DATA VALUATION AND LAW

JORDAN BARRY* & D. DANIEL SOKOL†

ABSTRACT

Data has become an increasingly valuable asset. Numerous areas of law—including contracts, corporate law, intellectual property ("IP"), antitrust, tax, privacy, and bankruptcy—require parties and courts to determine the value of assets, including data. Unfortunately, data valuation has been hindered by a lack of clarity over what data is and why it is valuable. This lack of clarity also increases the chances of legal decisionmakers valuing data in inconsistent ways, which would create further confusion, inefficiencies, and opportunities for regulatory arbitrage.

This Article proposes a unified framework for valuing data that will promote consistent valuations across fields of law. It begins by conceptualizing data as building blocks: It is of little value on its own. But when placed in skillful and creative hands, it can unlock choices for its holders—choices they would not otherwise have—that can generate tremendous profits. Thus, data constitutes what is known as a "real option." This Article shows how using real options to value data can significantly improve upon existing data valuation practices.

TABLE OF CONTENTS

* John B. Milliken Professor of Law and Taxation, USC Gould School of Law.

† Carolyn Craig Franklin Chair in Law, Professor of Law and Business, USC Gould School of Law and USC Marshall School of Business, and Senior Advisor, White & Case LLP.

We wish to thank the students at the law review for their excellent work as well as participants in the USC Faculty workshop.

INTRODUCTION

The rise of data analytics has been staggering. In 2021, 1.134 trillion megabytes were created every day, totaling 74 zettabytes for the year.[1] As large as this is, projections for 2022 are over 25% higher.[2] Big data and new information technology are changing the tools, business models, operations, and mindset that firms, nonprofits, and governments use every day, quietly transforming business and society.[3]

These changes come with challenges. A variety of legal regimes govern economic activity; in many instances, those legal regimes must determine the value of owning or using particular assets, including data.

For example, one area in which data valuation plays an important role is in contracting. Firms contract with each other daily with regard to the sale of data. This includes first-party data sales, such as when Target sells data that it has collected to Proctor & Gamble, as well as third-party data sales, in which data aggregators or brokers sell data that others have collected. If one party breaches the contract, what remedies are available to their

1. *See* Louie Andre, *53 Important Statistics About How Much Data Is Created Every Day*, FINS. ONLINE (July 16, 2023), https://financesonline.com/how-much-data-is-created-every-day [https://perma.cc/RKL6-9L8S].

2. Approximately 94 zettabytes of new data were projected to be created during 2022. *Id.*

3. *See generally* GEOFFREY G. PARKER, MARSHALL VAN ALSTYNE & PAUL SANGEET CHOUDARY, PLATFORM REVOLUTION: HOW NETWORKED MARKETS ARE TRANSFORMING THE ECONOMY AND HOW TO MAKE THEM WORK FOR YOU (2016); MARCO IANSITI & KARIM R. LAKHANI, COMPETING IN THE AGE OF AI: STRATEGY AND LEADERSHIP WHEN ALGORITHMS AND NETWORKS RUN THE WORLD (2020); AJAY AGRAWAL, JOSHUA GANS & AVI GOLDFARB, POWER AND PREDICTION: THE DISRUPTIVE ECONOMICS OF ARTIFICIAL INTELLIGENCE (2022).

counterparty?[4] In corporate law, target boards have fiduciary duties to make sure their shareholders are being appropriately compensated during mergers and acquisitions. This requires having a handle on the value of the target firm's assets, including its data.[5] In tax, the taxation of intangible assets and specifically of data is a growing issue of concern.[6]

These questions can potentially be even thornier when specific aspects of data must be valued, rather than full ownership. To take another example, suppose that one firm's negligence results in another firm's proprietary data leaking to the public. To award damages, a court must determine how much the damaged firm lost from having the data become public—but how much is that?[7] Similarly, in antitrust, when control of data plays an important role in anticompetitive behavior, is it ownership of the data itself that creates the problem, or the use of the data?[8] Does sharing the data with competitors make matters better or worse?[9] The rise of generative artificial intelligence ("AI"), which requires data for its machine learning models, may create additional concerns as to the value of various data usage rights.

Unfortunately, the difficulties of conceptualizing data have hampered law's attempts to incorporate the data revolution into multiple legal doctrines. This has opened the door to confusion, inconsistency, and inefficiency. Decisionmakers have confused data with algorithms, and struggled with how to apply certain doctrines to the legal rights that data owners and data users possess. This increases the risks that regulators in different substantive areas of law, as well as in different jurisdictions, will take inconsistent approaches. This creates inefficiencies as parties subject to multiple regimes work to navigate them. Different legal regimes also creates opportunities for regulatory arbitrage, in which regulated parties take advantage of divergent regulatory rules to achieve the regulatory treatment they want while making only minor changes to their economic activities.

To address these concerns, this Article offers a general framework for valuing data based on real options valuation. The financial economics

4. Cemre Bedir, *Contract Law in the Age of Big Data*, 16 EUR. REV. CONT. L. 347, 362–64 (2020).

5. Doron Nissim, *Big Data, Accounting Information, and Valuation*, 8 J. FIN. & DATA SCI. 69, 70 (2022).

6. Young Ran (Christine) Kim & Darien Shanske, *State Digital Services Taxes: A Good and Permissible Idea (Despite What You Might Have Heard)*, 98 NOTRE DAME L. REV. 741, 797–798 (2022).

7. D. Daniel Sokol & Tawei Wang, *A Review of Empirical Literature in Information Security*, 95 S. CAL. L. REV. 95, 109 (2021).

8. *See* Tilman Kuhn, Kristen O'Shaughnessy, Tobias Pesch, Jaclyn Phillips & D. Daniel Sokol, *Big Data and Data-Related Abuses of Market Power*, *in* RESEARCH HANDBOOK ON ABUSE OF DOMINANCE AND MONOPOLIZATION 438, 438–55 (Pinar Akman, Or Brook & Kristianos Stylianou eds., 2023) (providing an overview of cases in the United States and European Union).

9. *Id.*

literature pioneered the use of real options to better assess business decision-making under uncertainty.[10] This approach has since been extended beyond finance to address other areas of uncertainty.[11] Real option analysis provides a better path forward than the current patchwork of doctrinal and analytical approaches. A real options approach is conceptually correct and thus has the potential to ameliorate the confusion, inconsistency, and inefficiency of existing approaches. To our knowledge, this is the first article to utilize real options as a method to value data, in law or otherwise.

Along with its potential benefits as a method of data valuation, real options analysis does have its drawbacks. Real options theory is complicated, which creates implementation challenges that must be overcome, or at least managed, to achieve the benefits described above. That said, real options analysis is an improvement over existing approaches. Applying a more unified theory also allows for a more standardized approach that can then be tailored to specific doctrines and areas of law.

This Article proceeds as follows. Part I provides context regarding the big data revolution and the growing importance of data. In doing so, it reviews the extant theoretical and empirical literatures on data valuation. Part II identifies the implications of data valuation for law by providing some case studies across fields. It includes vignettes demonstrating the types of issues that emerge and some current legal approaches. Next, in Part III, the Article explores how real options analysis offers a viable potential solution to the current patchwork of legal approaches. The Article concludes on how agencies and courts would benefit from such an approach, notes limitations on the use of real options, and offers avenues of future research.

I. THE DATA REVOLUTION AND THE VALUE OF DATA

To understand the importance of data valuation methods to the law, one must understand two other, related points. First, one must have a grounding in why and how data is used in the modern economy. Second, one must consider how to think about how those use cases translate into value estimates.

10. *See generally* AVINASH K. DIXIT & ROBERT S. PINDYCK, INVESTMENT UNDER UNCERTAINTY (1994).

11. *See, e.g.*, Joseph A. Grundfest & Peter H. Huang, *The Unexpected Value of Litigation: A Real Options Perspective*, 58 STAN. L. REV. 1267, 1282–91 (2006); Andrew Chin, *Teaching Patents as Real Options*, 95 N.C. L. REV. 1433, 1434–35 (2017).

A. DIGITAL TRANSFORMATION

To understand the role of data in the modern economy, one must consider three related points: (1) The increase in AI techniques that can generate value from data; (2) The increase in data to which such AI techniques can be applied; and (3) The amount of value that these techniques are creating. Understanding these dynamics allows us to explore specific case studies that apply these insights across a number of areas of law.

1. Generating Value from Data with AI

As a starting point, companies across the economy have moved to increasingly digitized, AI-enabled business strategies, producing profound effects on value creation and innovation.[12] We use the term AI broadly here, as a way to encompass algorithms that improve prediction and decision-making.[13] There are different approaches to AI, such as neural networks and machine learning, among others.[14]

When thinking about data and AI, it can be helpful to consider a simple, three-tier vertical model of how companies and other actors use data and AI to further their goals.

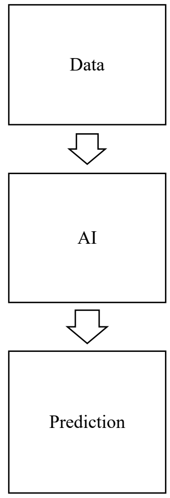
---

12. IANSITI & LAKHANI, *supra* note 3, at 28–40; AJAY AGRAWAL, JOSHUA GANS & AVI GOLDFARB, PREDICTION MACHINES: THE SIMPLE ECONOMICS OF ARTIFICIAL INTELLIGENCE 11–13 (2018); Hau L. Lee, *Big Data and the Innovation Cycle*, 27 PROD. & OPERATIONS MGMT. 1642, 1645–46 (2018); Hal R. Varian, *Big Data: New Tricks for Econometrics*, 28 J. ECON. PERSPS. 3, 7–25 (2014) (analyzing the uses of big data in economics). Many companies have become platforms, where the ability to create economies of scale and scope have allowed for a generation of "new opportunities to create, appropriate, and deliver value for firms and [users] . . . ." D. Daniel Sokol, *Technology Driven Government Law and Regulation*, 26 VA. J.L. & TECH. 1, 2 (2023).

13. For applications in law, see for example, Amy L. Stein, *Artificial Intelligence and Climate Change*, 37 YALE J. ON REG. 890, 895–900 (2020); Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1829–32 (2019); W. Nicholson Price II, *Regulating Black-Box Medicine*, 116 MICH. L. REV. 421, 432–37 (2017).

14. Xiao Liu, Dokyun Lee & Kannan Srinivasan, *Large-Scale Cross-Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning*, 56 J. MKTG. RSCH. 918, 924–25 (2019) (using neural networks in marketing research); Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 871–80 (2016) (discussing machine learning in a legal context).

FIGURE 1.



At the first stage is data. If AI is the product or output, data serve as the input. Data feed the needs of AI-enabled technologies. Data underlie machine learning and prediction models, and it is data that has fueled digital transformation.[15] Without sufficient quantity and quality of data, many current AI techniques simply cannot produce very good results.

Data often is the input to the next stage—powering an algorithm. The algorithm itself is not the end of the production. Rather, the algorithm simply enables better prediction. It is at the stage of prediction where there are outputs to AI—outputs that can generate tremendous value.

For example, when a user types terms into a search engine, that engine might consider data about what sites other users who typed in similar terms ultimately clicked on (among other data) when deciding what results should appear. Diagnostic software might compare a patient's MRI to millions of MRI images that have already been analyzed by doctors to estimate the likelihood that the patient has breast cancer. Data drives the AI, the AI makes predictions, and those predictions enable better decision-making, which creates economic value.

15. Marshall Fisher & Ananth Raman, *Using Data and Big Data in Retailing*, 27 PROD. & OPERATIONS MGMT. 1665, 1666–67 (2018); Anindya Ghose & Vilma Todri-Adamopoulos, *Toward a Digital Attribution Model: Measuring the Impact of Display Advertising on Online Consumer Behavior*, 40 MGMT. INFO. SYS. Q. 1, 2–3 (2016).

2. Increase in Data

While many facets of AI are themselves not new, the speed of data collection and processing have significantly improved these tools' impact.[16] Data is vast and the various ways to use it have grown significantly, such that there are distinct data-related strategies that firms may adopt.

The data ecosystem is worth exploring briefly. Data can be bought and sold like many other inputs.[17] It can be acquired from public sources. It can be collected from what can be termed data suppliers. For example, first-party companies such as Netflix or Spotify can sell their data and databases to other companies—firms regularly sell large quantities of this type of data through basic business transactions.[18] Third-party data brokers, apps and internet service providers ("ISPs") that can provide locational or other data, and data aggregators also play significant roles in the data ecosystem.[19] Data brokers buy and sell data, thereby allowing firms to acquire new data to make better predictions.[20] This increase in data sources is an important change, as it makes data more widely available. This both enables more actors to put it to use and to experiment and innovate with it.[21]

Indeed, data has become both a make and buy decision.[22] That is, firms have significant opportunities to generate their own data—such as Target keeping track of what consumers buy at Target—and to acquire third-party data from other actors. This is especially true with respect to end-consumer data.[23]

16. Ajay Agrawal, Joshua Gans & Avi Goldfarb, *Prediction, Judgment, and Complexity: A Theory of Decision-Making and Artificial Intelligence*, *in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE 89, 93 (Ajay Agrawal, Joshua Gans & Avi Goldfarb eds., 2019).

17. Maryam Farboodi & Laura Veldkamp, *Data and Markets* 1 (Mass. Inst. of Tech. Sloan, Research Paper No. 6887–22, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4284192 [https://perma.cc/M4JS-4Y2A].

18. Firms also sell "exhaust" data; this is data sold for what are unrelated to business transactions but have a secondary purpose for other kinds of business.

19. Llewellyn D.W. Thomas & Aija Leiponen, *Big Data Commercialization*, 44 INST. ELEC. & ELECTRONICS ENG'RS: ENG'G MGMT. REV. 74, 80 (2016).

20. *See* Nico Neumann & Catherine Tucker, *Data Deserts and Black Boxes: The Impact of Socio-Economic Status on Consumer Profiling* (February 27, 2023) (unpublished presentation) (on file with the Southern California Law Review); Arion Cheong, D. Daniel Sokol & Tawei Wang, *Cookie Intermediaries: Does Competition Leads to More Privacy?* 2–5 (April 16, 2023) (unpublished manuscript) (on file with *Southern California Law Review*).

21. To the extent that data is accessible from many sources, that weakens arguments that data access is a key barrier to entry.

22. *See* Jordan M. Barry & Victor Fleischer, *Tax and the Boundary of the Firm* 2–7 (Aug. 28, 2023) (unpublished manuscript) (on file with *Southern California Law Review*). *See generally* R.H. Coase, *The Nature of the Firm*, 4 ECONOMICA 386 (1937).

23. *See* Alessandro Bonatti, Munther Dahleh, Thibaut Horel & Amir Nouripour, *Selling Information in Competitive Environments* 4–5 (Mass. Inst. of Tech. Sloan Sch. of Mgmt., Working Paper No. 6532-21, 2022), https://arxiv.org/pdf/2202.08780 [https://perma.cc/7MWJ-AZNQ]; Anja Lambrecht

3. Amount of Value

What is this power of data? Typically, data is defined across four "V's": velocity, veracity, volume and variety.[24] Combined, these four Vs create data value. Velocity is the speed at which data is collected and used. Volume is the sheer amount of data that is generated, which (at least at present) overwhelms our ability to process it; there is more data than ever before and every day we create 328.77 million terabytes of new data.[25] Veracity goes to the increasingly important issues of data accuracy and trustworthiness. Finally, variety reflects the diversity of data types that can be collected and used, such as e-mails, PDFs, and videos.

Data may come from many sources. The general rule of data is that the more the data, the greater the ability to feed AI and the better the ability to improve prediction,[26] although there are limits to what data alone can do.[27] Data must be processed, via AI or otherwise, to reap benefits.[28] When properly processed, big data allows firms to improve their products and services and to develop new such products and services.[29]

The academic and practitioner literature on data valuation is complex. First, there is the literature on data brokers. In some senses, the costs of data are lower now than ever before.[30] The reduced cost of data allows for the creation of a wide variety of sophisticated algorithms that can produce

& Catherine E. Tucker, *Can Big Data Protect a Firm From Competition?*, COMPETITION POL'Y INT'L ANTITRUST J. (Jan. 17, 2017), https://www.competitionpolicyinternational.com/can-big-data-protect-a-firm-from-competition [https://perma.cc/JK39-W2CR]; Thomas & Leiponen, *supra* note 19, at 80.

24. *See* A.B.A. SECTION OF ANTITRUST, ARTIFICIAL INTELLIGENCE & MACHINE LEARNING: EMERGING LEGAL AND SELF-REGULATORY CONSIDERATIONS (PART ONE) 2 (2019), https://www.americanbar.org/content/dam/aba/administrative/antitrust_law/comments/october-2019/clean-antitrust-ai-report-pt1-093019.pdf [https://perma.cc/F9S2-8P5Q].

25. *See* Petroc Taylor, *Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2020, with Forecasts from 2021 to 2025*, STATISTA (Sept. 8, 2022), https://www.statista.com/statistics/871513/worldwide-data-created [https://perma.cc/LZ5B-CSFM].

26. IANSITI & LAKHANI, *supra* note 3, at 16–27; Andrei Hagiu & Julian Wright, *Data-Enabled Learning, Network Effects and Competitive Advantage* 3 (May 2021) (unpublished manuscript), https://app.scholarsite.io/julian-wright/articles/data-enabled-learning-network-effects-and-competitive-advantage-3 [https://perma.cc/6J8A-L8MU].

27. *See, e.g.*, Carmelo Cennamo, *Building the Value of Next-Generation Platforms: The Paradox of Diminishing Returns*, 44 J. MGMT. 3038, 3039–41 (2018) (identifying diminishing returns to data); Hanna Halaburda, Mikolaj Jan Piskorski & Pinar Yildirim, *Competing by Restricting Choice: The Case of Matching Platforms*, 64 MGMT. SCI. 3574, 3574–76 (2017) (identifying network saturation allowing for competition through differentiation in platforms); D. Daniel Sokol & Roisin Comerford, *Antitrust and Regulating Big Data*, 23 GEO. MASON L. REV. 1129, 1135–40 (2016) (illustrating that it is not the data but what you do with them that matters as well as other limits to data).

28. Ron Berman & Ayelet Israeli, *The Value of Descriptive Analytics: Evidence from Online Retailers*, 41 MKTG. SCI. 1074, 1076 (2022) (finding that e-commerce data analytics dashboards increase weekly revenues between 4%–10%).

29. Sokol & Comerford, *supra* note 27, at 1134.

30. Avi Goldfarb & Catherine Tucker, *Digital Economics*, 57 J. ECON. LITERATURE 3, 3 (2019).

insights that would elude unassisted humans.[31] The ability to utilize data to feed AI allows for opportunities to better create, appropriate, and deliver economic value not merely for AI-driven firms but for the different users of digital platforms such as advertisers, complementors, and customers.[32]

This transformation creates significant economic value, but the drivers of that value are not well understood by courts and regulatory bodies. In some cases, regulation might stymie the use of data and chill innovation and investment.[33] In other cases, the potential portability of certain types of data has motivated increased legislative and regulatory action.[34] In other situations, courts have held that owners of certain types of data have certain rights, such as the right to exclude others from such data. The exact value—either of the underlying data itself or of the rights to exclude others—may not always be clear.[35] There are yet other areas in which data-related transactions occur on a regular basis, but which have not produced judicial decisions to date.[36]

It is these sorts of complexities as to law and data to which we next turn.

B. DISAGREEMENTS ON HOW TO THINK ABOUT DATA CREATING VALUE

Valuing data presents conceptual challenges because data is unlike other assets, including other intangible assets. The first problem is to understand how even though data is a building block for constructing a final product, data is not like traditional tangible assets such as bricks and steel used to make a factory. Data can be collected and mixed in a number of different, complex ways. Further, unlike bricks, data is non-rivalrous; more than one firm can use the same data.[37] For instance, someone's driving

31. IANSITI & LAKHANI, *supra* note 3, at 62–70.

32. Ron Adner, Phanish Puranam & Feng Zhu, *What Is Different About Digital Strategy? From Quantitative to Qualitative Change*, 4 STRATEGY SCI. 253, 258 (2019); Michael G. Jacobides, Carmelo Cennamo & Annabelle Gawer, *Towards a Theory of Ecosystems*, 39 STRATEGIC MGMT. J. 2255, 2257 (2018); Geoffrey Parker, Marshall Van Alstyne & Xiaoyue Jiang, *Platform Ecosystems: How Developers Invert the Firm*, 41 MGMT. INFO. SYS. Q. 255, 259 (2017).

33. *See* Jian Jia, Ginger Zhe Jin & Liad Wagman, *The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment*, 40 MKTG. SCI. 661, 677 (2021) (finding a decrease in venture capital investment as a result of GDPR); Rebecca Janssen, Reinhold Kesler, Michael E. Kummer & Joel Waldfogel, *GDPR and the Lost Generation of Innovative Apps* 1 (Nat'l Bureau of Econ. Rsch., Working Paper No. 30028, 2022) (finding a reduction of apps by one third as a result of GDPR).

34. ORG. FOR ECON. COOP. & DEV., DATA PORTABILITY, INTEROPERABILITY AND DIGITAL PLATFORM COMPETITION 42 (2021).

35. Francesco Decarolis & Gabriele Rovigatti, *From Mad Men to Maths Men: Concentration and Buyer Power in Online Advertising*, 111 AM. ECON. REV. 3299, 3299–303 (2021) (discussing ad auctions).

36. *Id.*

37. Charles I. Jones & Christopher Tonetti, *Nonrivalry and the Economics of Data*, 110 AM. ECON. REV. 2819, 2834 (2020).

history can be used at the same time by multiple firms, in the same or different industries (for example, advertisers, insurance companies, credit card companies). As Jones and Tonetti explain:

> An analogy may be helpful. Because capital is rival, each firm must have its own building, each worker needs her own desk and computer, and each warehouse needs its own collection of forklifts. But if capital were nonrival, it would be as if every auto worker in the economy could use the *entire* industry's stock of capital at the same time. Clearly this would produce tremendous economic gains. This is what is possible with data.[38]

Thus, non-rivalry means that valuation may be harder across a number of the traditional measurements.

Further complicating data is that it is (mostly) non-exclusive.[39] For example, if someone collects public records about home purchases into a comprehensive database, that does not prevent others from collecting that same information in the same way. This is a stark contrast from some other intangible assets, including traditional forms of IP such as patents and copyrights, which create value by conferring exclusive rights on their holders.[40]

Both of these indicia suggest that the underlying value of the data, rather than that of the algorithm, may be small. When the input (data) is easily available to all, it is the actor's ability to make use of the input—that is, the algorithm—that creates the value, not the input itself. For example, a classic crème brûlée recipe has only four ingredients—cream, sugar, egg yolks, and vanilla. All of these items are widely available. The ability to charge a premium for the final product is a function of the baking skill of the pastry chef.

Beyond non-rivalry and non-excludability, some regulation, such as the European Digital Markets Act[41] requires fair, reasonable, and non-discriminatory ("FRAND") licensing. Even in IP and antitrust, FRAND terms are not always clearly understood.[42] It stands to reason that in data, with fewer cases to provide guidance across different areas of law, the nature

38. *Id.* at 2820.

39. *But see* AUTORITÉ DE LA CONCURRENCE, DÉCISION N° 14-MC-02 DU 9 SEPTEMBRE 2014 RELATIVE À UNE DEMANDE DE MESURES CONSERVATOIRES PRÉSENTÉE PAR LA SOCIÉTÉ DIRECT ENERGIE DANS LES SECTEURS DU GAZ ET DE L'ÉLECTRICITÉ (2014) (identifying unique data because of regulation as to customer data and contracts).

40. John P. Conley & Christopher S. Yoo, *Nonrivalry and Price Discrimination in Copyright Economics*, 157 U. PA. L. REV. 1801, 1818–19 (2009).

41. *Proposal for a Regulation of the European Parliament and of the Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act)*, COM (2020) 842 (Dec. 15, 2020) [hereinafter *Proposal for a Regulation*].

42. Herbert Hovenkamp, *FRAND and Antitrust*, 105 CORNELL L. REV. 1683, 1684 (2020).

of FRAND obligations is even less clear. Further, certain types of data have sharing requirements in practice that may change the valuation of data, such as requirements for data portability.

Data is also unlike some other intangible assets because of the speed at which data can become obsolete.[43] Much data gets stale over time.[44] This suggests that much data is a diminishing asset, something which IP such as patents or copyrights do not face nearly as quickly because those rights last for longer periods.

II. THE IMPLICATIONS OF DATA VALUATION FOR LAW

There are many areas of law for which valuation of various assets is important. Data is an increasingly valuable asset. Unfortunately, there is currently relatively little law on how to value data. Courts and regulators have generally avoided the question whenever possible, perhaps out of concern for the difficulty of the problem, or uncertainty on how to proceed, and often such cases get decided upon other grounds. This raises the chances that different legal areas will use different valuation methods. Such inconsistency creates dilemmas as to how to allocate legal rights and responsibilities. Perhaps the clearest way of understanding this tension across areas of law is to consider the purpose of damages. Damages exist to compensate a potential victim for violations of law and/or to deter the violator from doing so again.[45] Methods across areas of law might include: (1) a cost-based approach based on the replacement cost; (2) a market-based approach based on similar acquisitions of data (or companies with data); and (3) an income-based approach, to the extent that the data is producing income via sales or even licensing. To this, we add the importance of a fourth possibility, an options-based approach. Often, outcomes seem to be highly contextual rather than based on valuation methodology.[46] A lack of consistency is significant because of the growing stake of data as an important part of economic activity.

43. Ehsan Valavi, Joel Hestness, Marco Iansiti, Newsha Ardalani, Feng Zhu & Karim R. Lakhani, *Time Dependency, Data Flow, and Competitive Advantage* 10 (Harv. Bus. Sch., Working Paper No. 21-099, 2021) ("High perishability undermines the importance of data volume or historical data in creating a competitive advantage.").

44. Ehsan Valavi, Joel Hestness, Newsha Ardalani & Marco Iansiti, *Time and the Value of Data* 1 (Harv. Bus. Sch., Working Paper No. 21-016, 2020).

45. Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169, 172–73 (1968). There are other potential justifications for damages, such as retributivism, but these are the two justifications raised most frequently in the civil context.

46. Feng Chen, Kenton K. Yee & Yong Keun Yoo, *Robustness of Judicial Decisions to Valuation-Method Innovation: An Exploratory Empirical Study*, 37 J. BUS. FIN. & ACCT., 1094, 1097 (2010).

Which approach ultimately to take across areas of law such as IP, antitrust, mergers and acquisitions (M&A), bankruptcy, torts, and other areas of law varies. One important driver is what information courts and parties can easily measure. When contracts (and comparable transactions) are not easy to find, private negotiations between contracting parties in the shadow of the law are another important driver. These questions become more salient as we try to understand how issues involving big data reverberate across a number of areas of law and in terms of the value of data overall. The biggest question is how much value do we think is in big data?[47]

A. Valuation Is Important to Many Areas of Law

Below we offer some examples of how data valuation plays a role across various areas of law. We highlight these examples as a way to understand some of the complexity that requires a more generalized rethink as to valuation method of data. Understanding these complexities helps clarify the value of data as well as some of the struggles that different areas of law are currently experiencing as they seek to value data.

1. Antitrust

Antitrust has tried to address the questions of competition and the exercise of market power in two contexts—mergers and conduct cases. These produce two types of antitrust cases—those where data is an input and those in which data is a product. However, there is little caselaw in each area. Consequentially, the problem with both sets of circumstances is that we tend not to see litigated cases that get to the valuation issue of the data.

Antitrust primarily addresses behavior one of two ways. The first is through ex ante enforcement through merger control. Essentially, regulators can block mergers that are expected to produce antitrust problems. On the mergers side, most cases do not go to court, which means that litigated cases may not be representative. Even in those cases for which there is a judicial opinion, not all issues may get addressed. Scholars have expressed general frustration with what gets decided under the shadow of merger law.[48] Thus, the basis for decisions on many issues, including data valuation, is limited or incomplete. As Professors Katz and Shelanski lament, "The overall picture

47. We assume that data creates value. *See* Maryam Farboodi, Roxana Mihet, Thomas Philippon & Laura Veldkamp, *Big Data and Firm Dynamics*, 109 AM. ECON. ASSOC. PAPERS & PROC. 38, 42 (2019). We might also imagine that information is simply a byproduct of economic activity. *See* Pablo D. Fajgelbaum, Edouard Schaal & Mathieu Taschereau-Dumouchel, *Uncertainty Traps*, 132 Q. J. ECON. 1641, 1642 (2017).

48. D. Daniel Sokol & James A. Fishkin, *Antitrust Merger Efficiencies in the Shadow of the Law*, 64 VAND. L. REV. EN BANC 45, 45–46 (2011).

of current merger enforcement practice is, therefore, murky."[49]

Cases provide some guidance on how antitrust courts and agencies think about data, which gives some insight on how to think about data's value. Yet much uncertainty remains. As of this writing, no mergers have been blocked on data theory grounds in the United States. Nor have there been any decided cases that explain the valuation method used for such transactions that weigh the data rather than its use to a specific platform.

In the case of data, let us begin with mergers and the possibility that data is itself the market. One such deal that included data as the market is the 2014 CoreLogic-DataQuick merger.[50] In that transaction, the Federal Trade Commission cleared the transaction with a database divestiture but did not explain the valuation technique employed. Alas, this has been typical with regard to antitrust analysis of mergers that include data as a market. Similarly, people generally have not discussed mergers that include valuable data as an input (for example, Microsoft/LinkedIn, Apple/Shazam) as matters of valuation. At best, there are transactions that have received some sort of conditional approval such as Nielsen/Arbitron but without an explicit discussion of data valuation.[51]

Antitrust, through public and private enforcement, polices against anticompetitive conduct by one or more firms that harms competition. Conduct cases in antitrust involving data issues have not resolved the data valuation question, either. Complicating antitrust further is that duties to deal with competitors are limited, which means that such data sharing cases do not get to the data valuation stage of the case. Rather, these cases are decided based on the premise that data is not required to be shared in the first place. Yet, understanding such cases helps to explore the value of data because the discussion helps to inform the value of data use and ownership.

For example, Section 2 of the Sherman Act generally imposes no requirements to deal with one's competitors.[52] In *Aspen Skiing Co. v. Aspen Highlands Skiing Corp.*, the Supreme Court held that there are some limited circumstances under which Section 2 requires monopolistic firms to deal with their rivals.[53] Courts have further narrowed *Aspen Skiing*'s holding since. Most recently, the DC Circuit dismissed a monopolization case that forty-six states brought against Meta based on the court's narrow reading of

49. Michael L. Katz & Howard A. Shelanski, *Merger Analysis and the Treatment of Uncertainty: Should We Expect Better?*, 74 ANTITRUST L.J. 537, 547 (2007).

50. *See* Decision & Order at 5–8, *In re* CoreLogic, Inc., Docket No. C-4458 (F.T.C. May 21, 2014).

51. *See* Decision & Order at 5–7, *In re* Nielsen Holdings N.V., Docket No. C-4439 (F.T.C. Feb. 28, 2014).

52. Sherman Act, 15 U.S.C. § 2 (1982).

53. Aspen Skiing Co. v. Aspen Highlands Skiing Corp., 472 U.S. 585, 585 (1985).

Aspen Skiing.[54]

Cases brought under other provisions of the Sherman Act have also implicated the value of data. However, much like the Section 2 monopolization cases, courts examining Section 1 of the Sherman Act have offered little guidance on how to value data. For example, in *Authenticom, Inc. v. CDK Global, LLC*, Authenticom brought a claim against CDK for closing its system for data and thereby barring data scrapers from access. The Seventh Circuit ruled in favor of CDK on the basis that forced data sharing was inconsistent with precedent.[55] Because of this ruling, which dismissed the case on essential facilities grounds, the data valuation issue was never addressed. Of course, that does not mean that the data does not have value, merely that the court was able to dispose of the case without determining what the data's value was.

Similar to antitrust enforcement, competition regulation increasingly plays an important role in big data valuation. This comes up specifically in the case of the Digital Markets Act ("DMA"), the European approach to ex-ante regulation of data.[56] Regarding "gatekeeper" firms, the DMA states:

> The gatekeeper shall provide to any third-party undertaking providing online search engines, at its request, with access on fair, reasonable and non-discriminatory terms to ranking, query, click and view data in relation to free and paid search generated by end users on its online search engines. Any such query, click and view data that constitutes personal data shall be anonymised.[57]

Of course, data from a gatekeeper will not generate profits on its own; gatekeeper data must still be combined with some effort by recipients. But this reality makes it harder to assess the incremental profits the recipient earns as a result of having access to the data.[58]

2. Business Law

Business law increasingly confronts data valuation. Unfortunately, it does so in ways that do not always show the precision that we believe is necessary to unlock a more accurate value of data assets. For example, data

54. New York v. Meta Platforms, Inc., 66 F.4th 288, 305 (D.C. Cir. 2023). Guam and the District of Columbia were also plaintiffs in the litigation.

55. Authenticom, Inc. v. CDK Global, LLC, 874 F.3d 1019, 1025–27 (7th Cir. 2017).

56. *Proposal for a Regulation*, *supra* note 41, at 7. *See* Nicolas Petit, *The Proposed Digital Markets Act (DMA): A Legal and Policy Review*, 12 J. EUR. COMPETITION L. & PRAC. 529, 529–32 (2021) (providing an overview of the Digital Markets Act).

57. Digital Markets Act, 2022 O.J. (L 265) art. 6 ¶ 11.

58. Incremental revenue, which one might hope to observe, will overstate the benefits; one must also consider incremental costs.

valuation questions arise within the context of both mergers and acquisitions ("M&A") and bankruptcy. A number of factors arise in each context that make data valuation more difficult. Within the merger context, the purpose of valuation is to best help the acquiring and target boards to fulfill their fiduciary duties to ensure that the price paid for the acquisition is an appropriate one.

Overall, corporate law has grappled with how to account for intangibles. Many assets, including branding and intangibles such as IP, are lumped together under the heading of "goodwill." However, the goodwill from reputation and branding is different than goodwill that is the basis of a regenerative asset such as data. Further, how data is stored and how easily it can be processed and integrated make such a valuation more challenging.[59]

Different data sets may have different levels of privacy requirements, such as data that is protected under the Health Insurance Portability and Accountability Act ("HIPAA") versus commercial health data, which has less stringent requirements. Identifying what sort of data companies may keep, for how long, how stale such data get, and the potential liabilities of such data are complex.[60] Yet, there are very few cases that offer direct guidance on how to value data in the corporate and M&A setting. Thus, data valuation ends up a financial black box with potentially large implications if and when such cases go to litigation. This sort of uncertainty creates potential risk for deals, particularly those deals for which the underlying data may be a significant asset.[61]

Finally, unresolved issues include requirements of how to store data[62] as well as how to destroy data.[63] The lack of uniform federal privacy

59. Chengxin Cao, Gautum Ray, Mani Subramani & Alok Gupta, *Enterprise Systems and M&A Outcomes for Acquirers and Targets*, 46 MGMT. INFO SYS. Q. 1295, 1299–300 (2022) (identifying similar issues in the context of integration of business enterprise software in M&A).

60. Sometimes firms might unknowingly buy a data lemon, with liabilities that attach because of a data breach, such as Marriot's acquisition of Starwood's hotel chain. However, this is a somewhat different question than valuing the data set itself. Chirantan Chatterjee & D. Daniel Sokol, *Don't Acquire a Company Until You Evaluate Its Data Security*, HARV. BUS. REV. (April 16, 2019), https://hbr.org/2019/04/dont-acquire-a-company-until-you-evaluate-its-data-security [https://perma.cc/XH4E-BK6M].

61. Michel Benaroch, Yossi Lichtenstein & Karl Robinson, *Real Options in Information Technology Risk Management: An Empirical Validation of Risk-Option Relationships*, 30 MGMT. INFO. SYS. Q. 827, 828 (2006) (suggesting a risk management-based approach to address the uncertainty associated with data breaches).

62. Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 B.C. L. REV. 1687, 1706 (2020).

63. Some forms of data disposal have specific regulation. *See, e.g.*, *Disposing of Consumer Report Information? Rule Tells How*, U.S. FED. TRADE COMM'N (June 2005), https://www.ftc.gov/business-guidance/resources/disposing-consumer-report-information-rule-tells-how [https://perma.cc/RWW9-2EXJ].

legislation makes such analysis more difficult. Federal agencies, especially the FTC, enforce privacy protections,[64] but private actions also play a role.[65] Moreover, states can impose additional rules on top of the federal ones. For example, California took inspiration from the General Data Protection Regulation ("GDPR") and adopted the California Consumer Privacy Act ("CCPA") and the California Privacy Rights Act ("CPRA").[66]

This issue of data valuation similarly plays itself out in the bankruptcy setting. In some settings, the data itself, such as customers' spending behavior,[67] may be the asset. Take the example of the bankruptcy proceeding for Caesar's Entertainment Operating Corp casinos.[68] Creditors viewed the company's data (customer-specific data on spending habits) as one of the company's most important assets. Yet, as is often the case in bankruptcy proceedings, this issue was resolved through negotiations in the shadow of the law, leaving behind no case law to help shape future data valuation inquiries. On one side, there was a note by the bankruptcy court examiner that properties of Caesar's that were sold off were worse off because they could not leverage the data of the rewards program—but at the same time, the examiner recognized that it would be difficult to sell the rewards program to other buyers.[69] Thus, the court never ultimately decided how to value the data in light of these complexities. This is common in bankruptcy, where few decisions come in the form of a bankruptcy court ruling.[70]

3. Synthesis

These case studies lead to a number of conclusions. First, courts do not always get to valuation questions. This may be because cases are decided on other grounds for legitimate reasons or because judges feel uncomfortable getting to the actual valuation and so they rule on different grounds to avoid the exercise. Second, there is uncertainty of valuation methodologies across areas of law, as well as potential for some such issues to simultaneously

64. Ginger Zhe Jin & Andrew Stivers, *Protecting Consumers in Privacy and Data Security: A Perspective of Information Economics* 1 n.2 (May 22, 2017) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3006172 [https://perma.cc/N3E3-4NGV].

65. *See generally* DANIEL J. SOLOVE & WOODROW HARTZOG, BREACHED! WHY DATA SECURITY LAW FAILS AND HOW TO IMPROVE IT (2022) (discussing the shortcomings of data privacy and privacy laws).

66. California Consumer Privacy Act of 2018, Cal. Civ. Code §§ 1798.100–199.100 (2018); California Privacy Rights Act, Cal. Civ. Code §§ 1798.100–199.100 (2018).

67. Perhaps this is a more sophisticated version of a customer list, which gets trade secret protection under the Defend Trade Secrets Act.

68. James E. Short & Steve Todd, *What's Your Data Worth?*, 58 MASS. INST. TECH. SLOAN MGMT. REV. 17, 17 (2017).

69. *Id.*

70. Douglas G. Baird & Robert K. Rasmussen, *The End of Bankruptcy*, 55 STAN. L. REV. 751, 786–88 (2002).

emerge in multiple contexts (for example, M&A and antitrust, M&A and bankruptcy, antitrust and data privacy) that may employ different methodologies. Accordingly, we believe that a more consistent approach may better facilitate business certainty with regard to valuation models.

III. REAL OPTIONS AS A SOLUTION

Real options analysis provides a framework that can be used to value data across different contexts, including different areas of law. We provide a basic introduction to real options before discussing the advantages and disadvantages of using them to value data. We then discuss how this approach might be employed in the real world.

A. REAL OPTIONS

An option is the right, but not the obligation, to do something. For example, if Maria has the right to paint her house green, to travel to Paris, or to order pizza for lunch, those are all options.

In finance, the most well-known options give their holders the right to buy or sell a specific quantity of a particular asset at a specified time for a specified price. These options are known as financial options.[71] For instance, Jacinta might have the right to buy 1,000 shares of Apple stock in three months' time at a price of $150 per share. That right would be quite valuable if, three months from now, Apple stock is trading at $200 per share: Jacinta could buy 1,000 Apple shares for $150,000,[72] then immediately sell them to other investors for $200,000,[73] netting her $50,000 of profit.[74]

Real options, like financial options, reflect the value of being able to react to changing conditions. However, rather than representing merely the right to buy or sell, they can encompass one's ability to change one's behavior in all manner of ways.[75] This ability to change course can be extremely valuable. A pair of simple, stylized examples help illustrate this point.

Example 1. Suppose that you are an executive at a company, and you are considering whether the company should launch a new product. It is unclear how consumers will react to the product; they may love it (iPods) or

71. *See Investment Products: Options*, FIN. INV. REGUL. AUTH., https://www.finra.org/investors/investing/investment-products/options [https://perma.cc/J6VN-7GPR] (last visited Aug. 28, 2023).
72. 1,000 shares * $150 purchase price per share = $150,000.
73. 1,000 shares * $200 sale price per share = $200,000.
74. $200,000 revenue from sale of Apple shares – $150,000 paid for Apple shares = $50,000 profit.
75. Real options are also called strategic options. IVO WELCH, CORPORATE FINANCE 363 (3rd ed. 2014).

they may not (Zunes). Suppose that there is a 50% chance that the product will be a success, in which case it will generate $10 million of profits per year for the next ten years.[76] On the other hand, there is a 50% chance that the product will be a commercial failure, in which case it will cost the company $20 million per year for the next ten years.

Under the facts of Example 1, the company should not launch the product.[77] Half of the time, the product will produce $100 million of profit;[78] the other half of the time it will produce losses of $200 million.[79] On average, then, launching the new product will cost the company $50 million.[80]

Example 2. The facts are the same as in Example 1, except that now the company has the ability to stop making the new product after its first year on the market.

Under the facts of Example 2, the company should absolutely launch the product. When the product is a success, it will keep the product on the market. Everything will remain the same in that circumstance, and the company will earn $100 million of profit. But when the product is a commercial failure, the company can now cut its losses after one year. By doing so, the company will reduce its total losses when the product fails from $200 million to only $20 million.[81] On average, the new product will now generate $40 million of profit.[82]

Taken together, Examples 1 and 2 show how valuable the ability to change course can be. Simply having the ability to give up on the product when it is not profitable transforms a project that loses $50 million into one that earns $40 million—a $90 million swing.[83] Since the only difference between these two Examples was the real option to give up on the product after a year, that option is worth $90 million.

Real options come in a variety of common forms. Companies can expand or contract their businesses, such as by opening new locations or

76. For conceptual clarity, and to avoid complicating the example with issues related to time value of money and discount rates, we assume that all of the payment values discussed in this example are present values—that is, the profit you will earn in year one (or two, or three, or seven, etc.) is worth $10 million to you today.

77. For simplicity, this analysis assumes that you are risk-neutral. If you were risk-averse, the case against the project would be even stronger.

78. $10 million in annual profits * 10 years = $100 million in total profits.

79. $20 million in annual losses * 10 years = $200 million in total losses.

80. 50% * $100 million + 50% * -$200 million = $50 million + -$100 million = -$50 million. Equivalently, the net present value (NPV) of this project is -$50 million.

81. The difference is between 1 year of $20 million annual losses and 10 such years.

82. 50% * $100 million + 50% * -$20 million = $50 million + -$10 million = $40 million. Equivalently, the NPV of this project is $40 million.

83. $50 million – -$40 million = $90 million.

closing existing facilities. They can accelerate or delay projects, such as by hiring more workers to build a factory or pausing construction. They can switch production processes, trade-off between workers and automated processes, or shift production between in-house divisions and outside contractors. Taken together, real options encompass a wide range of actions spread across an expansive set of possible circumstances.

B. Real Options as a Model for Data Valuation

As a framework for valuing data, real option analysis has many virtues. First, the value of data is that it enables a person to take new actions that were not available previously.[84] Real option analysis is how finance values the ability to take new courses of action. Thus, as a conceptual matter, real option analysis is a natural fit for valuing data. Further, real option analysis is a flexible and expansive tool that can be used to model an extraordinarily wide range of scenarios and circumstances. This makes it capable of handling the range of new possible outcomes that data, paired with modern statistical analysis, can produce.

Moreover, as noted previously, current approaches to data valuation offer little guidance. This increases the potential for confusion, inconsistency, and regulatory arbitrage. In some instances, they assign data no value at all.[85] Applying real options analysis to data valuation would help ameliorate all of these problems. Real options analysis gives a clear theoretical framework, providing guidance and structure for those trying to determine data's value. This would help align and unify the disparate valuation approaches that have been employed to date. Improved alignment would also reduce the opportunities for regulatory arbitrage that can result when different regulatory regimes adopt inconsistent valuation methodologies.[86]

84. This feature is not unique to data. For example, the value of lumber comes from what you can build with it, or what someone will give you in exchange for it—which depends on what they can build with it or what they can sell it for, and so on.

85. Interestingly, this parallels the most common mistake that managers make with respect to real options. WELCH, *supra* note 75, at 368. In some instances, holding data can have negative expected value, even accounting for the real options it creates. This could happen if the uses for the data generate little profit (for example, if legislation narrowly circumscribes their permitted uses), but the firm would suffer large costs if the data leaks, and the chance of a leak remains significant even after the firm takes precautions.

86. *See* Victor Fleischer, *Regulatory Arbitrage*, 89 TEX. L. REV. 227, 230 (2010) (describing regulatory regime arbitrage); *cf.* Jordan Barry, Response, *On Regulatory Arbitrage*, 89 TEX. L. REV. *SEE ALSO* 69, 73–78 (2010) (arguing that regulatory regime arbitrage is a subset of economic substance arbitrage, and that true regulatory arbitrage is only possible in that context when at least one of the regulatory regimes in question is using a regulatory rule that does not track the relevant underlying economic substance).

While real option valuation offers a number of benefits, it also entails a significant drawback: correctly valuing real options is quite difficult. To do so precisely, one must anticipate, and then think through, all of the possible future states of the world, their respective likelihoods of occurring, how one would respond to them all, and how much one would ultimately reap as a result. From there, one can work backwards from these endpoints to determine the right course of action at each decision point and the scenario's expected value overall. This is a tall order—especially when valuing data, an asset whose value depends in part on future developments in statistical analysis.

To put a somewhat finer point on it, consider financial options once more. Valuing financial options is a difficult mathematical problem. Fischer Black, Myron Scholes, and Robert Merton's options pricing model was a watershed advance for the field, ultimately garnering a Nobel Prize in 1997.[87] Even with the solution in hand, the mathematics remain challenging. As important as options are to modern finance, many undergraduate finance courses do not cover the application of their formula, let alone its derivation.[88]

Valuing real options is even harder than valuing financial ones. There are more possibilities to consider, more actions available, and more variables of interest.[89] It would be extremely difficult to write and apply a regulation with a precise formula that generalized across different types of data from diffuse contexts and industries. The complexity of real options also poses challenges for parties, for judges, and for juries.

This is a serious problem. A valuation method that has attractive theoretical properties, but that is impossible to apply in practice, would seem to be of extremely limited value.

87. Fischer Black & Myron Scholes, *The Pricing of Options and Corporate Liabilities*, 81 J. POL. ECON. 637, 640–45 (1973); Robert C. Merton, *Theory of Rational Option Pricing*, 4 BELL J. ECON. & MGMT. SCI. 141, 162–71 (1973); Press Release, The Nobel Prize, Royal Swedish Academy of Sciences, The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 1997 (Oct. 14, 1997), https://www.nobelprize.org/prizes/economic-sciences/1997/press-release [https://perma.cc/AP7W-9Z4H].

88. *See, e.g.*, A. Craig MacKinlay, The Wharton School, U. Pa., *Finance 1000: Corporate Finance* (2022), https://apps.wharton.upenn.edu/syllabi/202230/FNCE1000001 [https://perma.cc/YV4T-3U7H]; Albers Sch. Bus. & Econ., Seattle University, *FINC 3400 Business Finance & FINC 3420 Intermediate Corporate Finance*, https://www.seattleu.edu/business/undergraduate/courses--syllabi/finance [https://perma.cc/W5DD-9C6N] (last visited on Aug. 28, 2023).

89. *See, e.g.*, Tom Copeland & Peter Tufano, *A Real-World Way to Manage Real Options*, HARV. BUS. REV. (Mar. 2004), https://hbr.org/2004/03/a-real-world-way-to-manage-real-options [https://perma.cc/BJL8-TE64] ("As many executives point out, options embedded in management decisions are far more complex and ambiguous than financial options. Their concern is that it would be dangerous to try to reduce those complexities into standard option models, such as the Black-Scholes-Merton model, which have only five or six variables.").

C. A Way Forward

Despite its complexities, we nonetheless believe that real options analysis holds great promise as a framework for valuing data. If one wants to value data accurately, one must have the right model. In our view, real options analysis captures what makes data useful, and thus offers the best framework to think about data's value. If data's value is complicated and depends on many factors, then this is not a fault of the model; the model can only help a user identify and focus on the things that matter, even if that's a long list.[90] Put another way, to get the right answer, one must ask the right question. The right question may be a hard one—but answering a different, easier question means avoiding the problem, not solving it.

Moreover, it is worth stating what may be obvious: the real options approach need not be perfect to be an improvement over existing practices.[91] Getting all interested parties asking the right question—or even the same question—would be valuable. It would reduce conceptual confusion, inconsistencies, and opportunities for regulatory arbitrage. Moreover, real options always have positive value.[92] Whenever taking an available course of action is profitable, one can do so; if that course of action is not profitable, one can simply decline to take that action.[93] Real options analysis would underscore the point that data has value and thus should not be ignored.[94] These combined benefits may be considerable.

Furthermore, if decisionmakers use real options analysis to value data, they may find ways to ameliorate the complexity problems over time. Trial and error can produce insights. As agencies and courts experiment with the framework, approximations may arise that are easier to calculate. Even if these approximations are not precisely accurate, they may be close enough to be useful. In particular, they may be significant improvements over existing data valuation methods.

---

90. The complexity of real options may not be an entirely bad thing. For example, complexity in the valuation process may impede parties' ability to strategically manipulate valuations.

91. Harold Demsetz, *Information and Efficiency: Another Viewpoint*, 12 J.L. & ECON. 1, 1 (1969) (identifying the nirvana fallacy of a first-best comparative institutional analysis).

92. This is also true of financial options.

93. This assumes that actors are rational. If that is not the case, then it may be beneficial to remove some of one's choices, such as Odysseus tying himself to the mast to avoid being lured by the Sirens' song. HOMER, THE ODYSSEY (Emily R. Wilson trans., W.W. Norton & Co. 1st ed. 2018). It can also be valuable to remove options from your choice set if that will change others' behavior in a way that is favorable to you. *See, e.g.*, Deepak Malhotra, *Six Steps for Making Your Threat Credible*, HARV. BUS. SCH.: WORKING KNOWLEDGE (May 30, 2005), https://hbswk.hbs.edu/item/six-steps-for-making-your-threat-credible [https://perma.cc/J58N-D7AS] (describing how, when playing chicken, the best strategy is to remove your steering wheel and throw it out the window; that way, your adversary knows that you cannot swerve even if you wish to, and must then act accordingly). *See also supra* note 85 and accompanying text.

94. *Cf.* WELCH, *supra* note 75, at 368.

That dynamic—of finding heuristics that are simpler but informative—has been borne out in other settings. For example, basic corporate finance theory teaches that profit-maximizing firms should use net present value analysis to allocate their resources.[95] Yet many firms, including large, sophisticated ones, analyze other metrics as well.[96] These metrics include the profitability index, which measures how much profit a project generates per dollar invested, and the payback rule, which considers how long it takes for a project to repay its startup costs.[97] Both of these simple rules have well-known flaws that can cause them to produce absurd results.[98] Why, then, do they remain common?

One possible answer is that these simple rules produce information about projects' real option value. For example, recouping one's initial investment means that those recovered dollars can be redeployed toward other purposes, increasing the range of decisions available to the firm.[99] Researchers have found that, under a variety of circumstances, such simple rules can allow firms to make nearly optimal decisions.[100] The relative accuracy of these rules, combined with their simplicity, may explain why firms use them more frequently than real options analysis.[101] These types of

95. *See, e.g.*, *id.* at 61–66.

96. *See* John R. Graham, *Presidential Address: Corporate Finance and Reality*, 77 J. FIN. 1975, 2038 (2022) (surveying corporate managers on how they make capital allocation decisions and finding that, among large firms, 64% use the payback method and 39% use the profitability index); John R. Graham & Campbell R. Harvey, *The Theory and Practice of Corporate Finance: Evidence from the Field*, 60 J. FIN. ECON. 187, 199 (2001) (finding that 57% used the payback method, 30% used the discounted payback method, and 12% used the profitability index).

97. WELCH, *supra* note 75, at 75–78.

98. Profitability index can produce the wrong decision rules because firms seek to maximize their total profits, not their profits per dollar invested. For example, consider two mutually exclusive projects: Project A costs $100 and produces $1000 in revenue. Project B costs $1 and produces $100 in revenue. Both projects are good, but if one must choose between them, Project A is clearly better; its $900 in profit dwarfs Project B's $99 profit. Yet Project B has a much higher profitability index ($100 / $1 = 100) than Project A does ($1000 / $100 = 10). *Id.* at 75–76.

The payback rule evaluates projects based on how long they take to return their initial costs. Discounted payback does the same, but discounts the project's future cash flows to account for the fact that they do not come immediately. Both have the same problem; they ignore any cash flows that the project generates after it has paid back its initial costs. Consider project C, which costs $100 today and returns $110 in a year, and project D, which costs $100 today and returns $1000 in a year and a day. Project D is clearly a superior project, but the payback method will select Project C instead. *Id.* at 77.

99. This assumes that capital markets are imperfect, which is true of real-world markets. *See id.* at 511–39.

100. *See* Robert L. McDonald, *Real Options and Rules of Thumb in Capital Budgeting*, *in* PROJECT FLEXIBILITY, AGENCY, AND COMPETITION 13 (M.J. Brennan & L. Trigeorgis eds., 2000); Achim Wambach, *Payback Criterion, Hurdle Rates and the Gain of Waiting*, 9 INT'L REV. FIN. ANALYSIS 247, 257 (2000); Glenn W. Boyle & Graeme A. Gutherie, *Payback and the Value of Waiting to Invest* 13–14 (Apr. 29, 1997) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=74 [https://perma.cc/8K39-B95L].

101. *See* Graham, *supra* note 96, at 1985 (finding that only 38% of large firms frequently use real options in decision-making, which was less frequent than profitability index (39%) or payback rule

heuristics, and others, may prove useful to valuing data.

Alternatively, real option analysis can inform other modes of valuation. One response to complicated valuation problems is the method of comparables: To determine an item's value, identify similar items whose values are known (that is, comparables), then make appropriate adjustments. This method is frequently employed to value items such as real estate, art, and active businesses.[102] Under the right circumstances, this method can produce accurate valuations.

The method of comparables can be tricky to apply to data for several reasons. First, it may be difficult to identify similar data sets with known values. Sale prices are often used as the measure of value for comparable items, and sale prices for data may not be public. But even when sale prices are available, data sets can differ from each other along a variety of dimensions. Which of those differences are important, and how much should value estimates be adjusted to account for these differences? For example, which is more valuable—a data set that is twice as large, or one that includes data drawn from twice as much time? Is data more valuable when the future is more uncertain or less? These are but a few of the dimensions one might wish to consider.

Real option theory sheds insight into some of these questions. It identifies a number of factors that directly affect real option value, and thus the value of data. These factors can then be considered and adjusted for when using comparables to value data.

One factor that informs a data set's value is its informational uniqueness. To what extent does that data tell its user something that they otherwise would not know? Having insights that no one else has can be extremely valuable. On the other hand, when competitors have access to comparably informative data, profitably exploiting the data gets harder, as competition among firms puts the firm's counterparties in a comparatively stronger position.

Two other factors stem from the payoffs available from exploiting data. Unsurprisingly, the higher the potential future profits that the data can unlock, the more valuable the data is. What is less obvious is that the value of data increases as the future becomes less certain. This is somewhat

(64%)); *see also* Graham & Harvey, *supra* note 96, at 188 (finding that payback rule was more commonly used than real options); H. Kent Baker, Shantanu Dutta & Samir Saadi, *Management Views on Real Options in Capital Budgeting*, 21 J. APPLIED FIN. 1, 8 (2011) (surveying Canadian firms and finding that only 10% often or always used real options analysis when deciding among projects, while 67% used the payback rule, 25% used the discounted payback rule, and 11% used the profitability index).

102. WELCH, *supra* note 75, at 431–36.

abnormal; in finance, safer cash flows are usually considered more valuable than riskier ones.[103] Options are an important exception to this general rule, however. Because options allow one to change behavior in response to different circumstances, they actually become more valuable when a project has a wider range of possible future payouts.[104]

Another important factor in real option valuation is the length of time over which one can continue to change one's behavior.[105] The longer that one can change direction, the more actions that one has available, and the more valuable the option. In the data context, this corresponds to the useful life of the data. As noted earlier, some data remains useful and informative for years or even decades; other data grows stale quickly.[106] All else equal, the former is more useful than the latter.[107]

Relatedly, interest rates affect the value of real options, and thus of data.[108] Profits earned in the future are more valuable when interest rates are low than when rates are high.[109] Interest rates have more of an effect on data with a longer useful life, and less of an effect on shorter-lived data.

How quickly and cheaply one can change one's behavior also affects a real option's value. The quicker one can act, the more nimble one is, the more ways in which one can profitably change one's behavior. Similarly, options that can be exercised at little cost are more valuable than those which are expensive to utilize.[110]

These factors are more amenable to forming legal standards than a strict formula for valuing real options would be. Accordingly, they may provide a path forward for data valuation.

Finally, real options theory could inform attempts to value data in a different way. Experience may convince policymakers that valuing data is simply too hard, and that they should act accordingly. Such actions could

103. *Id.* at 124, 197.

104. *Id.* at 364.

105. This is also an important factor in financial option valuation. *See generally* Merton, *supra* note 87.

106. Of course, distinguishing one from the other may be challenging in particular cases. The task gets easier when one at least knows to ask the question, however.

107. This factor relates to the first. If the data is informationally unique, or more unique, for a longer period of time, the firm possessing that data will have more attractive choices available to it for a longer period of time (that is, a longer-lived option).

108. This is also true of financial options. *See generally* Merton, *supra* note 87.

109. More precisely, firms should care about the discount rate they apply to future cash flows rather than about interest rates, but the two concepts are similar. In practice, the latter is easier to observe and may closely correlate with the former.

110. This is analogous to the strike price for a financial call option; all else equal, options with lower strike prices are more valuable.

take multiple forms.

One response to a difficult valuation problem is to simply exit the field as much as possible. Section 83 of the Internal Revenue Code provides a good example of this approach.[111] It addresses the questions of how much income a taxpayer has when they receive property in exchange for performing services, and when the taxpayer is taxed on that income. Section 83's general rule is that employees are taxed on property based on its fair market value, and they are taxed at the time it becomes clear that they will get to keep the property.

For example, startup companies frequently include some form of equity interest in the company as part of their employees' compensation packages.[112] These interests can come in various forms, including stock, restricted stock units, or stock options.[113] If employees leave their employer before a certain date—if they quit to take a new job or are fired—then they forfeit some or all of their equity interests. The date after which an employee gets to keep an equity interest, even if the employee leaves the firm, is known as that interest's vesting date. If an employee leaves the employer before the vesting date, they lose their unvested equity.

Under the general rule of Section 83, an employee is typically taxed on the value of their equity interest at the time those interests vest.[114] However, as noted previously, valuing stock options is difficult. Accordingly, Section 83 exempts stock options from its general rule—unless they have a visible market price (in which case they are easy to value).[115] Instead, employees who receive stock options generally are not taxed until they exercise those options, at which point they receive stock in their employer, which is easier to value.[116] This limits taxpayers' ability to take aggressive valuations of hard-to-value stock options.[117] Regulators can adopt similar tactics in the

111. 26 U.S.C. § 83 (2023).

112. *See, e.g.*, Abraham J.B. Cable, *Fool's Gold? Equity Compensation & the Mature Startup*, 11 VA. L. & BUS. REV. 613, 613 (2017).

113. *Id.*

114. 26 U.S.C. § 83 (2023).

115. 26 U.S.C. § 83(e) (2023); Treas. Reg. § 1.83–7(b) (as amended in 2004). Stock options can also have a readily ascertainable fair market value if they are not actively traded, but this is unusual; the relevant regulations recognize that the possibility of future price changes increases the value of an option and requires (among other conditions) that this component of value be measurable with reasonable accuracy. Treas. Reg. § 1.83–7(b)(2), (3) (as amended in 2004).

116. This assumes that the stock is vested. The general rule of Section 83 applies to the stock; if the employee may have to surrender the stock to the employer in the future if they do not continue their employment past a specified date, then the employee is not taxed on the value of the stock until the stock vests.

117. For example, absent these rules, an employee could assign a low value to a stock option, thereby recognizing little ordinary income at the time of the grant. They would then recognize greater gains on the eventual sale of their stock, but those gains would generally be long-term capital gains and

context of data valuation.

A potentially complementary approach would be to foster a market for data, with standardized features, in order to make private transaction prices more visible and data sets more easily comparable. In a number of instances, legislative and regulatory interventions have helped shift markets characterized by bespoke arrangements toward more commoditized features and greater transparency.[118] Such standardized markets can make the job of valuation much easier, and can also protect unsophisticated parties operating in those markets.[119]

CONCLUSION

While data has become increasingly valuable and important, the law's attempts to value data have lagged, remaining confused and underdeveloped. Situating data valuation law within an economic framework built on real options analysis would resolve conceptual confusion among courts, agencies, and legislatures. It would also create greater predictability among private actors, which in turn would reduce the risk of regulatory uncertainty and facilitate investment. A clearer legal approach that cuts across different areas of law and jurisdictions would limit opportunities for regulatory arbitrage across fields of law addressing data valuation. Furthermore, a consistent approach reduces politicization of results, preventing favored groups from shifting unclear legal rules in their favor when there is no economic basis for such a shift. A consistent approach also makes decision-making less opaque, thereby increasing the legitimacy of outcomes.

While the real options approach is not without potential problems, we believe that it is the least bad alternative available. Moreover, increased use of real options analysis over time may generate heuristics that simplify data valuation by courts and agencies. These heuristics may prove so effective that private parties incorporate them into arm's length transactions. Further research is needed to identify what heuristics work best in the data valuation context, as well as how to encourage more transparent and comparable pricing in burgeoning data markets worldwide.

would be subject to a significantly lower tax rate. Because options are hard to value, it could be difficult for the IRS to prove that the employee's valuation was too low.

118. Financial derivatives provide a useful recent example. *See* Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111–203, §§ 701–774, 124 Stat. 1376 (2010).

119. *See* BURTON G. MALKIEL, A RANDOM WALK DOWN WALL STREET: THE TIME-TESTED STRATEGY FOR SUCCESSFUL INVESTING 26 (2015) ("Taken to its logical extreme, it means that a blindfolded monkey throwing darts at the stock listings could select a portfolio that would do just as well as one selected by experts.").