Wiley & Sons 1958); Herbert A Simon, *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations* 4th edn (Free Press 1997); Daniel Kahneman, 'Maps of bounded rationality: psychology for behavioral economics' (2003) 93 American Economic Review 1449.

[15] F A Hayek, *Individualism and Economic Order* (University of Chicago Press 1948).

[16] Joseph E Stiglitz, 'The contributions of the economics of information to twentieth century economics' (2000) 115 Quarterly Journal of Economics 1441.

[17] Michael Power, *The Audit Society: Rituals of Verification* (Oxford University Press 1997); Ezra W Zuckerman, 'The categorical imperative: securities analysts and the illegitimacy discount' (1999) 104(5) American Journal of Sociology 1398; Wendy N Espeland and Mitchel Stevens, 'A sociology of quantification' (2008) 49 Archives of European Sociology 401, Marion Fourcade and Kieran Healy, 'Classification situations: life-chances in the neoliberal era' (2013) 38 Accounting, Organizations and Society 559.

[18] Eugene F Fama, 'Efficient capital markets II' (1991) 46 Journal of Finance 1575.

knows more and can make better decisions. Arguments about big data or surveillance capitalism rest upon the claim that the volume of information has vastly increased. Information asymmetries mean that some parties to a transaction know more than others; in the canonical example, the seller of a used car knows more than the buyer about the car's true underlying condition, and whether or not it is a 'lemon'. While it is useful to discuss the total amount of information in a market, or to determine who knows more than others, it is important to recognise qualitative differences as well. Some of these differences matter a great deal for the role that information plays in financial markets, and for how that role changes.

Here I apply a simple framework and consider some ways in which information varies. The first issue concerns the subjects of information, ie what or who is it about? Concerns about surveillance capitalism are clearly driven by the expansion of information about individuals and their activities. But information can also be about financial prices (eg how much does it cost to buy 100 shares of IBM common stock?) and quantities (eg how many IBM shares were traded yesterday on the New York Stock Exchange?). It can concern distinctive qualities, as classifications pervade financial markets.[19] For example, bond ratings place debt securities into a set of discrete ordered categories that measure credit risk. Other economic categories, like industry, are unordered. Analyst recommendations turn on how a company is classified: at first Amazon, for example, was categorised by some analysts as being in the book industry, and by others as belonging in the tech industry. The valuations of Amazon varied enormously depending on this disputed classification (which determined the benchmarks to which it was compared). A second set of issues concern those who create and process information: what is the extent of their ownership and control over information? What are the rights and responsibilities of those who possess information? Do intellectual property rules or privacy standards apply? Who has duties in regard to cybercrimes, or money laundering? Is consent required from data subjects, and what documents such consent? How freely can holders distribute the information they possess about others? Must holders ensure that information is accurate, and how are such obligations enforced? A third set of issues concern the pragmatics of information: how is it used, in what situations, and by whom? Is information utilised to predict behaviour? Does it guide private financial decision-making? Does information inform public regulations? Can it trigger private or public rules? And who constitutes the primary audience for information? Is it directly actionable or does it just provide background and context?

[19] More generally, categories undergird the process of uncertainty absorption.

Currently, several laws govern the production and use of information in US financial and credit markets to ensure orderly markets or to protect the public interest. Depending on the subjects of information, the creators, users or holders of information may be encumbered with obligations and responsibilities. Innovation has often challenged and circumvented these regulations. In fact, some innovations are intended to evade regulation, and regulators necessarily play a game of catchup. New products and processes may not meet strict regulatory definitions and so are not subject to oversight. Social media platforms, for example, fit badly into older industrial categories, as an anomalous combination of software company, media outlet, and telecommunications firm. Or innovators exploit regulatory loopholes to avoid compliance. Consider, for example, the emergence of a 'shadow banking system' which performs many banking functions but which is not subject to banking regulation because the entities that comprise it do not meet the official definition of 'bank'. Or recall the over-the-counter (OTC) financial derivatives market, which grew dramatically during the 1990s and 2000s as public regulators left it alone.[20] Today, fintech firms sit uneasily in the regulatory system because their activities are both hybrid and innovative.

New uses of information challenge existing standards, in part because the type of information that is now available, as well as its breadth and detail, were simply inconceivable when laws like the Fair Credit Reporting Act of 1970 (FCRA), for example, were passed. Relevant information now increasingly accumulates outside of credit reports and traditional credit rating agencies. Similarly, data security breaches now occur on a scale that was impossible when information was stored in paper files. A lone hacker can, in a single breach, abscond with confidential information about hundreds of millions of individuals, and sell it on the dark web. In 2017, Equifax, one of the biggest consumer credit rating agencies, experienced a security breach and confidential information on more than 145 million individuals was accessed and extracted.[21] Similarly, privacy standards confront unprecedented challenges, like the capacity to track people continuously in time and space, or to record proximity so that trackers can know who was with whom and when, or to document all that a person does with their mobile apps. Even ordinary household objects can become

[20] Bruce G Carruthers, 'Diverging derivatives: law, governance and modern financial markets' (2013) 41 Journal of Comparative Economics 386.

[21] General Accounting Office, *Consumer Data Protection: Actions Needed to Strengthen Oversight of Consumer Reporting Agencies, GAO-19-196* (General Accounting Office 2019) 1, 8.

monitoring devices, thanks to the 'internet of things'.[22] The very largest repositories of personal data are now owned and controlled by big tech firms like Google, Amazon, and Facebook, without prior public debate or policy consideration. And even as corporations exploit new information technologies, they have had to invest in cybersecurity or risk theft of their own confidential information. In 2013 the US store Target was breached and information on 41 million consumers stolen, a catastrophe for Target and a sobering example for other mass retailers.

BIG DATA IN THE PAST

Although current developments in big data are heralded as if they were without precedent, the role of private firms in the accumulation, analysis and commodification of large volumes of financial information began in the nineteenth century, using traditional paper-and-ink information technology. Starting in the 1840s, for-profit 'mercantile agencies' in the US began systematically to gather information about firms nationwide and to provide credit reports, and later credit ratings, to their clients.[23] Reports and ratings were then used by clients to reduce their own uncertainties and vulnerabilities in making credit decisions. Typically, wholesalers extended short-term, unsecured trade credit to their customers, who would be supplied with goods and then paid after a conventional period of time like 90 or 180 days, or even settling their accounts once a year. After they received goods but before they paid, customers were indebted to their suppliers. Suppliers faced a difficult situation: it was hard to sell to customers without providing credit, but extending credit to the wrong customer risked losing money. So mercantile agency clients were especially keen to know who was genuinely creditworthy. Local customers were part of the supplier's own community, and so local social networks made it easy to determine someone's reputation or ascertain past behaviour. But as commerce expanded, firms increasingly dealt with customers from other parts of the country, and so traditional reputation-based methods to assess creditworthiness did not work. Social networks grew threadbare at greater distance, and so suppliers sought alternative forms of information.

The Mercantile Agency was founded in 1841 by a businessman whose own failure during the 1837 crisis underscored the importance of credit.

[22] Dan Feldman and Eldar Haber, 'Measuring and protecting privacy in the always-on era' (2020) 35 Berkeley Technology Law Journal 197.

[23] Barry Cohen and Bruce G Carruthers, 'The risk of rating: negotiating trust and responsibility in 19th century credit information' (2014) 1-93 Sociétés contemporaines 39.

Lewis Tappan created an organisation based in New York City, initially serving New York wholesalers, and establishing a national network of confidential informants to provide information about firms around the country. Mostly, the informants were local attorneys, who typically knew a great deal about business dealings in their own community. In exchange for referrals for collection work, informants reported on local business and responded to queries, mailing information to the Agency's headquarters in New York. Their letters were transcribed into the Agency's proprietary ledgers and then destroyed to maintain the confidentiality of the source. Informants would discuss the state of a business, estimate its net worth, summarise the proprietor's reputation and history of dealings, and provide periodic updates. Out of this accumulation of largely qualitative and impressionistic information, the Agency produced reports and provided them to clients interested in a particular firm's creditworthiness. After the 1850s, the Agency and its competitors started to publish bound volumes containing summary ratings of tens of thousands of firms. Agency subscribers would regularly receive an updated version of the 'manual', containing an alphabetical listing of firms from a particular city or region, a brief statement of their line of business (eg saloon, tailor, dry goods), and then a rating that classified the firm into a discrete ordinal category. The categories looked like modern bond ratings, with some version of 'AAA' denoting the highest level of creditworthiness. Clients could then consult the manual, look up a firm to learn its rating, and judge the risk of extending credit. And the rating system made it easy to make quick decisions, in part because the seemingly precise ratings overlooked the complexities, ambiguities and equivocations contained in ledger information.[24]

Both the Mercantile Agency, later known as RG Dun, and its chief rival, Bradstreet's, expanded over the nineteenth century.[25] Dun augmented its informants with a growing number of branch offices, located both domestically and abroad, so it could use employees to gather information. The total amount of information also grew, and by the end of the century Dun manuals provided ratings on more than a million firms in every part of the US. The manuals were issued annually at first, then twice a year, and then on a quarterly basis. They provided useful information about firms that lacked a national profile, and which were typically of small or medium size. Consultation of rating manuals became part of standard business practice, and even financial organisations specialising in the provision of credit, ie banks,

[24] This pattern is typical of 'uncertainty absorption'.

[25] Bruce G Carruthers, 'From uncertainty toward risk: the case of credit ratings' (2013) 11 Socio-Economic Review 525.

became subscribers. Businesses managed their uncertainties by using information provided by rating agencies.

The rating agencies profited from the subscription fees they charged their customers, so in effect they adopted a 'user pays' business model. More subscribers meant higher revenues. And in order to maintain subscriptions, the agencies vigilantly protected their intellectual property. The manuals were not to be copied or shared with anyone except the authorised subscriber. Yet even as agencies maintained ownership of information, their ability to control its use and diffusion was uneven. Plagiarism and unauthorised replication, for example, was an ongoing concern, and the agencies struggled to prevent out-of-date manuals from circulating privately. The agency continually updated its contracts with customers to discourage them from sharing information with others, but the problem was not easily solved: valuable information was hard to manage.

As credit ratings gained importance, as more people used them to make credit decisions, and as the coverage of firms became more extensive, the agencies came under fire from two groups while the legal aspects of credit information were worked out. Agencies faced lawsuits from firms given a low rating: such firms sometimes sued claiming that the rating was mistakenly low, and that because others had withheld credit, the low rating had harmed the plaintiff. In effect, those bringing suit argued that low ratings acted like self-fulfilling prophecies. Agencies were also sometimes sued when they gave a high rating. If a client used the rating and granted credit to a firm that subsequently defaulted, then they might sue on the grounds that the rating was mistakenly high, and that the agency misled the lender and caused them to lose money. For some decades, the agencies faced a worrisome amount of litigation in state courts, but eventually the law settled on the idea that ratings, as information, were akin to opinions. They could be neither true nor false and were constitutionally protected as a type of free speech. Nevertheless, legal worries meant that the agencies carefully promoted their informational products on the grounds that they were generally useful, but not that they were literally true.

The ratings model was successfully transplanted to a very different credit situation in the early twentieth century. Mercantile agencies continued to provide information about small and medium-sized businesses (and eventually Dun and Bradstreet's merged in the 1930s). But their methods and model were copied starting in 1909 when John Moody began to rate railroad bonds. Bond issuers were among the largest and most capital-intensive firms, and they required long-term capital. Moody and his competitors classified railroad bonds (and later utilities, corporates and sovereigns) into an ordered category system in order to measure the riskiness of a bond. 'AAA' was the highest rating, and connoted the lowest credit risk. Thanks to the mercantile agencies this format for credit information was already very familiar to the business community, although the application was new. Moody charged subscribers for his ratings, published in a manual issued annually, and so adopted the same 'user pays' business model. And the bond rating agencies became as important for long-term debt as the mercantile agencies were for short-term finance: virtually any entity issuing bonds would get rated by Moody's, Standard & Poor's, or Fitch. An unrated bond issue had a hard time attracting buyers, and a 'ratings downgrade' was much to be feared. Bond rating agencies also protected their ratings as intellectual property by putting constraints on their customers, but found the situation increasingly untenable after the invention of the photocopier. Once this device spread in the late 1960s, it became too easy for people to make unauthorised copies of the volumes that Moody's and its competitors published. Quickly, all the rating agencies shifted to an 'issuer pays' business model. Henceforth, an entity wishing to borrow by selling bonds paid the rating agency to rate the bonds. Despite these changes, ratings continued to be used by investors to gauge credit risk, just as they were used by financial regulators. Ratings also found new service when they were incorporated into private regulations, like the standardised contractual language created by ISDA (the International Swaps and Derivatives Association) for the OTC financial derivatives market.[26] Ratings were routinely used by derivatives market participants to calibrate and mitigate credit risk.

By the mid-twentieth century, ratings had become a ubiquitous type of information applied to short-term credit for small firms, long-term lending for large firms, and playing a key role in market governance. But matters did not stop there. Starting in the 1950s, national credit rating agencies compiled credit records for individual consumers, often expanding or merging local agencies that had previously serviced retail merchants in particular cities.[27] After the development of FICO scores in the 1960s, credit card companies, retail businesses, department stores, banks and other lenders could consult a single summary measure of an individual's creditworthiness and use it to decide whether or not to make a loan or offer credit, and at what price. FICO scores governed access to consumer credit, but eventually scores were applied in other contexts as well. Insurance companies, landlords, and employers have

[26] Jon Gregory, *Counterparty Credit Risk: The New Challenge for Global Financial Markets* (Wiley 2010); Joanne P Braithwaite, 'Standard form contracts as transnational law: evidence from the derivatives market' (2012) 75 Modern Law Review 779.

[27] Josh Lauer, 'Plastic surveillance: payment cards and the history of transactional data, 1888 to present' (2020) 7 Big Data and Society 1.

used them to gauge potential customers, tenants, and employees.[28] Mortgage lenders and home equity lenders use FICO scores. Credit scores now regulate much more than just consumer credit, and are readily incorporated into algorithmic decision procedures.

Some recent developments were foreshadowed by the invention of automated underwriting for home mortgages in the 1990s. Given the highly institutionalised market for home mortgages in the US, and given the standardisation both of the underwriting process and mortgages themselves, it was easy to use desktop computer technology to automate some aspects of the process.[29] Mortgage underwriting had traditionally been a complicated labour-intensive process involving many documents and much opportunity for discretionary decision-making and potential bias by loan office personnel. One of the virtues attributed to automated underwriting was its ability to curtail discrimination and ensure that all loan applicants were treated equally, and its development was led by Freddie Mac (the Federal Home Loan Mortgage Corporation).[30] It was also supposed to help speed up the approval process, cut costs, and 'democratise' access to credit. Since federal government agencies had set standards in home mortgage lending since the 1930s, it was unsurprising that a government-sponsored entity like Freddie Mac would take the lead in standardising mortgage underwriting through its Loan Prospector software product. But aspirations of more equitable and democratic credit were not realised.

The upshot was that towards the end of the twentieth century, decades before the current era of big data and surveillance capitalism, remarkable amounts of information about individuals, small and medium-sized businesses, large businesses, and any entity seeking to borrow by issuing bonds, was gathered and distributed on a global scale, and used to address the uncertainty that afflicted financial decision-making. Users sought this information to make predictions about the subjects of information, and the goal was to ensure that debtors met their obligations.

The legal status of these analytical activities was fairly well settled: bond ratings were like opinions, and their veracity entailed little legal risk for those who issued them. Instead, bond ratings supposedly faced a market test: there would be demand for them so long as they were sufficiently useful. And some of that demand derived from the official role bond ratings played in national and state-level prudential

[28] Akos Rona-Tas, 'The off-label use of consumer credit ratings' (2017) 42 Historical Social Research 52.

[29] John W Straka, 'A shift in the mortgage landscape: the 1990s move to automated credit evaluations' (2000) 11 Journal of Housing Research 207.

[30] Ibid 208.

regulations (particularly the distinction between 'investment grade' and 'below investment grade'). The bond rating agencies possessed a unique status as NRSROs (nationally recognised statistical rating organisations), bestowed by the Securities and Exchange Commission, although there was little content to this status, nor even a set of well-defined criteria for how to achieve it. After 2008 the bond rating agencies were widely criticised for failing in their role as evaluators of structured financial instruments like mortgage-backed-securities and collateralised debt obligations, which prompted changes passed in the Dodd-Frank Act of 2010.[31] But they have maintained their central role as creators of key information.

For credit information about individuals, legal rules were put in place to govern how such information could function. Explicit usage of some information, like the race of the borrower, was forbidden. The Fair Housing Act of 1968 outlawed discrimination in home mortgage lending and prohibited 'disparate treatment' in real estate credit transactions. Consumer credit was regulated by Equal Credit Opportunity Act of 1974, which also prohibited discrimination on the basis of protected classes including race, religion, sex, and marital status. Thus, although certain kinds of information about consumers could readily be collected (eg their race, sex, age, etc), conditioning the extension of credit on that information was problematic. And consumer credit ratings were covered by the FCRA of 1970, which set standards for how information could be collected, stored, consulted and presented in a consumer credit report. Among other things, FCRA stipulated that credit information could only be shared with designated parties who had a legitimate business reason to obtain it. At first, however, rating agencies had no obligation to share their information with subjects, nor to ensure that the information was accurate.

Ample research shows that the Fair Housing Act and the Equal Credit Opportunity Act failed to erase discrimination in credit markets and in practice lenders continued to use information about race and gender.[32] Additional laws were passed, like the Home Mortgage Disclosure Act of 1975 and the Community Reinvestment Act of 1977, but these did not solve the problem either. Unequal access persisted, partly because many people did not have a substantial credit record, or because they

[31] On the inefficacy of these regulatory changes, see Giulia Mennillo and Timothy J Sinclair, 'A hard nut to crack: regulatory failure shows how rating really works' (2019) 23 Competition and Change 266.

[32] Devah Pager and Hana Shepherd, 'The sociology of discrimination: racial discrimination in employment, housing, credit, and consumer markets' (2008) 34 Annual Review of Sociology 181; Chloe N Thurston, *At the Boundaries of Homeownership: Credit, Discrimination, and the American State* (Cambridge University Press 2018).

lived in areas that were 'under-banked'. FCRA was updated to ensure that consumers could know the content of their own credit reports and would have some recourse when that content was inaccurate. In general, the legal arrangements governing the generation and deployment of credit information at the end of the twentieth century were imperfect at best, even as such information remained critical to the operation of credit markets. But now, a dramatically new set of circumstances threatens to destabilise these arrangements even more.

BIG DATA TODAY

Online commerce, internet usage, widespread adoption of mobile phones, and social media saturation have now created vast depositories of detailed information about billions of individuals, largely in the hands of a small number of very large tech firms. How to exploit big data for financial purposes is an ongoing project, but the scope and frequency of measurement now goes far beyond what might have been envisioned in a 1980s-era credit report. Nevertheless, the fundamental puzzle remains: how to predict the creditworthiness of persons and firms? How to resolve uncertainties faced by lenders and investors? And the generic answer to this puzzle also remains: gather more information.

What has changed is the volume, variety and velocity of information (number of variables, number of cases, frequency of measurement), which now exceed human comprehension. Thus, data analysis is increasingly done via ML and AI, and substantial amounts of modelling, interpretation and simplification are necessary before putting information before human eyes.[33] ML and AI algorithms search for patterns and optimise pre-specified outcomes, and users hope that information newly analysed is predictive of outcomes that interest them: will a borrower default on a loan? What is the likelihood of repayment? How profitable might a particular customer be to the lender? Such questions are being answered using non-traditional information, ie data that does not come from an ordinary credit report. Berg et al find that several readily accessible 'digital footprints' significantly augment traditional credit variables in predicting the likelihood of default.[34] These include features like the subject's computer operating system (Apple's iOS is better) and hardware (desktop computer, laptop, tablet or mobile phone), keystroke errors, and whether the subject's personal name is contained in their email address. Jagtiani and Lemieux list

[33] Witness the growing importance of 'data visualization' techniques.

[34] Tobias Berg et al, 'On the rise of FinTechs: credit scoring using digital footprints' (2020) 33 Review of Financial Studies 2845.

a number of features that fintech firms exploit, but which are not in credit reports, including bill payment histories, medical and insurance claims, education, social networks, and so on.[35] But there are millions more variables, waiting to be 'mined' for their value in predicting outcomes. The spectrum of accessible data and what it measures is now much wider than before. And the statistical associations that large-scale data analysis uncovers may not be intuitively obvious even if they are predictive (who knew that use of iOS made a difference?). Unexpected associations may mean that a previously overlooked relationship has been unearthed, but it can also reflect spurious correlation.

As the statistical model-building becomes elaborate and automated, the models become increasingly opaque, and even incomprehensible, to humans. Complex models involving millions of variables can deliver better results, but when these are used to inform credit decisions, it is difficult to explain the outcome in any straightforward manner, and such inexplicability presents a regulatory challenge.[36] Opaque algorithms readily baffle and stymie human subjects, and even experts may not understand their own digital tools.[37] When a loan officer consults an applicant's traditional credit file, finds a history of defaults and then denies the loan, it is easy to justify the adverse decision: the individual's payment record had too many blemishes, and those were visible both to the loan officer and the applicant. Furthermore, how the applicant might improve their chances of receiving a loan is also obvious: avoid blemishes. But if a complicated algorithm determines that an applicant is too great a risk and should be denied a loan, there is no easy way to explain how the decision was reached. The analytical process is a black box, and those whose applications were denied may well be unsatisfied with a rationale that amounts to saying: the computer decided. This is particularly problematic since compliance with FRCA requires firms to provide consumers with an 'adverse action' notice if consumer report

[35] Julapa Jagtiani and Catharine Lemieux, 'The roles of alternative data and machine learning in fintech lending: evidence from the LendingClub consumer platform' (2018) Federal Reserve Bank of Philadelphia Working Paper Series WP 18-15.

[36] Bryan Casey, Ashkon Farhangi and Roland Vogl, 'Rethinking explainable machines: the GDPR's 'right to explanation' debate and the rise of algorithmic audits in enterprise' (2019) 34 Berkeley Technology Law Journal 143; Talia B Gillis and Jann L Spiess, 'Big data and discrimination' (2019) 86 University of Chicago Law Review 459, 474.

[37] Hatim Rahman, 'The invisible cage: workers' reactivity to opaque algorithmic evaluations' (2021) Administrative Science Quarterly 3, 8; Callen Anthony, 'When knowledge work and analytical technologies collide: the practices and consequences of black boxing algorithmic technologies' (2021) 66 Administrative Science Quarterly 1173.

information is used to deny access to credit, employment, insurance or some other service.[38]

Similar problems arise in other settings. For example, the financial services industry is turning to 'chat bots' as a way to offer cheap large-scale advice to clients.[39] Instead of a face-to-face meeting with a human financial advisor, the client interacts with a natural language processing algorithm that dispenses financial advice. From the standpoint of the service provider, this is much cheaper and readily scalable. And an algorithm presumably has the advantage of being even-handed and bearing no animus. Yet, if a client seeks to know why they have received a particular piece of automated advice, or why some alternative investment strategy was not recommended, it may be difficult to offer a meaningful explanation. The operation of an opaque algorithm is not easy to explain.

Big data creates big problems. The last several years have witnessed multiple instances where repositories of information have been breached and their contents stolen. The stakes of cybersecurity are high indeed when hackers can seize highly personal information about hundreds of millions of individuals in a single hack.[40] Among other things, these kinds of breaches increase the possibility of large-scale identity theft even as compliance with anti-money-laundering (AML) and know-your-customer (KYC) laws require financial institutions to determine the identity of their clients. The portability of electronic information is one of its great virtues, but this feature also facilitates its theft. Furthermore, critical information systems can fail through a denial-of-service (DoS) attack, or be held hostage to ransomware. Legitimate users can be 'phished', so that they inadvertently provide confidential information like passwords or other identifiers. And some data subjects wish 'to be forgotten', ie to have their past digital footprints completely erased. That way, evidence of wayward youthful behaviour or inappropriate expressions, richly documented on Instagram and Twitter, will not haunt an individual for the rest of their life. Generally, the legal responsibilities of electronic information holders have not caught up with current realities, and this remains an area of consequential flux. Firms concerned about potential liabilities can now obtain 'cyber insurance', but not all relevant cyber-risks are

[38] Federal Trade Commission (n 2 above) 14.

[39] Tom Baker and Benedict Dellaert, 'Regulating robo advice across the financial services industry' (2018) 103 Iowa Law Review 713; Allen et al (n 12 above) 29.

[40] David Maimon and Eric R Louderback, 'Cyber-dependent crimes: an interdisciplinary review' (2019) 2 Annual Review of Criminology 191.

insurable, and the adoption of firewalls', anti-virus software, secure passwords, and duo-factor authentication is no guarantee of security.[41]

The problem worsens when information passes quickly through many hands. Data aggregation has become a common practice in the tech industry, and many companies turn to vendors to supply both data and analytical services (rather than develop such capabilities internally). Firms known as 'data brokers' specialise in collecting and pooling information from multiple sources, and then providing it to clients. For example, a newsfeed from one of the wire services could be merged with government economic statistics, meteorological data, and Twitter postings to make predictions about demand for portable generators or home equity loans in a particular region, and then to send out targeted advertising on social media to potential buyers and borrowers, or to advise potential sellers and lenders. Much of this tracks individuals who have no idea of the scope of information that has been gathered about them, and who have no easy way to learn its extent or its provenance.[42] A welter of private contracts govern these activities, and often stipulate who owns data, or who may license its use for a period of time. The contracts rarely address the accuracy of the information, nor do brokers reliably ensure that their clients conform to the terms of use.[43] And the high volume of exchange among brokers would make it hard for subjects to identify erroneous information about themselves, or the source of the problem, and have it corrected. These arrangements are not subject to rigorous public oversight, or indeed any oversight at all.

Most data subjects 'consent' to use of information about themselves through end-user-agreements that insufficiently inform them about what can or will be done with such information. These agreements are typically standard-form contracts offered on a take-it-or-leave-it basis: to use software or an internet platform service, for example, the user must accept the terms dictated to them. With each new software release, the terms are updated and modifications are easy to overlook. Frequently, the service is offered 'free' to the user, so it will appear to be a good deal. A combination of economies of scale and network externalities make it hard for users to find viable alternatives, so to obtain the service they must accept the terms as given. But what users fail to appreciate, and what they are not told, is how much monetisation of the data streams created by use of the platform will benefit the

[41] Shauhin A Talesh, 'Data breach, privacy, and cyber insurance: how insurance companies act as "compliance managers" for businesses' (2018) 43 Law and Social Inquiry 417.

[42] Federal Trade Commission, *Data Brokers: A Call for Transparency and Accountability* (Washington DC 2014) vii, 14.

[43] Ibid 17, 41.

host company: by contributing content, users enable the platform to make money. Nor do they appreciation the frequency with which their information will be shared widely for any number of purposes. Issues of privacy, data ownership or data security are pushed to the side as aggregators strive to discover new ways to monetise the information they assemble, and the consent they obtain from data subjects is inadequately informed.[44]

Sometimes, individuals become data subjects without consent. Social media platforms vacuum up so much information that they can develop 'shadow profiles' about people who are not users and who therefore did not consent to the terms of use.[45] Such non-users might be, for example, the mutual friends of users, who show up in the list of contacts that the users uploaded onto the platform. Their names, phone numbers, postal and email addresses will become known to the platform. Non-users may be identified in the photos that users like to share, and their images included in facial recognition software. Time-stamped and geo-coded pictures ascertain time and place for the physical movements of users and non-users alike.

It took decades for the legal status of credit ratings and reports to settle in the late nineteenth century, as the judicial system grappled with the problems posed by (then) unprecedented amounts of credit information. A wave of regulation occurred in the 1960s, when policymakers recognised the significance of personal credit information and the potential for discrimination. The current period of expansion and innovation poses a new set of challenges, and again the legal/regulatory system struggles to catch up.

BIAS AND BIG DATA

The incompleteness of legal rules is particularly problematic for how information affects discrimination in markets. Data aggregators, like other tech firms, do not fit the strict definition of the credit rating agencies regulated by FCRA and so escape its oversight. Yet, increasingly, their informational products bear directly on credit markets. Similarly, the information in which they traffic is used by others for employment decisions, but since they themselves are not making the actual decisions, their obligations under Title VII of the Civil Rights Act are unclear. Antidiscrimination laws prohibited lenders from basing decisions on an applicant's race, sex, age, marital status,

[44] Sylvia Zhang, 'Who owns the data generated by your smart car?' (2018) 32 Harvard Journal of Law and Technology 299.

[45] Nicholas Diakopoulos, *Automating the News: How Algorithms are Rewriting the Media* (Harvard University Press 2019) 214.

or other protected category. One of the ostensible virtues of a computer algorithm is that it does not bear animus against people because of their race, gender, nationality, religion or other personal characteristics.[46] But does this mean algorithms are unbiased?[47] How to be sure that an automated decision does not turn on any of these features? The obvious solution is to exclude measures of race, sex, age, etc from the data upon which the algorithm operates. But in the context of big data, this is not enough. If there are variables included in the dataset that are correlated with any of these protected categories, singly or in combination, then an algorithm could discriminate in effect.[48] Given the opacity of the modelling process, it would be extremely difficult for an observer or regulator to know if such discrimination were occurring. For starters, algorithms are usually proprietary: they are part of the intellectual property belonging to the fintech firm. In addition, many of these algorithms are 'trained' on proprietary datasets. If the variables include geographic measures (postal codes, street addresses, or geocodes) then given the high level of residential segregation in most US communities it would be very easy to measure race indirectly. Similarly, given the homophilous nature of informal social networks, calculating an individual's social media connections could provide a proxy measure for race (or any number of other characteristics).[49]

The use of a broader set of information to make credit decisions, beyond the traditional credit report, makes it challenging to comply with rules prohibiting discrimination. And yet, exploitation of 'alternative information' also holds out the possibility of greater inclusion, faster decisions, and the extension of credit and financial services to the millions of 'unbanked' individuals.[50] A fintech lender can consider an applicant's on-time rental payments (which do not appear in a traditional credit record), their educational credentials, online behaviour, internet browser history, or information about an applicant's friends and associates.[51] Fintech may be able to exploit

[46] Matthew Adam Bruckner, 'The promise and perils of algorithmic lenders' use of big data' (2018) 93 Chicago-Kent Law Review 2, 5.

[47] Eirini Ntoutsi et al, 'Bias in data-driven artificial intelligence systems – an introductory survey' (2020) 10 WIREs Data Mining and Knowledge Discovery 1.

[48] Gillis and Spiess (n 36 above) 464, 469.

[49] Miller McPherson, Lynn Smith-Lovin and James M Cook, 'Birds of a feather: homophily in social networks' (2001) 27 Annual Review of Sociology 415.

[50] Bruckner (n 46 above) 6, 18.

[51] General Accounting Office, *Financial Technology: Agencies Should Provide Clarification on Lenders' Use of Alternative Data GAO-19-111* (Washington DC 2019) 33, 34; Congressional Research Service, *Alternative Data in Financial Services CRS IF11630* (Washington DC 2020).

opportunities that have been overlooked by traditional financial institutions.[52]

The necessary use of historical data to 'train' an algorithm risks reproducing historical biases. The general procedure is for some optimisation algorithm (eg one that identifies the most creditworthy borrowers) to be developed on an existing dataset, and then applied to new applications as they are submitted. Development usually involves estimation of parameters and coefficients that best link input information with some outcome that the developers care about (eg minimising loan defaults), and the bigger the dataset, the better. However, if biases have operated historically, for example if home mortgages were in the past extended in a discriminatory fashion, then the algorithm may well reflect those historical biases and reproduce them when applied to new loan applications. And because of the opacity of the algorithm, such bias may be hidden from those who use it. Safiya Noble gives the example of the Google search engine, trained on the billions of searches done by Google users to develop its 'auto-complete' algorithm (which offers suggestions to the user for how to complete their search phrase).[53] Because of racial biases in the user population, certain search phrases were 'auto-completed' in a racially biased manner, until the offensive pattern was called to the attention of Google, and its algorithm was modified. Unfortunately, algorithmic bias is seldom as obvious as it was for the auto-complete feature of Google. Another type of bias can arise when the training dataset itself is skewed, under-representing some population subgroups and over-representing others. If so, the algorithm may be good at estimating some associations, but bad at others, and will inadvertently misrepresent minority populations.

An additional complication stems from the fact that, as information from different sources gets aggregated, it is also being put to new uses. Often, data gathered for one purpose can be redeployed in an entirely different direction, and in a manner that evades existing rules. This means, for example, that a big tech firm getting involved in the provision of financial services is not necessarily subject to the regulations that normally govern banks because it was not founded as a bank and does not take deposits.[54] Or consider that a firm that harvests data about

[52] Julapa Jagtiani and Catharine Lemieux, 'Do fintech lenders penetrate areas that are underserved by traditional banks?' (2018) Federal Reserve Bank of Philadelphia Working Paper Series WP 18-13; Jagtiani and Lemieux (n 35 above).

[53] Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York University Press 2018).

[54] Xavier Vives, 'Digital disruption in banking' (2019) 11 Annual Review of Financial Economics 243.

online search behaviour and provides it to employers (who use it for their hiring decisions) may not be aware of how antidiscrimination standards like 'disparate impact' work. The dramatic repurposing of information renders older forms of regulation less relevant because it puts data in the hands of organisations that escape oversight or for whom compliance is unfamiliar.[55] One of the requirements of FCRA is that consumers are entitled to see the contents of their credit file at least once a year. This measure allows individuals to ensure the accuracy of fateful information about themselves. Yet, as alternative information becomes increasingly important in credit decisions, and as its volume and opacity grow, annual disclosure to the individual becomes less and less useful. What is an ordinary individual to make of a deep neural net model built out of millions of variables that governs their access to home equity loans? How to indicate the decisive information that led a lender to reject their loan application?

The discussion thus far has focused on how the rise of big data changes the volume and use of information about debtors, whether they are individuals or firms. As compared to the past, much more can be known about debtors and potentially it flips the traditional asymmetry between the two sides. In the market-for-lemons model, debtors know their own willingness and ability to repay but their creditors do not. Increasingly, however, creditors know more about debtors, even to the point where they know more than debtors do about themselves. This change presumably means that lenders are becoming much better at identifying truly creditworthy borrowers, including ones that previously would have been overlooked, although it raises thorny issues about privacy and non-discrimination.

BIG DATA AND COLLATERAL

Security is another key component of credit. In order to reduce their vulnerability, many lenders insist not only that the borrower have a good credit record, but also that the borrower provide collateral so that in the event of a default, the lender can seize the debtor's asset and use it to recover the unpaid balance of the debt. The collateral stays in the possession of the borrower for the duration of the loan, and so is available for use, but the lender has a security interest. Collateral has two effects: the possibility of its loss provides an incentive for the debtor to repay, and in the event the debtor does not do so, its liquidation helps to compensate the creditor. In the past, secured loans

[55] Mark T Andrus, 'The right to be forgotten in America: have search engines inadvertently become consumer reporting agencies?' (2016) May Business Law Today 1.

usually involved tangible assets like land, buildings, or large durable goods (eg automobiles). If the borrower defaulted on their home mortgage, for example, the bank could seize the home and sell it, using the proceeds to cover its losses. The mortgage identified the specific piece of property that functioned as collateral using information from a public land registry, which ensured that the borrower had clear title. Or if someone defaulted on their car loan, the finance company would repossess the car, knowing that a specific vehicle collateralised the loan because each car had a unique VIN (vehicle identification number). Registration of a security interest also ensured that lenders would know of any prior or senior liens and could adjust the terms of their loan accordingly.

New information capabilities create the possibility of expanding secured lending because it is becoming easier to register and track large numbers of valuable assets.[56] With a suitably elaborate information infrastructure in place, highly mobile, numerous, dynamic assets can be identified, registered, and tracked reliably enough that they can serve as collateral.[57] The 'internet of things' holds out the possibility that even mundane personal items can be tracked electronically and so can function like pawns in a pawnshop (except that they remain in the possession of the debtor).[58] In principle, the expansion of collateral could encourage more lenders to lend, knowing that they have been able more effectively to minimise their risks.

Despite this potential, the success of electronic registries is not assured. The case of MERS (the Mortgage Electronic Registration System) offers a sobering reminder that new informational infrastructures can face unexpected dysfunctionality. MERS was created in the 1990s as a private membership organisation to track home mortgages. With the development of securitisation, lenders moved away from the 'originate to hold' model and towards an 'originate to distribute' model. In the past, mortgage lenders typically held the debt until maturity, on their balance sheets. With the loan secured by real estate, they possessed the right to seize collateral in the event of a default and would register their security interest with a public administrative agency (often the Secretary of State's office). Securitisation meant that large numbers of home mortgages were transferred to new ownership (perhaps to a special purpose entity),

[56] Jacob Muirhead and Tony Porter, 'Traceability in global governance' (2019) 19 Global Networks 423.

[57] Roy Goode, 'Asset identification under the Cape Town Convention and Protocols' (2018) 81 Law and Contemporary Problems 135; Charles W Mooney Jr, 'fintech and secured transactions systems of the future' (2018) 81 Law and Contemporary Problems 1.

[58] Samuel Greengard, *The Internet of Things* (MIT Press 2015).

pooled together, and new securities were issued against that pool of assets. Securitisations became more complicated as different tranches were issued against the pool, varying by seniority. Nevertheless, it remained essential to track who held the lien, and therefore who had the right to foreclose on a defaulting mortgage, even when mortgages were blended together, repackaged in order of priority, and then distributed to multiple investors.

It was the purpose of MERS to track security interests through all the financial engineering. As growing numbers of mortgages changed hands, during the securitisation process and later in secondary market transactions, it was difficult and expensive to register every transfer with the public authorities. Instead, MERS as an entity became the nominal mortgage holder in relation to the outside world, and internally MERS tracked exactly who among the members held which rights over which mortgage.[59] As massive numbers of homeowners defaulted on their mortgages in 2008 and 2009, secured lenders moved to assert their rights and begin foreclosure. But in a significant number of legal cases, judges in different states refused to recognise their claims and ruled that the assertion that MERS operated as a kind of unchanging nominee on behalf of a changing group of lenders and their assignees was defective. MERS indeed kept track of the transfer of mortgages among its membership, but did not do so in a way that was recognised by the courts, and so foreclosure rights were not properly transferred. When creditors could not foreclose, the MERS system had clearly failed to function as its architects had intended, and one of the basic protections for creditors did not work.

The MERS experience showed how an innovative tracking system based on new information technologies could completely malfunction under pressure. Despite all the advantages of shifting from paper to electronic files, notwithstanding the ambition to reduce creditor costs, this system failed to articulate with the legal system in a manner that dependably supported the legal rights that creditors believed they possessed. Creditors thought they could foreclose on a mortgage in default, but they could not. Clearly, as finance moved to exploit new information technologies, including big data, it was critical to remain firmly anchored in the legal system. Financial claims have little manifestation except through law, so if their legal efficacy disappears, so does their value.

[59] David P Weber, 'The magic of the mortgage electronic registration system: it is and it isn't' (2011) 85 American Bankruptcy Law Journal 239; Laura A Steven, 'MERS and the mortgage crisis: obfuscating loan ownership and the need for clarity' (2012) 7 Brooklyn Journal of Corporate, Financial and Commercial Law 251.

CONCLUSION

New information technologies have produced an increase in the volume, velocity and variety of information that prompts some to suggest that we are in a new era of surveillance capitalism. Financial markets are particularly sensitive to such developments because of their dependence on information. Financial relationships involve uncertainty and vulnerability in that one party's fate depends on another's future actions. Participants therefore gather information so they can anticipate the future and seek better outcomes. They reduce their vulnerability by exploiting new informational capabilities to earmark assets and collateralise loans more broadly than in the past. And if they cannot guarantee a positive outcome, when held to account they can at least show they tried to manage the uncertainties and vulnerabilities.

Financial market participants do not always create information for themselves. Frequently, they acquire it from a third party, who is neither the subject nor the user of that information, but whose interests shape its production, distribution, and format. The acute demand for information means that financial market participants have been among the earliest adopters of the most advanced information technology, quickly embracing the postal system, telegraph, telephone, computer, and internet as these became available.

Although information has increased in volume, it also varies by source, format, content and use. To view change as exclusively quantitative is to overlook much of significance. Information can be bespoke or highly standardised. Its provenance may be public, or private. It may consist of qualitative classifications or quantitative measurements, and its format has been as much shaped by historical precedent as by the demands of users. Information may come with substantial legal obligations, or none at all. The legal status of information has varied, changing over time and depending on its usage. And contrary to the efficient markets hypothesis, many act as if market prices did not fully summarise all available information. Financial market participants are interested in prices, to be sure, but they are always interested in many other kinds of information as well.

Information continues to play a central role in contemporary markets. But the latest big data information technology, captured by the idea of surveillance capitalism, poses new challenges, particularly in the area of consumer finance. The US's existing legal and regulatory framework is strained by the volume, variety, speed and ubiquity of information, and its use in unanticipated ways. The inadequacy of the 'user consent' model, which relies on standard-form agreements to obtain 'informed' consent from data subjects, is now apparent. The billions of users who generate online data have no way to comprehend how data are being used or by whom. Nor do they appreciate their exposure to privacy violations or cybersecurity risks. Their consent is largely fictitious, ill-informed, and ceremonial. Some usages of information are legally restricted so as to prevent discrimination in contexts like employment or credit. But as alternative data becomes increasingly important, as proxy measures become readily available, as 'black box' algorithms play more of a role, and as information is repurposed for use in new contexts, existing legal restrictions become less effective. Who owns big data, as opposed to who controls it, remains an important unanswered question. Should private property be the default, or is it better to circumscribe and maintain an informational commons? How to make algorithms accountable for the decisions they render? Should the software engineers who write code be accountable? How to ensure that AI and ML algorithms do not unfairly discriminate against protected groups? Much more deliberation is required before these questions can be properly resolved.

Incumbent financial institutions feel the effects of big data. Upstart fintech firms do not usually qualify as banks, but on the loan side they are undertaking activities traditionally dominated by banks. Fintech's ability to exploit alternative data enables them to identify borrowers who were overlooked by traditional lenders. And if fintech firms lend, instead of providing a platform between lenders and borrowers, they will have to comply with know-your-customer regulations that target money-laundering and terrorism financing. The big tech firms, which already possess big data and know how to exploit it, could offer a range of financial services to their large user bases, and would pose a serious threat to incumbent banks. Among other things, their provision of payment services and other media of exchange would weaken the ability of central banks to control the money supply using their traditional policy instruments.

Some organisations appear to be unthreatened by these changes, including some at the very centre of information production and distribution. The bond rating agencies, for example, continue to produce a distinctive type of categorical information and play a central role in global financial markets, despite their US origins and even though their failures during the 2008 global financial crisis were widely noted and criticised. Attempts to create rival rating agencies outside of the US have failed, and their part in long-term capital allocation, structured finance, OTC financial derivatives markets, and prudential regulation seems largely unaffected by surveillance capitalism.[60]

[60] Eric Helleiner and Hongying Wang, 'Limits to the BRICS' challenge: credit rating reform and institutional innovation in global finance' (2018) 25 Review of International Political Economy 573; Mennillo and Sinclair (n 31 above).

In similar fashion, the rating organisations for individual credit (eg TransUnion) also seem relatively secure. They were already in the big data business, and so it is relatively easy for them to partner with fintech startups and find ways to improve their credit-scoring formulae by adding alternative variables to their models. Furthermore, the discovery of new ‘off-label’ uses for consumer credit scores helps to expand demand for their products. Since these scores directly shape the life chances of millions of individual consumers, when regulators confront the problems created by big data, new regulations will quite likely affect how the credit rating agencies operate.

Zuboff claims that a distinctively new era of surveillance capitalism poses unprecedented challenges and opportunities. But in some respects, panoptic surveillance of and by the capitalists populating the US financial system has been underway since the mid-nineteenth century. The widespread adoption and exploitation of cutting-edge information technology to address uncertainty and reduce vulnerability in a market setting is not a recent invention, nor are the associated legal and regulatory challenges. The historical experience underscores the dynamism and complexity of information as it diffuses and gets applied in new and unexpected ways. Instances of deep institutionalisation, where particular types of information are incorporated into basic structures of market governance, ensure that these utilisations survive the passage of time and surmount the shocks induced by economic crisis. It further suggests that regulatory interventions can make a difference, but also that measures tied too strictly to inflexible rules will likely be circumvented by innovative market actors. The traditional belief that ‘more is better’ can be problematic in situations where sharp information asymmetries exist and ‘too much information’ can easily overwhelm individuals subject to bounded rationality. Mandatory disclosure as a regulatory intervention seldom levels the playing field. In fact, much depends on the qualities of information, and its place in the architecture of decision-making, not just on its quantity. The new world of big data has been announced with dramatic claims about its novelty, but key parallels with the past offer a way to see past the hype.