TAKING SCALE SERIOUSLY IN TECHNOLOGY LAW

Mark P. McKenna [*] and Woodrow Hartzog [**]

Issues of scale—the relationship between the amount of an activity and its associated costs and benefits—permeate discussions around law and technologies. Indeed, it's not much of an exaggeration to say that scale is the reason for most technology regulation.

But it's not always clear how lawmakers and judges conceptualize "scale" when approaching questions around automated technologies. Scale is often used intuitively, just to mean "more." But scale is not always just about more—scale can introduce new harms and benefits along different dimensions, not simply costs or efficiencies of greater magnitude.

In this Article, we argue for a more sustained interrogation of the role of scale in law, one that is more sensitive to the distinction between what we describe as "scale is more" and "scale is different." When lawmakers and judges fail to properly categorize the role of scale in a particular context, they risk ignoring or misidentifying harms, misdiagnosing the causes of those harms, and potentially focusing on the wrong policy tools, and even the wrong actors, in proposing solutions.

---

[*] Professor of Law, UCLA School of Law; Faculty Co-Director, UCLA Institute for Technology, Law & Policy; Of Counsel, Lex Lumina PLLC.

[**] Professor of Law, Boston University School of Law. The authors would like to thank Kabbas Azhar, Quincy Cason, Susan Hong, Benjamin Kline, and Janelle Robins for their research assistance.

TAKING SCALE SERIOUSLY IN TECHNOLOGY LAW

INTRODUCTION

Dozens of currently-pending lawsuits deal with the copyright implications of generative artificial intelligence (AI).[1] Those cases raise several different issues, but they have in common the question of whether it is fair use to train AI models on copyrighted works. Among other arguments, plaintiffs in these cases insist that the training uses are not fair because the AI developers could license use of the copyrighted works for training purposes. The feasibility of a licensing market depends, however, not just on whether the copyright owners stand willing to license, but also on the scale of licensing that would be necessary to train these systems effectively. Frontier generative AI models require a truly massive amount

[1] Kadrey v. Meta; Bartz v. Anthropic; Open AI cases

of training data—on the order of many *trillions* of tokens.[2] These systems learn by ingesting massive amounts of data and extracting patterns—patterns that do not emerge at millions or even hundreds of millions of tokens. The point here is not simply that training gets better with more data; it's that the training is basically worthless until some massive scale is achieved. Of course, the fact that these systems can't be trained effectively without using that much data does not fully determine fair use. Nor does it tell us that these generative AI systems are, on the whole, technologies that we want to enable. But to the extent the feasibility of licensing markets is relevant to the fair use question, decisionmakers have to reckon with the true effects of scale in this context.

Palantir advertises itself as providing "AI Powered Automation for Every Decision."[3] According to its website, Palantir's "software powers real-time, AI-driven decisions in critical government and commercial enterprises in the West, from the factory floors to the front lines."[4] How does it do that? By taking "big data analytics" to the next level. Palantir collects and combines massive amounts of data and learns patterns and connections that don't appear without that scale. Understanding the importance of scale to Palantir's model puts the government's plans to enable cross-departmental data access in a whole new light.

These are just two particularly salient current examples where scale is central to understanding the human values implicated by technology, and potential regulatory interventions. There are many other examples. Indeed, considerations of scale—the relationship between the amount of an activity and its associated costs or benefits—are everywhere in technology law and policy.[5] It's barely an exaggeration to say that scalability is *the*

---

[2] See, e.g., Brian Wang, *Deep Dive on DeepSeek and AI*, NEXT BIG FUTURE (Feb. 5, 2025), https://www.nextbigfuture.com/2025/02/deep-dive-on-deepseek-and-ai.html. "Tokens" are units of data processed by AI models during training. See Explaining Tokens—the Language and Currency of AI (https://blogs.nvidia.com/blog/ai-tokens-explained/) ("Tokens are tiny units of data that come from breaking down bigger chunks of information.").

[3] See Palantir.com, https://www.palantir.com/ ("AI-Powered Automation for Every Decision") (last visited August 13, 2025).

[4] *Id.*

[5] See, e.g., Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CAL. L. REV. 513, 538 (2015) ("Today's robots do a variety of tasks that people could do, but don't for reasons

reason lawmakers are so concerned about new technologies.[6] If technologies didn't scale and only affected a few people, then they would be less worthy of categorical regulatory attention. But despite the ubiquity of scale issues, lawmakers and judges usually deal with the concept intuitively—and primarily just to mean "more" of something.

---

of cost or preference. Moving more tasks into the category of automation could in and of itself cause legal issues at scale."); Julie Cohen, *How (Not) to Write a Privacy Law*, KNIGHT FIRST AMEND. INST. (Mar. 23, 2021), https://s3.amazonaws.com/kfai-documents/documents/306f33954a/3.23.2021-Cohen.pdf ("Current approaches to crafting privacy legislation are heavily influenced by the antiquated private law ideal of bottom-up governance via assertion of individual rights, and that approach, in turn, systematically undermines prospects for effective governance of networked processes that operate at scale.…[Arguments for user-governed data cooperatives] tend to ignore important qualifications that affect the ability of common-governance arrangements to scale.… Both arguments for bottom-up governance flowing from assertion of individual rights and arguments for commons-based cooperative governance of personal data collection and processing overlook the structural and temporal effects of design operating at scale.… the dysfunctions of the networked information economy reflect underlying problems of networked flow and scale that are distinct from existing patterns of market domination.… To be effective at all, regimes for privacy governance need to target order of magnitude problems in ways that enable oversight and enforcement to scale up and out commensurately."), Sarah Ciston, *A Critical Field Guide For Working With Machine Learning Datasets*, KNOWING MACHINES PROJECT (2022), https://knowingmachines.org/critical-field-guide ("The speed and scale of machine learning and massive datasets make "discrimination easier, faster, and even harder to challenge.… Whether designing a dataset from scratch or using one that has been around for years, decisions made at every step will inform your project outcomes. These decisions get scaled and compounded by machine learning models.") (citing RUHA R. BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE (2019)); Mike Ananny & Kate Crawford, *Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, 20 NEW MEDIA & SOC'Y 973 ("Sometimes, the details of a system will be not only protected by corporate secrecy or indecipherable to those without technical skills, but inscrutable even to its creators because of the scale and speed of its design."). Tech companies also regularly refer to issues of scale. *See Facebook, Social Media Privacy, and the Use and Abuse of Data: Hearing Before the S. Comm. On the Judiciary & S. Comm. on Com., Sci., & Transp.*, 115th Cong. (10 Apr. 2018), www.commerce.senate.gov/2018/4/facebook-social-media-privacy-and-the-use-and-abuse-of-data (quoting Mark Zuckerburg saying "We have gotten increasingly better at finding and disabling fake accounts.…This is thanks to improvements in machine learning and artificial intelligence, which can proactively identify suspicious behavior at a scale that was not possible before—without needing to look at the content itself.)

[6] See, e.g., Julie Cohen, *Platforms, Data Infrastructures, and Infrastructure Stacks*, GLOBAL GOVERNANCE BY DATA: INFRASTRUCTURES OF ALGORITHMIC RULE (Fleur Johns, Gavin Sullivan & Dimitri Van Den Meerssche, eds. 2024) ("Data infrastructures (are designed to) scale both vertically and laterally; they are nimble, flexible, and adaptable to new uses.…We will see that the emergence of data infrastructures has both enabled control and coordination of formerly distinct activities and afforded points of entry for new assertions of political and geopolitical power.").

Scale as "more" is often significant. In considering legal intervention, courts and regulators frequently compare the magnitude of potential harms associated with an activity with the costs of preventing that harm.[7] Higher levels of an activity often produce more harm, and so the amount of an activity matters to these calculations and can tip the balance in favor of legal liability or in favor of regulation. Similarly, more use of a legal tool might at some point change the incentives for affected parties and shift their behavior in ways that have offsetting costs. In the context of notice and takedown regimes like the Digital Millennium Copyright Act, for example, platforms might receive so many notices of infringement that dealing with each notice individually becomes enormously costly and burdensome. As a result, those platforms have strong incentive to create automated systems like Content ID—systems that might deal much more efficiently with notices of infringement but also block some legitimate uses copyrighted content.[8]

But sometimes scale does not just mean more—it does not relate simply to harms or benefits that are greater in magnitude, or even to situations where the increased magnitude creates a tipping point in our cost/benefit calculation. Sometimes scale creates entirely new effects along different dimensions. The dominance of social media platforms and "network effects" is a good example. At a certain scale, the size and popularity of a social media platform makes it meaningfully different, both in its perceived value to users and in the business models it enables. This is the point where people are attracted to a platform not primarily because of the platform's discrete features, but because of who else is there. Importantly, scale in this context creates a new collective action problem—social media platforms are more valuable precisely because of other users' presence on the platform, making it harder to switch from one platform to

[7] The most obvious example here would be the Hand Formula, which requires courts to compare the product of the probability of harm from some conduct (P) and the magnitude of the loss produced by that conduct (L) with the burden of preventing the harm (B). When PxL>B, it is negligent not to take the precaution that would prevent the harm. U.S. v. Carroll Towing, 159 F.2d 169 (2d Cir. 1947).

[8] Katharine Trendacosta & Corynne McSherry, *What Really Does and Doesn't Work for Fair Use in the DMCA*, EFF (July 31, 2020), https://www.eff.org/deeplinks/2020/07/what-really-does-and-doesnt-work-fair-use-dmca.

another. We have seen this phenomenon in real time with users' attempts to find replacements for Twitter (X).

Another example is the scale of IoT doorbells like Amazon's popular Ring cameras. IoT doorbells were first designed to provide a simple video feed of the area right in front of the door. Now they are being outfitted with AI-powered facial recognition and anomaly-recognition technologies and have a range of 1.5 miles.[9] If these always-on doorbells become fully outfitted at even a relatively modest scale, they could change the privacy risk from localized identification upon approaching a doorstep to ensuring that there were few public spaces left where anyone could remain anonymous.

In this Article, we argue for a more sustained interrogation of the role of scale in technology law, one that is more sensitive to the distinction between what we describe as "scale is more" and "scale is different." That distinction is particularly crucial in the context of data and artificial intelligence. When lawmakers and judges fail to properly consider the role of scale in a particular context, they risk ignoring or misidentifying harms, misdiagnosing the causes of those harms, and focusing on the wrong policy tools, and even the wrong actors, in proposing solutions.

In our terminology, "scale as more" refers to situations in which the effects of an activity increase in some relationship with the amount of that activity.[10] That relationship may not be linear, and it might not even be constant over time. But the point is that these are contexts where the same *types* of effects increase in relation to the amount of the activity. If an act causes x units of a particular kind of harm, then the total amount of that harm caused by the activity is some function of the number of instances of the activity. Again, that kind of scale is often important, because the

---

[9] Michael Brown, *The new Abode Edge Camera boasts 1.5-mile transmission range*, TECHHIVE (Jan. 12, 2024), https://www.techhive.com/article/2199931/the-new-abode-edge-camera-boasts-1-5-mile-transmission-range.html; Shira Ovide, *Amazon's Ring Plans to Scan Every Face at the Door*, WASHINGTON POST (Oct. 3, 2025), https://www.washingtonpost.com/technology/2025/10/03/amazon-ring-doorbell-facial-recognition-pricacy/.

[10] *See* Adrian Bridgwater, *What Is Technology Scalability?,* FORBES (19 Feb. 2020), www.forbes.com/sites/adrianbridgwater/2020/02/19/what-is-technology-scalability/?sh=9181da04f3f0 ("[W[e can probably assume that scalability in the IT platform and application sense refers to scaling upwards, to make a piece of technology bigger and more expansive.").

aggregate harm of an activity must be compared to its aggregate benefits and/or the costs of mitigating that harm. And that means it will often be important to understand the relationship between activity and harm, and how harms increase in relation to the amount of the activity.

But when "scale is different," it's not just that the effects increase as some function of activity; it's that *different types* of effects might emerge. Scale in this sense is not just about a change magnitude along the same dimensions; scale adds new dimensions. In this sense, scale is "n dimensional."[11] Importantly, when we use "scale is different," we are referring to the effects of the activity itself, not simply to our legal or regulatory response. There are many cases in which the scale of an activity reaches a tipping point where costs exceed benefits. Just because the legal response is different doesn't mean that the types of effects are different. What we mean to differentiate here are the cases where the *kinds of effects* change or emerge as the amount of the activity increases.

Scale in this sense might change the equation in at least four possible ways that should cause lawmakers and judges to think of the problem differently.

First, the population affected could change. For example, the data collected for machine learning doesn't just affect each individual whose data is collected in the sense that each suffers an individuated harm that we can simply multiply by the number of people whose data is used. At scale, that data provides population-level insights that can be used against different people within the same category, and different categories of people.[12]

Second, the scale of an activity can *create new problems* that didn't exist in small numbers. For example, if racial bias becomes encoded in all automated systems, individual instances of wrongful discrimination at scale can have the effect of shutting people out of entire career options and other important life decisions.[13] Likewise, the number of sidewalk robots in

---

[11] We thank Julie Cohen for this terminology.

[12] *See* Salomé Viljoen, *A Relational Theory of Data Governance*, 131 YALE L.J. 573, 577 (2021).

[13] *See* Katherine Creel & Deborah Hellman, *The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision-Making Systems*, 52 CAN. J. PHIL. 26 (2022).

use might fundamentally change the physical landscape: those robots might be annoying in small numbers but at scale can clog up sidewalks so much the sidewalks become unusable.

Third, the scale of activity *can challenge original assumptions about the costs and benefits of an activity*. Manipulation via dark patterns might always be wrongful, but the harms might seem *de minimis* when viewed from the perspective of individual users. Scale can make the nature of the problem more apparent. Using automated scraping tools to collect "publicly available" data to train machine learning systems might seem functionally equivalent to a person simply reading and writing down information that anyone could access if they were given the right link. But most of that information wouldn't be aggregated without the scraping tools because of the time and expense that would be required. The tools enable collection of information that otherwise would have been functionally obscure.[14]

Finally, scale can *affect the efficacy of solutions*, making certain institutional designs more effective and taking some legal, social, design, and market-based remedies and strategies entirely off the table.[15] When it comes to legal remedies, sometimes scale is "different" in that lawmakers cannot assume that expanding a certain type of enforcement or ratcheting up remedies would produce a proportionate increase in efficacy. For example, the automated nature of misinformation makes private lawsuits

---

[14] *See* Woodrow Hartzog & Evan Selinger, *Surveillance as Loss of Obscurity*, 72 WASH. & LEE L. REV. 1343, 1345–46 (2015) ("[W]e argue that the concept of "obscurity," which deals with the transaction costs involved in finding or understanding information, is the key to understanding and uniting modern debates about government surveillance."); Woodrow Hartzog & Evan Selinger, *Increasing the Transaction Costs of Harassment*, 95 B.U. L. REV. ANNEX 47 (2015); Evan Selinger & Woodrow Hartzog, *Obscurity and Privacy*, *in* SPACES FOR THE FUTURE: ROUTLEDGE COMPANION TO PHILOSOPHY OF TECHNOLOGY (Joseph Pitt & Ashley Shew eds., 2018), https://www.routledge.com/Spaces-for-the-Future-A-Companion-to-Philosophy-of-Technology/Pitt-Shew/p/book/9780415842969; *see also* Woodrow Hartzog & Frederic Stutzman, *The Case for Online Obscurity*, 101 CALIF. L. REV. 1, 5 (2013) ("We argue the case for obscurity for two reasons. First, we argue that obscurity is a common and natural condition of interaction, and therefore human expectation of obscurity will transfer to the domains in which we spend time, both physical and virtual. Second, we argue that obscurity is a desirable state because we are protected by an observer's inability to comprehend our actions, and therefore social practice encourages us to seek obscurity."); Woodrow Hartzog & Frederic Stutzman, *Obscurity by Design*, 88 WASH. L. REV. 385 (2013).

[15] *See generally*, Ryan Calo, *Code, Notice, or Nudge?*, 99 IOWA L. REV. 773 (2014).

to remedy individual instances of deception seem futile. Likewise, when scale is different, private enforcement mechanisms that rely predominantly on compensating individual harms are unlikely to address the systemic or structural harms that may only emerge beyond certain thresholds.

We conclude this article with a reflection on the nature of scale in existing legal frameworks for automated technologies and a call for more regulator nuance. We emphasize that more nuanced engagement with scale is not necessarily an argument for more regulation. Sometimes thoughtful consideration of the ways scale matters will have more to do with *how* we regulate than whether we do. Other times an appreciation of the ways scale is different might suggest *less* need for regulation. In some cases, legal intervention is needed to prevent or remedy harms that are the result of outlier behavior. When someone acts outside the norm, that party might cause unique harms that affected parties are not well situated to avoid. But when that same activity becomes the norm, there might be incentive for technological or legal adaptation that makes it less likely the individual harms will be visited in the same way. Those are cases where the scale of an activity changes our assessment of the harm caused by that activity because the scale produces (or is likely to produce) responsive measures that wouldn't exist at lower levels.

I. SCALE IS AN UNDERSPECIFIED CONCEPT

"Scale" is widely invoked in conversations about technology and its governance. Yet, remarkably, that concept is rarely explicitly defined or clarified when used in consequential settings. The Merriam-Webster dictionary defines scale as "something graduated especially when used as a measure or rule," "a graduated series or scheme of rank or order," and "a proportion between two sets of dimensions….a distinctive relative size, extent, or degree."[16]

Within the natural and social sciences, "scale" typically refers to "the spatial or temporal dimension of a phenomenon, and scaling is the transfer of information between scales."[17] Scientists often identify space,

---

[16] *Scale,* Merriam-Webster Dictionary, https://www.merriam-webster.com/dictionary/scale (last accessed Aug. 7, 2023).

[17] Jianguo Wu & Harbin Li, *Concepts of Scale and Scaling, in Scaling and Uncertainty Analysis*, *in* ECOLOGY: METHODS AND APPLICATIONS 3 (2006).

time, and organizational level as dimensions or kinds of scale.[18] In statistics, "scaling usually refers to a set of techniques for data reduction and detection of underlying relationships between variables."[19] Ecologists use scaling to predict and understand.[20] In technological circles, scalability is often conceptualized as "the capability of a system, network, or process to handle a growing amount of work, or its potential to be enlarged in order to accommodate that growth."[21] Economists often discuss "economies of scale," whereby an average cost falls as output increases.[22]

---

[18] *Id.* at 5 ("Space and time are the two fundamental axes of scale, whereas organizational hierarchies are usually constructed by the observer.").

[19] *Id.* at 9-10 ("In physical sciences, scaling usually refers to the study of how the structure and behavior of a system vary with its size, and this often amounts to the derivation of a power-law relationship. This notion of scaling has often been related to the concepts of similarity, fractals, or scale-invariance, all of which are associated with power laws. For example, a phenomenon or process is said to exhibit "scaling" if it does not have any characteristic length scale; that is, its behavior is independent of scale – i.e., a power law relationship."). To poorly paraphrase (and with apologies to statisticians), something is scalable where a relative change in one dimension results in a relative proportional change in the other dimension, independent of the initial aspects of those dimensions.

[20] *Id.*

[21] *Scalability*, NETWORK SECURITY, https://www.networxsecurity.org/members-area/glossary/s/scalability.html (last accessed Sept. 16, 2023) ("For example, [scalability] can refer to the capability of a system to increase its total output under an increased load when resources (typically hardware) are added. An analogous meaning is implied when the word is used in an economic context, where scalability of a company implies that the underlying business model offers the potential for economic growth within the company. Scalability, as a property of systems, is generally difficult to define and in any particular case it is necessary to define the specific requirements for scalability on those dimensions that are deemed important. It is a highly significant issue in electronics systems, databases, routers, and networking. A system whose performance improves after adding hardware, proportionally to the capacity added, is said to be a scalable system. An algorithm, design, networking protocol, program, or other system is said to scale if it is suitably efficient and practical when applied to large situations (e.g. a large input data set, a large number of outputs or users, or a large number of participating nodes in the case of a distributed system). If the design or system fails when a quantity increases, it does not scale….Scalability refers to the ability of a site to increase in size as demand warrants. The concept of scalability is desirable in technology as well as business settings. The base concept is consistent – the ability for a business or technology to accept increased volume without impacting the contribution margin (= revenue ? variable costs). For example, a given piece of equipment may have a capacity for 1–1000 users, while beyond 1000 users additional equipment is needed or performance will decline (variable costs will increase and reduce contribution margin.").

[22] *Economies of Scale*, SCIENCEDIRECT, https://www.sciencedirect.com/topics/social-sciences/economies-of-scale (last accessed Sept. 16, 2023).

Parties in litigation commonly use the language of scale, suggesting that something is "scaling up" or, conversely, "doesn't scale." In all of these cases, courts invoke scale in very general terms, referring in some way to the magnitude of some activity. For example, the court in a case alleging fraud over a tech company's products cited affidavits asserting that the defendant "is not currently competitive on large-scale parallel systems, as Sybase's database does not scale well past four CPUs."[23] In patent litigation, a judge wrote that the "[d]efendant was arguing that while the processes were successful for 'one-off tablets', a POSA would have sought a process that could be *scaled up*. Plaintiffs d[id] not make a plausible argument that a POSA would not want to develop a *scalable process*. Plaintiffs also d[id] not make a plausible argument that a POSA would have [had] options other than modifying Bartholomaus and McGinity if they wanted to produce hardened tablets *at scale*."[24]

In securities litigation, a judge wrote that "Talis also sought to capitalize on a rapidly closing window to sell a new COVID-19 test before demand cooled due to the FDA's approval of the Pfizer and Moderna COVID-19 vaccines in December 2020, and before competing tests captured the market. Talis would need to persuade investors that its product provided fast, accurate, reliable results and could be manufactured *at scale*."[25] In a lawsuit over allegedly fraudulent statements regarding Novavax's production of the COVID-19 vaccine, the court cited an executive's statement, "'We appear to have got past (certain) supply issues and are now being able to produce *at scale*.'"[26]

---

[23] In re Sybase, Inc. Sec. Litig., 48 F. Supp. 2d 958, 962 (N.D. Cal. 1999); *see also*, In re Cloudera, Inc. Sec. Litig., No. 19-CV-03221-MMC, 2022 WL 14813896, at *14 (N.D. Cal. Oct. 25, 2022) ("Cloudera's offerings provided "[c]loud and on-premises deployment at scale and across hybrid cloud environments[.]"); Indiana Pub. Ret. Sys. v. Pluralsight, Inc., No. 119CV00128JNPDBP, 2021 WL 1222290, at *9 (D. Utah Mar. 31, 2021), *aff'd in part, rev'd in part and remanded*, 45 F.4th 1236 (10th Cir. 2022) ("At the time, we had about 80 quota-bearing reps and little infrastructure around our sales reps....None of that infrastructure really existed at scale.").

[24] Purdue Pharma L.P. v. Accord Healthcare Inc., No. CV 20-1362-RGA, 2023 WL 2894939, at *6 (D. Del. Apr. 11, 2023).

[25] In re Talis Biomedical Corp. Sec. Litig., No. 22-CV-00105-SI, 2022 WL 17551984, at *2 (N.D. Cal. Dec. 9, 2022) (emphasis added).

[26] Sinnathurai v. Novavax, Inc., No. CV TDC-21-2910, 2022 WL 17585715, at *8 (D. Md. Dec. 12, 2022) (emphasis added).

Legal scholars have also invoked concepts of scale in their attempts to explain various legal doctrines. According to Richard Epstein, "the doctrine of efficient breach does not 'scale' as the number of parties increases."[27] Jonthan Adler wrote that the Clear Air Act's core provisions that "focus on locally concentrated pollutants and a cooperative federalism model…[do] not scale cleanly to the control of a ubiquitous and globally dispersed pollutant such as carbon dioxide."[28] Francis Fukuyama wrote about the creation of norms and values that support legal enforcement that "[s]pontaneous order does not scale well: the larger the group size, the lower the likelihood that free riders will be detected or punished."[29]

Law and tech scholars have frequently used the language of scale to describe problems related to the extent of an activity and that activity's costs or harms.[30] For example, David Post wrote regarding the growth of

---

[27] Richard A. Epstein, *Common Ground: How Intellectual Property Unites Creators and Innovators*, 22 GEO. MASON L. REV. 805, 815 (2015); Nicolas P. Terry, *The Opioid Litigation Unicorn*, 70 S.C. L. REV. 637, 667 (2019) ("Unfortunately, litigation is a blunt instrument that--to the extent it is effective at all--is best suited to well prescribed, narrow claims between individuals or between an individual and a corporation. Litigation *does not scale well*, and it is not a good tool for remedying mass social ills. It is also extremely inefficient both in its procedural costs (including attorneys' fees and other expenses) and a lack of timely resolution that almost guarantees that any recovery will be too late to help those who are currently suffering.") (emphasis added); Benjamin Ewing, *The Structure of Tort Law, Revisited: The Problem of Corporate Responsibility*, 8 J. TORT L. 1, 7 (2015) ("[I]t begins to look unfair that tort law *does not scale* the extent of tortfeasors' liability to their degree of culpability or to the foreseeable extent of the harm they cause. Although in negligence law defendants are generally liable only for categories of harm that were reasonably foreseeable, under the so-called "egg-shell skull rule" they are liable for the full extent of a reasonably foreseeable harm they cause, even if the extent of the harm far exceeds normal expectations because of a hidden and unusual vulnerability in the victim.") (emphasis added).

[28] Jonathan H. Adler, *The Environmental Protection Agency Turns Fifty*, 70 CASE W. RES. L. REV. 871, 876 (2020).

[29] Francis Fukuyama, *Differing Disciplinary Perspectives on the Origins of Trust*, 81 B.U. L. REV. 479, 490 (2001). Scholars have even referenced the concept of scale when criticizing The Bluebook, writing "that the core problem with The Bluebook is that it is unwieldy. It still applies a twentieth-century method in a much larger, twenty-first century world. What worked for The Bluebook with twenty-six pages in 1926 *does not scale* well to its current 511 pages and beyond." Stephen M. Darrow & Jonathan J. Darrow, *Beating the Bluebook Blues: A Response to Judge Posner*, 109 MICH. L. REV. FIRST IMPRESSIONS 92, 95 (2011) (citing Richard A. Posner, *The Bluebook Blues*, 120 YALE L.J. 850, 859 (2011) (emphasis added)).

[30] *See e.g.,* Douglas Lichtman, *Copyright as Innovation Policy: Google Book Search From a Law and Economics Perspective*, 9 INNOVATION POL'Y & ECON. 55, 72 (2008)

the Internet that turning small into big "can be a tricky proposition indeed, because scaling problems--the problems that arise solely as a consequence of increasing size or increasing numbers--can be profound, and profoundly difficult to solve."[31] Regarding the regulation of professional speech, Cassandra Burke Robertson and Sharona Hoffman wrote that, "[t]he scale of modern mass communication offers a much larger threat to the viability of traditional regulatory approaches."[32] And in the context of copyright infringement and enforcement, Annemarie Bridy wrote that "[w]ith each successive iteration, P2P network architecture has become not only more scalable and efficient, but also more perfectly adapted to 'massive infringement.' The key to effective online copyright enforcement in the P2P context is identifying and implementing enforcement strategies that are commensurately scalable."[33]

Scale is a common theme in privacy literature too. According to Daniel Solove, "[r]eading privacy notices is a task that does not scale. There

("In a world with a large and ever-changing list of opt-out projects, authors would be forced to invest substantial sums finding each project and notifying each about their desire to participate. The problem would be even worse if some of those opt-out programs were designed strategically to make things difficult on authors, for instance, imposing high standards of proof before acknowledging that an opt-out really came from the correct copyright holder. (Infringers have an incentive to do just that because in an opt-out system, infringers benefit if authors find it too expensive to actually engage in the mechanism of opting out.) Overall, then, the problem with an opt-out program is that it does not *scale*.") (emphasis added); Naomi Appelman & Paddy Leerssen, *On "Trusted" Flaggers*, 24 YALE J. L. & TECH. 452, 473 (2022) ("[T]rusted flagging *does not scale*. If third parties wish to influence content moderation as it is currently practiced, they must leverage its automation.") (emphasis added).

[31] DAVID POST, IN SEARCH OF JEFFERSON'S MOOSE: NOTES ON THE STATE OF CYBERSPACE 30 (2009); *see also* Jeffrey L. Vagle, *Tightening the Ooda Loop: Police Militarization, Race, and Algorithmic Surveillance*, 22 MICH. J. RACE & L. 101, 123 (2016) (noting that police departments often attempt to justify algorithmic surveillance by relying on the common trope that "an experienced and talented officer can apply their knowledge and analytical skills to attain an imperfect version of predictive policing, but that the model does not scale well.").

[32] Cassandra Burke Robertson & Sharona Hoffman, *Professional Speech at Scale*, 55 U.C. DAVIS L. REV. 2063, 2100 (2022).

[33] Annemarie Bridy, *Is Online Copyright Enforcement Scalable?*, 13 VAND. J. ENT. & TECH. L. 695, 736 (2011); *see also* Thomas C. Rubin, *Leveraging Notice and Takedown to Address Trademark Infringement Online*, 37 COLUM. J.L. & ARTS 585, 591 (2014) ("Forcing platforms to choose between uncertain but potentially enormous liability, or policing its users in a way that does not scale and that undermines the utility of the service, is no choice at all."); Doug Lichtman, *Google Book Search in the Gridlock Economy*, 53 ARIZ. L. REV. 131, 142 (2011) ("Thus, opt-out, while better than nothing, does not seem to justify a fair use finding. It simply does not scale.").

are simply too many privacy notices to read--people get notice fatigue."[34] Likewise, "[m]anaging one's privacy is a vast, complex, and never-ending project that does not scale; it becomes virtually impossible to do comprehensively."[35] Even one of us has used the concept without explaining it, writing that the concept of informed consent "does not scale without losing its legitimacy."[36]

Outside the legal literature, commentators describing challenges in regulating information technologies commonly focus on scale. A great example is content moderation. Journalist Mike Mansick, who runs the popular website Techdirt, wrote that it is sometimes "difficult to get across to people 'the scale' part when we talk about the impossibility of content moderation at scale. It's massive."[37] Journalist Helena Pozniak wrote, "[m]oderating content online is messy, arbitrary and expensive – a huge headache for lawmakers and social media companies alike. While automating such moderation is essential at scale due to the sheer volume

---

[34] Daniel J. Solove, *The Limitations of Privacy Rights*, 98 NOTRE DAME L. REV. 975, 996 (2023).

[35] Daniel J. Solove, *The Myth of the Privacy Paradox*, 89 GEO. WASH. L. REV. 1, 5 (2021) ("Second, the CCPA does not scale well. The number of organizations gathering people's data is in the thousands. Are people to make thousands of requests? Opt out thousands of times? People can make a few requests for their personal data and opt out a few times, but this will just be like trying to empty the ocean by taking out a few cups of water."); *see also* Tyler Prime & Joseph Russomanno, *The Future of FOIA: Course Corrections for the Digital Age*, 23 COMM. L. & POL'Y 267, 298 (2018) ("Currently, manual "sanitization" [of public records] is expensive, time-consuming, susceptible to disclosure risks and *does not scale* as the volume of data increases.") (emphasis added); George S. Geis, *Automating Contract Law*, 83 N.Y.U. L. REV. 450, 476 (2008) ("Manual tagging [of documents] also takes a lot of time and *does not scale*.") (emphasis added).

[36] Neil Richards & Woodrow Hartzog, *The Pathologies of Digital Consent*, 96 Wash. U.L. Rev. 1461, 1500 (2019); *see also* Woodrow Hartzog & Evan Selinger, *The Internet of Heirlooms and Disposable Things*, 17 N.C. J. L. & TECH. 581, 588 (2016) ("Bad defaults on IoT devices are common and most users cannot easily patch them. The process is usually complicated. What's worse is that the updating process for the IoT does not scale well.").

[37] Mike Mansick, *The Scale of Content Moderation is Unfathomable*, TECHDIRT (Nov. 2, 2021), https://www.techdirt.com/2021/11/02/scale-content-moderation-is-unfathomable/.

of traffic, it remains problematic."[38] The idea is that it's simply impossible to respond individually to the frequent posts of tens of millions of users.

Sometimes this same concept of scale has been invoked as an explanation of the limits of enforcement. Obama White House cyber security policy coordinator Howard Schmidt said in an interview that "[o]n cyber crime, we've always had an issue, as we have with other types of crime, which is there is oftentimes more than we can handle as law enforcement. [Our ability to respond] just does not scale."[39]

II. TWO MEANINGS OF SCALE

All of these uses of "scale" in some way refer to the relationship between the extent of an activity and its related effects. Often scale is invoked to help explain the difficulty of (cost-effectively) increasing some activity or precaution—this is usually what people mean when they say that a business model or a technological solution "won't scale." In general, these invocations of scale that focus on "increases" do little to explain the nature of the relationship between extent and effects. But that relationship is critical.

In this part, we focus on one important distinction in the meaning of scale, particularly in relation to technology law. Specifically, we draw out the distinction between ideas of "scale as more" and "scale as different." We use "scale is more" to refer to a dynamic where the effects of some activity increase as some function of the amount of that activity. "Scale is different" refers to situations where qualitatively new and different effects arise beyond some amount of the activity.

When "scale is more," the effects of an activity increase as some function of the number of instances of that activity. In the simplest example, the relationship is linear: If the amount of harm caused by each unit of activity is x, then the total harm is simply x multiplied by the number of instances of the activity. But the important point here is not the specific function by which effects of an activity increase, it's that "scale" in this

---

[38] Helen Pozniak, *Tackling the Impossible Problem of Content Moderation*, ENGINEERING AND TECHNOLOGY (April 18, 2023), https://eandt.theiet.org/content/articles/2023/04/tackling-the-impossible-problem-of-content-moderation/.

[39] 185 INTELLECTUAL PROPERTY COUNSELOR NL 10.

sense implies primarily an increase in magnitude of the *same kinds* of effects, not a qualitative change in the nature of those effects. People seem to rely upon "scale is more" logic often when thinking about whether an action can be increased indefinitely at an acceptable cost or when trying to set legal or policy thresholds. The idea behind "scale is more" logic is that, at some point, enough is enough.

In other contexts, scale is not simply more: "scale is different." In those cases, increases in magnitude do not only create more of the same kinds of effects; instead, they generate *new kinds* of effects that only emerge beyond some amount of the activity. In this sense, the relationship between the extent of an activity and its effects is n dimensional. That is, it involves or can be described by a number ("n")ofdifferent variables, parameters, or features, each one contributing to the complexity and scope of the problem[40]

We certainly don't contend scale as "more" is irrelevant. Indeed, "scale is more" commonly matters. Law frequently requires a comparison of costs and benefits, and particularly aggregate costs and benefits of repeated activities. In the tort context, courts have long evaluated negligence by considering the probability of harm, the magnitude of that harm, and the burden of preventing the harm..[41] Using that formula, the amount of an activity matters at least to the aggregate harm (the total loss) whenever the loss associated with the activity increases in relation to the amount of that activity.

Scale in this sense can also matter to the burden of preventing the harm to the extent the burden increases with the amount of the activity. Because the Hand Formula requires a comparison, the rate at which costs and benefits increase in relation to the amount of the activity is highly relevant to determining the point at which the burden outweighs the

---

[40] See, e.g., Warren Weaver, Science and Complexity, 36(4) AM. SCIENTIST 536 (1948); JOHN H. HOLLAND, ADAPTATION IN NATURAL AND ARTIFICIAL SYSTEMS: AN INTRODUCTORY ANALYSIS WITH APPLICATIONS TO BIOLOGY, CONTROL, AND ARTIFICIAL INTELLIGENCE (5th prtg., MIT Press 1998) (1975), https://ia601604.us.archive.org/2/items/holland-9780262275552/Holland_%209780262275552.pdf

[41] This is commonly referred to as the "Hand formula" because it was introduced by Judge Learned Hand in the famous case of *United States v Carroll Towing Co*., 159 F.2d 169 (2d. Cir. 1947).

discounted probability of loss. That is the point at which the legal conclusion flips and a party is no longer deemed negligent for failing to prevent the loss. That same kind of comparison is also very common in the regulatory context, despite the difficulty of quantifying all relevant costs and benefits related to most technologies.[42]

But that is not the only way costs and benefits can relate to the amount of an activity. Take, for example, vaccination rates. Public health experts have long understood that, given the efficacy rate of a particular vaccine and the infectiousness of the disease against which it inoculates, a certain percentage of the relevant population needs to be vaccinated to achieve "herd immunity." Herd immunity is the idea that, once that percentage of the population is vaccinated, the disease is effectively prevented from spreading, even though no vaccine is 100% effective for any particular recipient.[43] "Scale is different" when it comes to vaccines because the desired effect on a population doesn't exist at the individual level. Herd immunity is not achieved incrementally—it is not 80% achieved at 80% of the necessary level of immunity. It only appears once a certain magnitude threshold has been met.

There's a corollary to that idea that we unfortunately have seen in real time in the COVID era. The failure to achieve herd immunity means that the disease will continue to circulate, and some percentage of people will continue to be infected even when vaccinated. Here is where "scale is different" comes in: the failure to achieve herd immunity not only means that the same strain of COVID will continue to circulate, but the extent of that circulation also creates opportunities for new mutations and therefore new and different strains to emerge (ones not covered by the existing vaccines). Put differently, low vaccination rates don't just mean that more people will continue to get sick with the known disease (scale is more); it means that new and different harms will emerge (scale is different).

Of course, we are hardly the first to observe that new dynamics sometimes emerge at a certain level of activity. Social and political scientists, economists, engineers, and people from a variety of backgrounds

---

[42] CASS R. SUNSTEIN, THE COST-BENEFIT REVOLUTION (2018).

[43] https://my.clevelandclinic.org/health/articles/22599-herd-immunity (" Herd immunity means that enough people in a group or area have achieved immunity (protection) against a virus or other infectious agent to make it very difficult for the infection to spread.").

have demonstrated this fact repeatedly, and they have often incorporated it into the general wisdom of their fields. It's not even foreign to legal scholarship. Scholars have long understood that technologies can have "network effects" – the phenomenon where the value or utility a user derives from a good or service depends on the number of other users of that good or service.[44] That concept has been a particularly powerful way of understanding the value of networked technologies. Indeed, network effects are one explanation for natural monopolies – circumstances where the value of a service depends on number of users, and the number of users necessary to achieve sufficient value can't realistically be achieved by multiple parties.[45]

Scholars studying complex systems have observed that scale is transformative. John Holland's foundational 1975 book, Adaptation in Natural and Artificial Systems detailed four defining characteristics of systems which become complex and adapt at scale (often called "complex adaptive systems" or "CAS"). These systems:

1. Have a large numbers of parts whose local interactions produce global phenomena that do not follow linearly from interaction parameters (emergence);
2. Impact that results from aggregate behavior, which feeds back to the individual parts (feedback loops);
3. Have interactions, the nature of which evolves over time, creating perpetual novelty and eschewing equilibrium (dynamism); and
4. Have individual parts that develop "rules" to anticipate the consequences of their own responses (adaptation).[46]

---

[44] *See* Catherine Tucker, *Network Effects and Market Power: What Have We Learned in the Last Decade?*, ANTITRUST (Spring 2018), https://sites.bu.edu/tpri/files/2018/07/tucker-network-effects-antitrust2018.pdf.

[45] *See* Christopher S. Yoo & Daniel F. Spulber, *Antitrust, the Internet, and the Economics of Networks*, *in* OXFORD HANDBOOK OF INTERNATIONAL ANTITRUST ECONOMICS (2014) ("A given production technology is said to exhibit natural monopoly characteristics if it has a subadditive cost function, i.e., a single firm can supply the entire market demand at lower cost than could two or more firms.").

[46] JOHN H. HOLLAND, ADAPTATION IN NATURAL AND ARTIFICIAL SYSTEMS: AN INTRODUCTORY ANALYSIS WITH APPLICATIONS TO BIOLOGY, CONTROL, AND ARTIFICIAL

In their comprehensive study of "ultra large-scale" (ULS) systems, Northup and co-authors argue that "scale changes everything."[47] The scholars distinguish large monolithic systems from systems of systems according to characteristics such as independence of the elements, emergent behavior, and geographic distribution. They further identify seven characteristics of ULS systems arising specifically due to their scale:

1. **Decentralized control**: Top-down control of ULS systems is infeasible.
2. **Conflicting and unknowable requirements**: Different ULS system components will have differing and evolving needs.
3. **Continuous evolution and deployment**: Integration, removal, and modification of components must occur while the system is operating.
4. **Heterogeneous, inconsistent, and rapidly changing components**: Extensions and repairs to a ULS system will inevitably preclude uniformity of its parts.
5. **Blurred human–machine boundaries**: People are part of ULS systems, not merely users of them.
6. **"Normal" persistent failure**: Operations that individually have infinitesimal likelihood of error are guaranteed to produce errors when iterated the number of times a ULS system requires.
7. **The need for new acquisition and governance paradigms**: Unpredictability of stakeholder motives and needs precludes the

---

INTELLIGENCE (5th prtg., MIT Press 1998) (1975), https://ia601604.us.archive.org/2/items/holland-9780262275552/Holland_%209780262275552.pdf

[47] Linda Northrop et al., ULTRA-LARGE-SCALE SYSTEMS: THE SOFTWARE CHALLENGE OF THE FUTURE (Bill Pollak ed., Carnegie Mellon Univ. Software Eng'g Inst. 2006), https://apps.dtic.mil/sti/tr/pdf/ADA610356.pdf ("The primary characteristic of ULS systems is ultra-large size on any imaginable dimension—number of lines of code; number of people employing the system for different purposes; amount of data stored, accessed, manipulated, and refined; number of connections and interdependencies among software components; number of hardware elements; etc. But to understand the nature of ULS systems, we must go beyond just the concept of size; we must understand the effects of scale and the demands that ULS systems are likely to place on technologies and processes. Issues that are not significant at smaller scales become significant at ultra-large scales. The problems introduced by scale require new solution approaches and new concepts of system design, development, operation, and evolution. In short, scale changes everything.").

possibility that any "prime contractor" can exert centralized control over a ULS system.[48]

So while other fields of scholarship are coming to understand that scale changes everything, we think that law and technology scholars do not always sufficiently consider the variety of ways in which scale can matter. We emphasize the more general distinction between "scale is more" and "scale is different" because attention to that distinction is important to determining the appropriate policy responses. Once we've identified a problem, we tend to conceive of the solution set in reference to the original framing of that problem. If we see privacy violations as instances of individualized harm perpetrated on the particular individuals whose information has been used, the legal frameworks are likely to be designed to remedy those individualized harms, even if at "scale" in the sense that there are a lot of those individualized harms. Unless policymakers are open to the idea that scale can create new and different problems that may require different kinds of solutions, the natural tendency will be to miss the real effect of scale in some contexts.

In exploring the concept of "scale" in tech regulation, Paul Ohm has argued that "[m]ost laws either treat all regulated actors the same or assume that twice as large means only twice as powerful and twice as harmful."[49] So, for example, "penalties for causing harm often multiply the number of individuals harmed by a set dollar figure, assessing $10,000 for each victim wiretapped or around $40,000 for each child monitored without parental consent."[50]

Ohm's critique is about the tendency to treat scale simply as more. Here, if an act causes x amount of harm, when it is done at scale (treating scale as "more") then the total harm caused by that act is x multiplied by the number of instances. One privacy violation is bad. A thousand privacy violations are worse because it's an additional 999 instances of harm. That way of thinking tends to produce responses of the same structure: if the

---

[48] Linda Northrop et al., ULTRA-LARGE-SCALE SYSTEMS: THE SOFTWARE CHALLENGE OF THE FUTURE (Bill Pollak ed., Carnegie Mellon Univ. Software Eng'g Inst. 2006), https://apps.dtic.mil/sti/tr/pdf/ADA610356.pdf.

[49] Paul Ohm, *Regulating at Scale*, 2 GEO. L. TECH. REV. 546 (2018).

[50] *Id.* (citing 18 U.S.C. § 2520 (2002), Adjustment of Civil Monetary Penalty Amounts, 16 C.F.R. pt. 1 (increasing FTC civil penalties to account for inflation)).

penalty for 1 violation is x, then the penalty for 1000 violations is just 1000(x). Ohm argues persuasively that a linear approach to scale is misguided, primarily because it fails to properly account for power dynamics. "Linearly bound regulation fails to reflect how the power and harm of some digital actors increase at much more than a linear, proportional rate. In at least three important ways, a platform with one billion users is more than one hundred times more powerful and potentially harmful than a company of ten million users."[51]

In our terminology, the problem with linearly bound regulation is that it ignores the ways that scale can be different. As Ohm says:

> [A] linear model fails to offer a proper moral accounting of the way human misery scales. We might feel more impelled to prevent a small harm affecting one million victims out of one billion users than we would to prevent the same harm affecting only ten victims out of ten thousand users, even though they reflect the same rate of injury with the only difference being the size of the injurer. Second, purely digital platforms expand automatically into any territory that the Internet touches, meaning platform providers need not attend to local regulators and regulations. Third, size begets power, particularly for artificial intelligence, meaning we can expect more from globe-spanning digital platforms.[52]

We agree with Ohm that "[m]assive digital platforms thus raise significant concerns of potential harm that calls for a regulatory response that accounts for effects of size. From privacy to tort to contract to consumer protection to intellectual property laws, we should better account for the power and potential harm of size."[53] What was once a salesperson's attempt to wheedle you into buying that shirt now becomes a structured and systematized user interface that simultaneously affects billions. What was once a conspiracy theory exchanged at the bar becomes amplified to billions.

However, as we argue below, we think that recognizing the ways scale can be different does even more than allow us to account for the

---

[51] Ohm, *supra* note 38, at 546–47.
[52] *Id.*
[53] *Id.*

magnitude of power accumulation. It's not just that lawmakers and judges are getting the math wrong when thinking about scale too simplistically and linearly. Sometimes when the instances of something related to information technologies significantly increase, whole assumptions about actions and consequences must be challenged.

III. A MORE COMPLETE ACCOUNT OF SCALE

In this part, we describe at least four ways that scale can mean "different" and not just "more." We do not claim that this is an exhaustive account of the effects of scale, nor do we argue that the consequences we describe below are entirely distinct from each other. We describe these effects to highlight the ways that scale can be different, and to help guide policymakers toward more nuance in considering the effects of scale and the corresponding range of policies regarding new technologies.

A. The Population Affected Could Change

One important way in which scale is different is that the scale of an activity might change the population that is affected by that activity. In isolation, certain practices only seem to implicate those actors that are directly involved. For example, when a company collects a person's information, we might assume that only that person's privacy and autonomy was at risk. Your browsing history probably doesn't directly reveal anything about me, so Google's collection of that information is a "you" problem. This isn't always true, of course, even in isolated cases. For example, if your family member takes a DNA test and gives that information to a company, you are exposed because of the strong overlaps in familial DNA.[54] But generally speaking, our default frame of analysis for isolated actions focuses only on the people involved, either directly or

[54] Law enforcement officers have recently solved a number of "cold" cases using forensic genetic genealogy – matching the DNA profile of the suspect to living family members whose genetic profiles are known, often because those family members voluntarily tested with a commercial ancestry testing company like 23andMe. *See, e.g.*, *Multiple Cold Cases Solved with Assist from Attorney General's Dna Forensic Genetic Genealogy Program*, ATTORNEY GENERAL OF WASHINGTON (July 11, 2022), https://www.atg.wa.gov/news/news-releases/multiple-cold-cases-solved-assist-attorney-general-s-dna-forensic-genetic; Joe Hernandez, *Genealogy DNA is used to identify a murder victim from 1988—and her killer*, NPR (Sept. 8, 2022), https://www.npr.org/2022/09/08/1121542171/genealogy-dna-murder-stacey-lyn-chahorski-henry-frederick-wise-michigan-georgia.

because they have "skin in the game" by being somewhere in the supply chain or otherwise standing to gain or lose something as a result of the action.

At scale, someone's actions might implicate not just related third parties, but the interests of entire populations with shared characteristics. For example, Salome Viljoen has argued that "data-collection practices of the most powerful technology companies are aimed primarily at deriving (and producing) population-level insights regarding how data subjects relate to others, not individual insights specific to the data subject. These insights can then be applied to all individuals (not just the data subject) who share these population features."[55] According to Viljoen,

> This population-level economic motivation matters conceptually for the legal regimes that regulate the activity of data collection and use; it requires revisiting long-held notions of why individuals have a legal interest in information about them and where such interests obtain. The status quo of data-governance law, as well as prominent proposals for its reform, approach these population-level relational effects as incidental or a byproduct of eroded individual data rights, to the extent that they recognize these effects at all. As a result, both the status quo and reform proposals suffer from a common conceptual flaw: they attempt to reduce legal interests in information to individualist claims subject to individualist remedies, which are structurally incapable of representing the interests and effects of data production's population-level aims. This in turn allows significant forms of social informational harm to go unrepresented and unaddressed in how the law governs data collection, processing, and use.[56]

Something similar can be said about AI training sets. Since the goal of training is for the system to learn patterns, especially patterns that were not visible to human observers, the size and representativeness of the training set matters enormously to the functioning of the AI system. Indeed, many of the documented problems of bias in AI systems are attributable to training sets that were not sufficiently diverse. For our

---

[55] Viljoen, *supra* note 7, at 578.
[56] *Id.*

purposes here, the point is that these systems aren't useful primarily because of individual bits of information they learn from specific inputs; their real value is in recognition of patterns that are only learnable when the data set is of a certain size. Those population-level insights are then frequently baked into algorithms in ways that have much more systemic effect than do the bits of information themselves.

B. Emergent Problems

One of the most obvious ways that scale can be different is that new and qualitatively different kinds of problems can emerge at certain thresholds. That is what we described with respect to the insufficient uptake of COVID vaccines: the lack of herd immunity allowed the virus to circulate at a scale that didn't just lead to more people being infected with the same variant, it enabled the emergence of new variants that would affect even the vaccinated.

Kathleen Creel and Deborah Hellman have described the ways that algorithmic decision-making at scale can produce meaningfully different problems as compared to individualized decisions on the same issues.[57] Specifically, Creel and Hellman argue that arbitrary individualized decisions (hiring decisions based on irrelevant characteristics, for example) generally don't rise to the level of moral concern because there's no strong interest in any individual decision being non-arbitrary (as opposed to non-biased). But, they argue, widespread adoption of an algorithmic system is different: whereas individual human decision-makers tend to be differently arbitrary, an algorithmic system locks in a single arbitrary choice, systematically locking people out of opportunities (jobs, credit, etc.).[58] Arbitrariness at scale creates a new and different problem that isn't just the sum of the harms of individual decisions.

Julie Cohen's work on platforms and infrastructure also demonstrates how economies of scale can surface problems that do not

---

[57] Kathleen Creel & Deborah Hellman, *The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision-Making Systems*, 52 CAN. J. PHIL. 26 (2022).

[58] *Id.*

exist in small numbers or with more limited affordances.[59] Because platforms are so large and have so many multi-sided relationships with both buyers and sellers, at a certain tipping point platforms are able to exploit their dominance in ways that perpetuate their market power but may be more difficult for regulators to detect. For example, "[b]ecause the economics of platforms permit so many different arrangements, pricing ceases to be a reliable sign of market power, and courts and regulators lose a previously reliable metric for determining whether power has been abused."[60] The scalability of platforms also places them at the center of market exchange, which allows these platforms to create market dependencies and hide the ways in which they engage in self-preferencing.[61]

Misinformation is also a good example of "scale is different." Individual pieces of misinformation are, of course, potentially harmful, because they can affect the behavior of those who receive it. People who believed President Obama was not born in the United States were more likely to vote against him and to be skeptical of anything his administration supported. That harm is surely multiplied as more misinformation circulates, which means there's an important "scale is more" effect in this context. But there are also important ways in which misinformation at scale is different. For one thing, the perceived credibility of any particular bit of misinformation might be impacted by the extent of that misinformation's circulation. This might be an example of where the effect of scale is to change the population affected: people who would be skeptical of a piece of misinformation when that misinformation was not widely circulated might become more likely to credit the misinformation when it circulates at greater scale. Scale even plays a key factor in distinguishing the idea of misinformation from disinformation. Ryan Calo, Chris Coward, Emma Spiro, Kate Starbird, and Jevin D. West have helpfully distinguished the two concepts along the lines of intent and scale:

---

[59] See, e.g., JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019); Julie E. Cohen, Infrastructuring the Digital Public Sphere, 25 Yale J.L. & Tech. 1 (2023); Julie E. Cohen, Oligarchy, State, and Cryptopia, 94 Fordham L. Rev. (forthcoming).

[60] COHEN, BETWEEN TRUTH AND POWER at 25.

[61] *Id.*

> [Misinformation is the] erroneous or misleading information to which the public may be exposed, engage with, and share….Disinformation refers to a purposive strategy to induce false belief, channel behavior, or damage trust. Misinformation is usually discrete or standalone, as when a neighbor shares a false rumor or overhears a misleading exchange. Disinformation tends to take the form of a multifaceted campaign with a predetermined financial, political, or other objective. Disinformation campaigns blend orchestrated action and organic activity, relying on the participation of willing but unwitting online audiences.[62]

Misinformation can exist in isolation, but disinformation requires scale for success.

There's another sense in which scale is different when it comes to misinformation. Specifically, there's a point at which there's so much misinformation, particularly in certain places or among certain groups, that it threatens destruction of belief in the idea of truth. That is a widely recognized feature of Russian disinformation: it is intended not just to convince people of the specific claims in individual pieces of misinformation, but to sow chaos and create doubt that there is any such thing as truth, particularly in official information.[63] That "flood the zone" strategy is premised entirely on the recognition that scale is different: beyond some point, the problem isn't really the specific misinformation, it's the epistemic free-for-all.

Facial recognition is another context in which scale is different. Individual uses of facial recognition technology can cause a variety of harms. If the technology is trained on disproportionately white faces, that technology is much more likely to misidentify non-white people, causing any number of discrete harms to the people misidentified. Those people

---

[62] Ryan Calo, Chris Coward, Emma Spiro, Kate Starbird, & Jevin D. West, *How Do You Solve a Problem Like Misinformation?*, SCIENCE ADVANCES (Dec. 8, 2021), https://www.science.org/doi/epdf/10.1126/sciadv.abn0481.

[63] Why Russian Disinformation Matters (https://www.cmu.edu/ideas-social-cybersecurity/events/ideas2024_paper_16.pdf

might be denied entry to a concert or sporting event,[64] and they might even be wrongly arrested.[65] More extensive use of such biased technology might repeat that harm over many people, increasing the aggregate harm. In that sense, scale is more. But widespread deployment of facial recognition technology across a range of settings also does something different and more insidious: it threatens a total surveillance society and a complete loss of obscurity.[66] In that sense, scale is very different.

Website scraping is a similar example. The owners of publicly-available websites (those unrestricted by passwords or privacy settings) should reasonable expect their websites to be accessed by all kinds of people as part of their normal use of a computer. But when bots scrape social media websites like LinkedIn and Twitter and preserve snapshots of those same websites at scale, things get weird. Not only can that level of continuous access crash a server, but once scraped, bits of information become sortable, cheaply stored, easily aggregated, effortlessly shared, perfectly preserved, and repurposed. As a result, information can be aggregated to paint pictures of human behavior that were unlikely part of people's threat modeling when they originally posted on social media. For example, someone polishing up their LinkedIn profile might not suspect that an their employer was using an automated bot scraping LinkedIn

---

[64] Kashmir Hill and Corey Kilgannon, Madison Square Garden Uses Facial Recognition to Ban Its Owner's Enemies, N.Y. Times (Jan. 3, 2023), https://www.nytimes.com/2022/12/22/nyregion/madison-square-garden-facial-recognition.html.

[65] Kashmir Hill, Wrongfully Accused by an Algorithm, N.Y. Times (June 24, 2020), https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

[66] Woodrow Hartzog & Evan Selinger, *Surveillance as Loss of Obscurity*, 72 WASH. & LEE L. REV. 1343, 1345–46 (2015) ("[W]e argue that the concept of "obscurity," which deals with the transaction costs involved in finding or understanding information, is the key to understanding and uniting modern debates about government surveillance."); Woodrow Hartzog & Evan Selinger, *Increasing the Transaction Costs of Harassment*, 95 B.U. L. REV. ANNEX 47 (2015); Evan Selinger & Woodrow Hartzog, *Obscurity and Privacy*, *in* SPACES FOR THE FUTURE: ROUTLEDGE COMPANION TO PHILOSOPHY OF TECHNOLOGY (Joseph Pitt & Ashley Shew eds., 2018), https://www.routledge.com/Spaces-for-the-Future-A-Companion-to-Philosophy-of-Technology/Pitt-Shew/p/book/9780415842969; *see also* Woodrow Hartzog & Frederic Stutzman, *The Case for Online Obscurity*, 101 CAL. L. REV. 1, 5 (2013) ("We argue the case for obscurity for two reasons. First, we argue that obscurity is a common and natural condition of interaction, and therefore human expectation of obscurity will transfer to the domains in which we spend time, both physical and virtual. Second, we argue that obscurity is a desirable state because we are protected by an observer's inability to comprehend our actions, and therefore social practice encourages us to seek obscurity."); Woodrow Hartzog & Frederic Stutzman, *Obscurity by Design*, 88 WASH. L. REV. 385 (2013).

profiles at scale to find evidence of which employees were a "flight risk," that is, looking for another job.[67] On top of that, their photos can power databases that turn your face into a tracking beacon, obliterating our collective practical anonymity and ability to hide in plain sight. These are bad times at scale.

C. Challenge the Assumption of the Original Problem

Actions at scale might also cause us to challenge the original assumptions regarding effects of an activity and various parties' incentives. For example, for years policymakers have considered information privacy issues to be primarily an issue of the dignity or interest of individual people. Public revelations of private information might cause emotional distress, chilling effects, or financial harm. Leaked health information might cause others to act differently towards you. If your credit card number gets out, a thief might wrongfully make charges to your account. The harms were individual harms visited on particular people, and the costs and benefits of legal responses or of tools for avoiding those harms were understood in

---

[67] See Maureen K. Olhausen & Peter Huston, hiQ v LinkedIn: A Clash Between Privacy and Competition, The Evolution of Antitrust in the Digital Era: Essays on Competition Policy, https://www.bakerbotts.com/~/media/files/thought-leadership/publications/2020/november/hiq-v-linkedin-a-clash-between-privacy-and-competition.pdf?la=en&hash=7C8DDF672A38EAC363C5CC1CBBAE4F72F459DC59Meet Jill. She's not happy at work. Her employer doesn't pay her what she's worth and her boss is a jerk. She decides to start looking for a new job, discretely. As a first step, Jill wants to make sure her profile on LinkedIn, the popular on-line professional network, is sparkling. She updates her list of accomplishments, polishes up the description of her experience, solicits some peer recommendations, and sends out a round of invitations to join her network. To keep her plans private, she double-checks her LinkedIn settings to make sure that each change she makes to her profile is not broadcast to her connections, which include several work colleagues. Unfortunately, Jill's goal of keeping her job search covert is not shared by hiQ Labs, a data analytics company. HiQ's automated bots scrape data from LinkedIn's servers and run it through the hiQ algorithm. HiQ determines that Jill is a "flight risk." For a fee, and unbeknownst to Jill, hiQ presents this determination to her employer. At this point, things could veer in a couple of different directions. Maybe Jill's employer, armed with hiQ's "flight risk" conclusion, realizes how valuable she is, offers her a raise and fixes the issues that caused her to be dissatisfied in the first place. On the other hand, maybe Jill's boss demotes her, makes her life even more miserable, and sabotages her chances of finding another job. Either way, she did not consent to hiQ's analysis and use of her LinkedIn data and her life is altered from the course she planned. Such a scenario is at the heart of litigation now pending between hiQ and LinkedIn."); Kevin Moss et al., The Legal Battle Between hiQ and LinkedIn Over Public User Data, Bloomberg Law (Oct. 18, 2017), https://news.bloomberglaw.com/legal-ops-and-tech/the-legal-battle-between-hiq-and-linkedin-over-public-user-data.

relation to the nature of the individual harms. As Salome Viljoen has argued, "[p]rivacy and data-governance law have traditionally governed forms of private interpersonal exchange in order to secure the benefits of data-subject dignity or autonomy.

But at scale, lawmakers might (and should) conceive of the risk of harm differently.

> [A]s data collection and use become key productive activities (i.e., economic activities that define the contemporary economy as an information economy), new kinds of information-based harm arise. There is growing evidence of the role that digital technology plays in facilitating social and economic inequality. Digital-surveillance technologies used to enhance user experience for the rich simultaneously provide methods of discipline and punishment for the poor. Algorithmic systems may reproduce or amplify sex and race discrimination. Even seemingly innocuous data collection may be used in service of domination and oppression. The pursuit of user attention and uninterrupted access to data flows amplifies forms of identitarian polarization, aggression, and even violence. Such evidence suggests that social processes of datafication not only produce violations of personal dignity or autonomy, but also enact or amplify social inequality."[68]

As a result, "alongside traditional concerns over individual autonomy, the social inequalities that result from data production are also forms of informational harm."[69]

We might say something similar about the structuring of our regulatory system around the value of choice, particularly though not exclusively in relation to privacy. It's not that each individual choice is hard, but once we have adopted a system that prioritizes choice, it throws consumers into a world where they're drowning in choice. This is the fundamental problem with notice and consent as a model of privacy regulation. There's a sense in which this might be considered a scale as

---

[68] Salomé Viljoen, *A Relational Theory of Data Governance*, 131 YALE L.J. 573, 580–81 (2021).

[69] *Id.* at 582.

more problem: each of these choices has some cost, and there's just a tipping point beyond which the costs overwhelm the benefits of choice. But we think this is an example of the ways that scale can illuminate problems with the original understanding of the costs and benefits of a choice-focused model. It's not just that the costs of each choice will mount. It's that, the model produces an environment that is not conducive to meaningful choice even in the individual instances.

Dark patterns might be another example like this. Dark patterns aredesign practices meant to influence people using technologies though manipulative, coercive, and deceptive means.[70] . Think of the additional steps intentionally inserted into the user experience of trying to cancel an account or the "I Agree" button highlighted and made prominent while the "x" or "close" button is small, easy to ignore, and hard to click.[71] Often these design techniques seem more like minor annoyances when viewed in isolation, but in the aggregate they can pollute the entire digital environment.[72]

What is important about these examples is that the regulatory model works outwardly from characterization of the individual instance, dismissing harms as *de minimus* or perhaps even seeing each instance as

[70] Commission Regulation 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277) 18 ("Dark patterns on online interfaces of online platforms are practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make autonomous and informed choices or decisions. Those practices can be used to persuade the recipients of the service to engage in unwanted behaviours or into undesired decisions which have negative consequences for them."

[71] Colin M. Gray, Cristiana Teixeira Santos, Nataliia Bielova & Thomas Mildner, An Ontology of Dark Patterns Knowledge: Foundations, Definitions, and a Pathway for Shared Knowledge-Building, in CHI '24: PROCEEDINGS OF THE CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS (2024), https://dl.acm.org/doi/10.1145/3613904.3642436.

[72] Johanna Gunawan et. al., *Dark Patterns As Disloyal Design*, 100 IND. L.J. 1389, 1405–06 (2025) ("[B]oth the European Union and the United States commonly rely on unfairness tests to determine violations. However, such tests commonly require thresholds to be met and place a burden of proof for articulating resultant harms or risk of harm. This may present too high a standard by which to regulate de minimis issues arising from dark patterns, particularly when examining dark patterns at an individual scale.").

net beneficial, but only because each action is viewed in isolation.[73] At scale, things look very different, even in terms of how we see individual instances.

Scale can also have the effect of normalizing practices, discouraging public resistance, and encouraging conformity, which might cause people to reevaluate their initial resistance to tthose practices. In research with Evan Selinger and Johanna Gunawan, one of us has argued that the ubiquity and ultimate mundanity of *de minimus* privacy encroachments can, at scale, both distort and bypass our ability to critically reflect upon the danger of exposure.[74]

> Two normalization dynamics that revolve around repeated exposure, "unexceptional habituation" and "favorably disposed normalization," might also play important roles in shaping how people view surveillance. *Unexceptional habituation* occurs when people in liberal Western democracies take ubiquitously encountered surveillance systems for granted—seeing them as so commonplace and mundane they are not worth thinking about critically….The psychological dynamic of *favorably disposed normalization*, whereby the routine experience of being surveilled inclines people to view surveillance as acceptable, if not desirable, might significantly influence what people believe is appropriate privacy policy."[75]

---

[73] *See generally* Max L. Veech & Charles R. Moon, *De Minimis Non Curat Lex*, 45 MICH. L. REV. 537 (1947); Frederick G. McKean Jr., *De Minimis Non Curat Lex* , 75 U. PA. L. REV. 429 (1927).

[74] Woodrow Hartzog, Evan Selinger, and Johanna Gunawan, *Privacy Nicks: How the Law Normalizes Surveillance*, 101 WASH. U. L. REV. 717 (2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4384541.

[75] *Id.* (citing Clare Southerton & Emmeline Taylor, *Habitual Disclosure: Routine, Affordance, and the Ethics of Young Peoples Social Media Data Surveillance*, 6 SOC. MEDIA & SOC'Y (2020); Evan Selinger & Judy Rhee, *Normalizing Surveillance*, 22 N. EUR. J. PHIL. 49 (2021)). We expanded upon this idea, writing:

> One plausible psychological basis for favorably disposed normalization is the impact of believing something is normal. Thinking something is normal does not necessarily entail a commitment to deeming that thing ethical. Nevertheless, normality judgments often are accompanied by positive affective experiences. For example, imagine someone believes using Facebook is ethically problematic but normal. That person might feel less badly about using Facebook than someone who believes the practice is ethically problematic and abnormal. The difference in how people feel has

Scale can also change incentives about the problem. Consider tort law's economic loss doctrine, which precludes liability for monetary harms that are not derivative of harms to person or physical property.[76] That rule is typically justified by reference to scale—specifically, the claim is that a different rule would make the scale of liability unmanageable for parties and even for the court system itself.[77] In other words, if your conduct causes enough harm to a wide enough range of people, the system is inclined to deny liability altogether. Anticipation of the scale of liability changes how we might think about the problem, which leads us to the final way scale can be different and not just more.

> implications for governance. The person with a stronger felt sense of discomfort might have a greater incentive to quit the platform. After all, people frequently complain about ethical violations. But taking the next step of committed action can require more than intellectual awareness that change is needed. Given the practical value of heightened moral motivation for rectifying injustice, in some circumstances, "beliefs about normality might be more important than moral beliefs." But how do people develop the belief something is normal? According to experiments conducted by philosophy and cognitive science professor Joshua Knobe and psychology professor Adam Bear, both prescriptive and descriptive information matter if people know how good something is perceived and how prevalent it is. Nevertheless, simply "increasing the frequency of something occurring," such as surveillance more becoming more prevalent, can lead people to perceive it as "more normal," not just increasingly widespread. Supporting evidence for this thesis exists in the experimental literature on environmental messaging. Alternatively, one might explain the dynamic of favorably disposed normalization through the psychological process of rationalization. From this perspective, people generally are motivated to see themselves positively, as moral, intelligent, and in control of their lives. To maintain this narrative and minimize inconsistency when making decisions that seem unethical, stupid, or unfree, they often subconsciously turn to rationalization. Put otherwise, being aware of a gap between how we would like to act and how we actually behave can be stressful because it creates cognitive dissonance. Rationalization is ameliorative because it can minimize or dispel cognitive dissonance. Rationalization provides people with a means to convince themselves they should see their situation differently—that seemingly troubling behavior is justifiable, tolerable, and in some cases, even laudable.

Id. (citing Nathanael J. Fast & Arthur S. Jago, *Privacy Matters…or Does It? Algorithms, Rationalization, and the Erosion of Concern for Privacy*, 31 CURRENT OP. PSYCH. 44 (2020)).

[76] See, e.g., Robins Dry Dock & Repair v. Flint, 275 U.S. 303 (1927).

[77] See Aikens v. Debow, 208 W.Va. 486, 492 (W. Va. 2000) (discussing "the danger of expanding the concept of duty in tort to include economic interests and consequent exposure of defendants 'to a liability in an indeterminate amount for an indeterminate time to an indeterminate class. The hazards of a business conducted on these terms are so extreme as to enkindle doubt whether a flaw may not exist in the implicating of a duty that exposes to these consequences.'").

D. The Solution Set Can Change

For lawmakers, the most important implication of distinguishing when scale is different is that, in those cases, the solution set might well be different than it is when scale is just more. For example, at scale, we might see that certain remedies that sound good in isolation, like data access and deletion rights in privacy law or even private law remedies such as breach of contract, are not nearly as effective as other measures such as infrastructural or institutional design remedies.[78] We mean here not only that regulatory intervention becomes more important at a certain scale, but that the types of interventions and even the identity of the regulatory actors might be different. Ryan Calo identified a good example in the context of electric cars. Electric cars initially posed a new kind of danger to pedestrians: because those cars do not have internal combustion engines, they are much quieter, and pedestrians were much less attuned to their presence. But rather than blanketing sidewalks with signs attempting to warn pedestrians about these silent vehicles, regulators turned to a form of "visceral notice": "requiring fake engine noises that change depending on the distance of the car as a natural warning embedded in the pedestrian's experience."[79] The scale of electric vehicle adoption affected human behavior to the point where a design solution that took advantage of the societal expectation that the way to tell if a car is coming is to listen for the sound of an engine.

This kind of visceral notice solution only becomes possible when the relevant technology is deployed at a certain scale, because a few electric

---

[78] Daniel J. Solove, The Limitations of Privacy Rights, 98 Notre Dame Law Review 975 (2023); Woodrow Hartzog and Daniel Solove, Privacy as Contract?, Harv. J.L. & Tech (forthcoming 2025); Ari Ezra Waldman, Privacy's Rights Trap, 117 Nw. U.L. Rev. Online 88, 89 (2022); Julie E. Cohen, Infrastructuring the Digital Public Sphere, 25 Yale J. L. & Tech. 1, 8-11 (2023) ("Platformized communication systems have posed two types of persistent and confounding challenges to that understanding of the digital public sphere and its governance mechanisms….First, content governance programs need to be implemented at scale within large, complex organizations that also have other priorities. Second, and relatedly, speaking about targeting and removal in the fairly absolute terms suggested by the ideas of "control" and "censorship" papers over a state of systemic, technical complexity in which far more fine-grained tuning of content flows at scale is the norm."); see also Edward J. Oughton et al., *Infrastructure as a Complex Adaptive System*, 2018 COMPLEXITY, File No. 3427826.

[79] M. Ryan Calo, *Against Notice Skepticism in Privacy (and Elsewhere)*, 87 NOTRE DAME L. REV. 1027, 1036 (2022).

cars driving around here and there are less likely to shape our collective expectations. But the scale of adoption is also related to the existence of the problem: if all vehicles were electric, people probably wouldn't have learned to identify cars by the sound of their engines, and if electric vehicles were widely adopted, at some point people might stop listening for them. . This presents a potential threshold legal intervention question for lawmakers. At what point should the law intervene or stop caring? Scale can help us understand not just when we need more regulation, but also when we need less.

Greater precision about the meaning of scale can also be important for identifying regulatory choke points, or points when certain kinds of enforcement is likely to become futile. Secondary liability in copyright, for example, is often justified on the ground that the scale of direct infringement by distributed infringers is likely to make enforcement practically impossible for rights holders. For that reason, it is sometimes important that rights holders be able to hold liable those that provide the means and instrumentalities of infringement, which might effectively cut off the ability of downstream infringers to act. Likewise, attention to the effects of scale might lead decisionmakers to try to determine the extent to which harm is attributable to many actors or a small number of large actors. For example, it might be important to know whether most misinformation on a platform is attributable to a particular bstate-sponsored purveyor of disinformation using bots. Even if there are billions of instances of an activity, lawmakers' solution set should depend upon whether there are 100 or 1 million bad actors.

Complexity science scholarship counsels moving beyond the notion that ex ante rule issuance alone can effectively govern behavior in real-world complex adaptive systems (CAS).[80] As such, lawmakers might better

[80] Warren Weaver, *Science and Complexity*, 36(4) AM. SCIENTIST 536 (1948), https://ia601007.us.archive.org/28/items/weaver_27826254/Science-and-Complexity.pdf; Thomas C. Schelling, *Dynamic Models of Segregation*, 1(2) J. MATHEMATICAL SOCIO. 143 (1971), https://www.suz.uzh.ch/dam/jcr:00000000-68cb-72db-ffff-ffffff8071db/04.02_schelling_71.pdf; Linda Northrop et al., ULTRA-LARGE-SCALE SYSTEMS: THE SOFTWARE CHALLENGE OF THE FUTURE (Bill Pollak ed., Carnegie Mellon Univ. Software Eng'g Inst. 2006), https://apps.dtic.mil/sti/tr/pdf/ADA610356.pdf; JOHN H. HOLLAND, ADAPTATION IN

recognizethat regulations must evolve dynamically to adapt to systems whose structure and behavior cannot be fully predicted, thereby mimicking the systems themselves.[81]

IV. THE FAILURES OF IGNORING "SCALE IS DIFFERENT"

We are, of course, not the first to observed that new dynamics sometimes emerge at a certain magnitude of activity. Scholars have long recognized that technologies can have "network effects"–the phenomenon where the value or utility a user derives from a good or service depends on the number of other users of that good or service. That concept has been a particularly powerful way of understanding the value of networked technologies.

But we think the legal discourse has not fully appreciated that network effects are a species of a broader category where scale is different, and we argue that law and technology scholars and regulators are not always sufficiently attentive to the variety of ways in which scale can matter. As we elaborate below, there are at least three ways that failure to distinguish between these different kinds of scale effects can negatively affect our regulatory responses in the context of technology.

A. Recognition Failure

When lawmakers only conceive of scale as more, they might fail to recognize effects caused by some practice or activity because the effects of that activity are small enough in individual instances that they seem safe to ignore, and they regard the overall effects as just some multiple of the minor individualized effects. Seeing effects in that frame might lead lawmakers or judges to fail to recognize ways that scale might lead to qualitatively different harms. An example might be manipulative user

---

NATURAL AND ARTIFICIAL SYSTEMS: AN INTRODUCTORY ANALYSIS WITH APPLICATIONS TO BIOLOGY, CONTROL, AND ARTIFICIAL INTELLIGENCE (5th prtg., MIT Press 1998) (1975), https://ia601604.us.archive.org/2/items/holland-9780262275552/Holland_%209780262275552.pdf; MELANIE MITCHELL, COMPLEXITY: A GUIDED TOUR (Oxford Univ. Press 2009), http://home.iscte-iul.pt/~jmal/mcc/Complexity_-_A_Guided_Tour.pdf.

[81] J.B. Ruhl, *Complexity Theory as a Paradigm for the Dynamical Law-And-Society System: A Wake-Up Call for Legal Reductionism and the Modern Administrative State Authors*, 45(5) DUKE L.J. 849 (1996); J.B. Ruhl, Regulation by Adaptive Management—Is It Possible?, 7(1) MINN. J.L. SCI. & TECH. 21 (2005); J.B. Ruhl & Robert L. Fischman, Adaptive Management in the Courts, 95 MINN. L. REV. 424 (2010).

interfaces known as "dark patterns" that interfere with people's decisions, distract them, and extract both time and labor from users. In small doses, the diverted labor and attention might just be an annoyance. But consider how scale might result in more cognizable harms by significantly interfering with our ability to concentrate and complete tasks. Dark patterns might pollute the entire online environment, making users generally distrustful or less willing to engage online environments. Regarding labor theft, getting someone to do work for you in small bits might be fine. But at what point does it result in an opportunity cost in terms of time or what we would consider to be wrongful exploitation?

An increasingly large part of technology law hinges upon assessing risk.[82] Many new rules require data processors, designers, and even the deployers of technologies to perform algorithmic impact assessments. If lawmakers are not sensitive to the affordances that can give rise to a "scale is different" problem, they might unintentionally create incentives for shortsighted impact assessments and pave the way for unacceptably dangerous technologies to be adopted and normalized because those technologies seem benign in isolation or when used in small doses. As Margot Kaminski wrote, "risk regulation typically assumes a technology will be adopted despite its harms. [And] while aspects of risk regulation may be effective at certain kinds of harm mitigation, risk regulation as a legal interface elides, or renders invisible, both certain kinds of harms (typically, those that are less readily quantifiable) and certain individuals and populations (typically, marginalized individuals and populations) harmed by AI."[83]

For example, categorizing the risk of data collection and processing by type is very difficult, because data is endlessly combinable.[84] Lawmakers should not be timid in forecasting the risk of data practices in a world where "everything reveals everything" and industry and government have every

---

[82] See, e.g., Margot Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. 1347 (2023); Andrew Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J.L. & TECH. 117 (2021)

[83] Kaminski, supra note ^ at 1352.

[84] See, e.g., Daniel Solove, *Data Is What Data Does: Regulating Use, Harm, and Risk Instead of Sensitive Data of Sensitive Data*, 118 NORTHWESTERN U. L. REV. 1081 (2024); Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. 1125 (2015).

incentive to build systems to entrench power, control populations, and profit.[85]

B. Framing Failure

Once people identify a problem, they tend to conceive of the solution set in reference to the original framing of that problem. This has effects within a particular kind of regulatory framing (we'll call that an internal framing issue), and in terms of the broader conception of a problem (we'll call that an external framing problem).

Our systems of privacy regulation reflect precisely what we mean by an "internal" framing issue. If we see privacy violations as instances of individualized harm to the individuals whose information has been used, then it should be no surprise that the legal frameworks are designed to remedy those individualized harms—even if the harms at issue derive from uses of technologies that implicate privacy at "scale" in the sense that there are a lot of those individualized harms. But that institutional design leaves scale-is-different problems out of view. Predictive algorithms are a good example here: if we conceive of the concerns about those algorithms in terms of the collection and use of individuals' data, we are likely to frame out of view the effects of those algorithms when used to deny credit or exclude people from labor markets. Those effects are only likely to materialize when the algorithms are adopted widely enough that they are used not just to make predictions about those whose data they collect, but to make predictions about others. The algorithms can't do that kind of prediction without lots of data.

What we call an "external" framing issue has to do with the legal categories we recognize as being implicated by some activity. Here the issue isn't that we fail to see certain kinds of privacy-related harms by not recognizing where scale is different; ; the issue is that, viewed through the lens of scale as more, we only see the privacy-related harms and leave other kinds of issues totally out of the frame. Consider the regulation of crypto currency: an initial framing of the regulatory concerns about cypto

[85] See, e.g., COHEN, BETWEEN TRUTH AND POWER; Paul Ohm and Scott Peppet, *What if Everything Reveals Everything?, in Big Data is Not a Monolith* (Cassidy R. Sugimoto, Hamid R. Ekbia, Michael Mattioli, eds. 2016); Julia Angwin, *'This Is What We Were Always Scared of': DOGE Is Building a Surveillance State*, N.Y. TIMES (April 30, 2025), https://www.nytimes.com/2025/04/30/opinion/musk-doge-data-ai.html.

currency might focus primarily on the kinds of risks and harms that we would associate with regulation of financial instruments. That might mean that we focus on particular actors as the relevant regulators (in this example, probably the Securities and Exchange Commission, at least before the Trump administration began leveraging cryptocurrency for corrupt purposes). That approach leaves out of view the massive environmental costs of crypto mining (different kinds of costs that are associated with the scale of crypto mining), and therefore ignoring other regulatory actors with more expertise and relevant tools for addressing the ignored harms (the Environmental Protection Agency, for example).

C. Intervention Failure

As our framing discussion suggests, failure to differentiate types of scale effects can misdirect judgments about *who* the relevant regulatory actors are and, importantly, about *when* regulatory intervention is called for. The general inclination of lawmakers is to foster innovation by allowing technologies to be developed, refined, and deployed with as few regulations as possible. But recognizing that scale can be different points up the danger of waiting too to fully understand the social impacts of technologies—when clarity finally arrives, those tools and systems might already be too entrenched to resist. In STS scholarship this is referred to as the "Collinridge dilemma," and it gives more nuance to what some law and tech scholars describe as the "avocado ripeness" problem. (Not yet…not yet…not yet…too late.)[86]

As Ryan Calo has said, "[t]ry to intervene too soon, and policymakers risk misunderstanding the social impacts of emerging technology and hence doing more harm than good. Try to intervene too late, however, and technology will have already become intertwined in the fabric of everyday life."[87] The result, in Calo's description, is often a kind of "constant state of watchful paralysis."[88] Recognizing that entirely new kinds of harms can arise at a certain scale (scale is different) presents

---

[86] DAVID COLLINGRIDGE, THE SOCIAL CONTROL OF TECHNOLOGY (Frances Pinter 1980).

[87] Ryan Calo, *The Scale and the Reactor* (Apr. 9, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4079851.

[88] I*d.* (citing DAVID COLLINGRIDGE, THE SOCIAL CONTROL OF TECHNOLOGY (Frances Pinter 1980).

important additional support for the precautionary principle--the idea that, where there is uncertainty about the effects of some activity but there is significant potential harm, regulators should err on the side of preventing the harm. The precautionary principle is even further justified if lawmakers were to periodically revisit rules, injecting regular democratic deliberation into lawmaking to fight harmful creep.[89]Facial recognition is a great example of this dynamic and a cautionary tale for how lawmakers are currently treating generative AI. Regulators around the world are hesitant to regulate facial recognition without being able to a specific individual harm such as emotional distress, financial loss, a diminished reputation, or significant denial of autonomy and dignity through lack of consent. Sometimes facial recognition leads to these kinds of harms. But other times, the real cost of these surveillance systems is social, involves the creation of a power imbalance and eventual exploitation of that power, and is hard to see at the individual level.[90] Meanwhile the most dangerous surveillance tool ever created is becoming entrenched in the digital systems that run our lives and is being normalized with every Face ID scan, Snapchat filter, airline check-in, and IoT doorbell.[91] We are in a brief window where it the cost of substantive prohibitions on these tools would be acceptable, but the more we come to rely upon them, the greater the cost. At some point, we will have no choice but to tolerate tools that have irrevocably exposed us and permanently diminished our privacy with virtually no democratic accountability.

V. HOW TO TAKE SCALE SERIOUSLY IN LAW AND POLICY

A more developed concept of scale would have significant implications for technology law and policy. The most fundamental change might be to the way scholars and policymakers reason through problems involving data, algorithms, sensors, and actuators. Metaphors and threat modeling are common ways of thinking about technology regulation, but metaphors can only be appropriate, and threat modeling only accurate, if

---

[89] *See* BRETT FRISCHMANN AND EVAN SELINGER, RE-ENGINEERING HUMANITY (2018); Hartzog, Selinger, & Gunawan, *supra* note 53.

[90] Hartzog, Selinger, & Gunawan, *supra* note 53.

[91] *Id.; see also* Daniel Wroclawski, *Facial Recognition Is Coming to Your Neighborhood Through Home Security Cameras and Video Doorbells*, CONSUMER REPORTS (May 2, 2023), https://www.consumerreports.org/electronics/privacy/facial-recognition-and-home-security-cameras-video-doorbells-a9500287020/.

we properly account for the effects of scale.[92] To that end, we join scholars like Ryan Calo who have called for law and technology to adopt a more sophisticated approach to technology and its relationship to humans and human goals by drawing from science and technology studies (STS) and related disciplines.[93]

STS scholars have explored how human behavior can change how a technology works at scale for decades.[94] Failing to deeply engage with STS has cost the field of law and technology wisdom and nuance. Indeed, law and tech scholarship has often fallen into some of the very traps STS aimed to avoid--adopting too strong a sense of technological determinism and the misguided idea that technology will shape behavior in one single way and no other.[95] This wisdom can also help lawmakers better project how and in what situations scale might be different, and not just more.

We recommend a simple rule of thumb for all policymakers and scholars approaching law and technology issues: *start with scale*. People studying and working in law and technology often seem to think about these technologies by starting with individual or atomized instances of technological deployments and working outward only later, if at all. Privacy is a great example. Over the past fifty years, it seems that lawmakers have

---

[92] See Ryan Calo, *Modeling Through*, 71 DUKE L. J. 1391 (2022); Ryan Calo, *Robots as Legal Metaphors*, 30 HARV. J. L. TECH. 209 (2017). It is worth nothing that some AI regulatory frameworks in the European Union have explicitly incorporated considerations of scale, such as the European Union's AI Act, which regulates general purpose AI models "where such an AI model is trained with a large amount of data using **self-supervision at scale**" and targets AI systems that pose "systemic risk," which the act defines as "a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be **propagated at scale** across the value chain." Regulation (EU) 2024/1689, art. 3(63), 3(65), 2024 O.J. ^, https://artificialintelligenceact.eu/article/3/ (emphasis added).

[93] Ryan Calo, *The Scale and the Reactor* (Apr. 9, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4079851.

[94] See, e.g., Trevor Pinch & Wiebe Bijker, *The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other*, 14 SOC. STUD. SCI. 399 (Aug., 1984); ROBERT MCGINN, SCIENCE, TECHNOLOGY, AND SOCIETY (Prentice Hall 1991); Sheila Jasanoff, *A Field of Its Own: The Emergence of Science and Technology Studies*, in THE OXFORD HANDBOOK OF INTERDISCIPLINARITY (Robert Frodeman ed., Oxford University Press 2ed. 2017).

[95] RYAN CALO, LAW AND TECHNOLOGY: A METHODOLOGICAL APPROACH (forthcoming 2025).

based most of privacy law around giving people control over their personal information. Control is a laudable goal in theory and in isolation. It serves our interests in autonomy, one of the most foundational values in nearly all Western legal frameworks. But informational self-determination fails at scale. We think basing privacy law and policy on concepts like consent and individual data subject rights is the wrong starting point because it ignores how these approaches work, change, and ultimately fail at scale. Consent models start with the efficacy of an individual choice and then work outward.

But we would be better off if lawmakers were to assume that scale is inevitable for all issues implicating the use of technology, and that scale could have several different kinds of effects. If lawmakers had started with scale for privacy law, they might have embraced more structural, social, and relational approaches that focused on mitigating abuses of power instead of prioritizing control. They might have better recognized that consent is easily extracted through manipulative design at scale, and that exercising any meaningful control is overwhelming in the aggregate, and that our perceived agency is typically illusory in mediated environments.[96] They also might have recognized that the collective wisdom from trillions of individual self-motivated decisions might not reflect or account for collective and societal concerns.

Beyond changing the starting point for analysis of law and technology problems, we think a more developed conceptualization of scale would have three important implications. First, lawmakers should assume that regulator approaches should be continually (or at least periodically) reassessed to confront how the popular adoption of new tools changes costs and benefits. Additionally, we argue that a better conception of scale supports a greater adoption of the precautionary principle. Finally, we argue that scale could shape how legal institutions are designed and the choice of remedies in law and technology disputes.

---

[96] WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES (2018); Woorow Hartzog, *The Case against Idealising Control*, 4 EUR. DATA PROT. L. REV. 423 (2018); Neil Richards & Woodrow Hartzog, *Pathologies of Digital Consent*, 96 WASH. U. L. REV. 1461 (2019); Evan Selinger & Woodrow Hartzog, *The Inconsentability of Facial Surveillance*, 66 LOYALA L. REV. 101 (2019).

When it comes to technology, it's been clear for some time that rules should be periodically revisited. Complex systems scholars have argued that it is functionally impossible to determine ex ante all of the operations that will take place within a ULS system, much less the ways in which parts of the system might fail.[97] Effective management of ULS systems therefore require an increased emphasis on resilience after failures, since complete prevention of failures is impossible.[98] Technologies work within society to change practices and people's perceptions. Expectations and laws that were based on technological practices that existed in 1985 (and business models that leverage those technologies) no longer make sense in 2023. But technological development isn't the only reason our rules related to technology need to be continually updated.

Sometimes it's not clear how scale is different until it manifests. Even when it is clear how law and technology will interact if everyone adopted them, policymakers often do not feel motivated to act upon speculation. But the reality of scale can be compelling, as we've seen with the plague of misinformation and disinformation on social media. Mass deception was always possible with social media, but lawmakers didn't take it seriously until it was widespread enough to be a serious threat to undermining elections at scale (and they arguably have yet to meaningfully respond). Acting upon scale concerns would be a way to interrupt regulatory inertia by requiring a periodical reassessment of the costs and benefits of both rules and tools. It's a way to build policy responses to anticipate that the changes of scale will happen.

Scale being different can also justify a precautionary approach to new technologies. For so long advocates of innovation have criticized early legal intervention where technology is involved because they claim it could hinder the development of new and useful tools. But it is clear now that issues of scale mean that at a certain point, there's no going back. Given the

---

[97] See, e.g., Linda Northrop et al., ULTRA-LARGE-SCALE SYSTEMS: THE SOFTWARE CHALLENGE OF THE FUTURE (Bill Pollak ed., Carnegie Mellon Univ. Software Eng'g Inst. 2006), https://apps.dtic.mil/sti/tr/pdf/ADA610356.pdf.

[98] Nortthup et al., supra note ^; J.B. Ruhl, *Complexity Theory as a Paradigm for the Dynamical Law-And-Society System: A Wake-Up Call for Legal Reductionism and the Modern Administrative State Authors*, 45(5) DUKE L.J. 849 (1996), https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=3307&context=dlj

existential threats digital tools potentially pose to democratic institutions, free expression, privacy, financial security, and the habitability of the planet, there is good reason to apply significant regulatory friction to the development and deployment of technologies to minimize irreparable harm and disempowerment death spirals and ensure innovation is safe and sustainable.

Finally, scale compels important questions about institutional design and legal remedies. Specifically, should policymakers address a problem through regulation, or is it better revolved through private litigation? Additionally, is the best approach to a problem that involves scale to seek monetary relief, or would an injunction address the problems of scale better? Lastly, what about other strategies to enact policy that go beyond liability rules, such as taxation, property interests, and human rights law?

The answers to these questions will vary, and scores of scholars have volumes to say on when and why certain strategies are desirable over others. Our point here is simply to emphasize that the different consequences of scale should be a part of this calculus. For example, if scale changes the population affected by a set of actions to include third parties otherwise unrelated to the relevant actors, then litigation alone might not be the best response because people besides the plaintiffs and the defendants will be affected. This is true even if class action relief is possible. Class actions respond to "scale is more." They simply aggregate the harm of all the class members. There is no obligation in class action lawsuits to address externalities or accommodate unrelated (but incidentally affected) third parties. Scale also might affect the remedies sought in litigation, counseling an injunction that affects everyone potentially affected in the future instead of monetary relief which only directly benefits the plaintiffs.

Issues of scale might also affect the structure and grant of authority to regulatory agencies. If problems only emerge (or appear to emerge) at scale, it's possible that federal agencies might need rulemaking power that doesn't hinge upon a showing of individualized harm. They might also need better information disclosure rules to achieve more transparency, a superstructure to encourage collaboration with researchers to improve expertise, since issues of scale might not be apparent through individualized case studies, past litigation, and anecdotes. Problems that

emerge at scale might also cut across various domains like health, finance, the environment, and consumer protection, necessitating rules to encourage harmony and collaboration, or possibly even a new regulatory agency designed to collect information, provide expertise, and assist with enforcement efforts.[99]

Scale might also direct lawmakers to go beyond the standard suite of regulatory liability rules and embolden property rights to better enable market dynamics (though we remain skeptical of property rights in information as a way to protect people's privacy).[100] Or lawmakers might consider a human rights approach that is less likely to wilt as part of a cost/benefit analysis or political compromise.[101]

Because some problems only manifest at scale, lawmakers might craft legislation that only kicks in at scale. We're already seeing examples of this at the federal and state levels. Senators Elizabeth Warren and Lindsey Graham have targeted "dominant platforms" in legislation that imposes, among other things, robust duties of loyalty, care, confidentiality, and mitigation upon only those businesses that among other things, have more than 50 million US-based monthly active users, 1 billion users worldwide, or an annual revenue of more than $550 billion.[102] If enacted, this law would only affect those operating at the largest scale. A California senator has proposed a legislative framework that would regulate only those "frontier" AI systems that operated at the largest and most robust scale, capturing those problems that exist at the most extreme edges of

---

[99] *See* Ryan Calo, *The Case for a Federal Robotics Commission*, BROOKINGS (2014), https://www.brookings.edu/wp-content/uploads/2014/09/RoboticsCommissionR2_Calo.pdf; Woodrow Hartzog, *Unfair and Deceptive Robots*, 74 MD. L. REV. 785 (2015); *see also* Chris J. Hoofnagle, Woodrow Hartzog, & Daniel J. Solove, *The FTC can rise to the privacy challenge, but not without help from Congress*, BROOKINGS (Aug. 8, 2019), https://www.brookings.edu/articles/the-ftc-can-rise-to-the-privacy-challenge-but-not-without-help-from-congress/.

[100] *See* Pamela Samuelson, *Privacy As Intellectual Property?*, 52 STANFORD L. REV. 1125 (2000); Ignacio Cofone, *Beyond Data Ownership*, 43 CARDOZO L. REV. 501 (2021).

[101] *See* Guido Calebrasi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV 1089 (1972).

[102] Digital Consumer Protection Commission Act of 2023, S. _, 118th Cong. (2023), https://www.warren.senate.gov/imo/media/doc/Tech%20Bill_Full%20Text.pdf.

artificial intelligence and overlooking those that operate at the smallest or more modest scales.[103]

Even better, concerns over scale should encourage lawmakers to look to corporate governance, taxation, and other fiscal approaches to better capture negative externalities of a practice or particular design.[104] For example, Julie Cohen has targeted the dual-class stock ownership structure as one way that individual founders and executives gain and abuse scalable power in the tech sector.[105] Grants, deductions, taxable items, and more all reflect policy preferences that can and should be sensitive to issues of scale. Lawmakers could make it more expensive to use a technology as scale increases or create rules that don't activate until a particular size or different scale threshold is met.

VI. CONCLUSION

Discussions of scale abound in law and policy discussions related to automated technologies. But the concept feels underspecified in ways that might matter. Intuitively, scale means simply "more." But in this essay we've argued scale can also mean "different." More or different

---

[103] Billy Perrigo, *Exclusive: California Bill Proposes Regulating AI at State Level,* TIME (Sep. 13, 2023), https://time.com/6313588/california-ai-regulation-bill/ ("It proposes that systems that require above a certain quantity of computing power to train—a threshold not specified by the bill—be subject to transparency requirements. It proposes establishing legal liability for "those who fail to take appropriate precautions" to prevent unintended consequences and malicious uses of advanced AI systems.").

[104] *See, e.g.*, Julie E. Cohen, Oligarchy, State, and Cryptopia, 94 FORDHAM L. REV. (forthcoming)("[T]ech oligarchs' power derives partly from legal entrepreneurship related to corporate governance and partly from the infrastructural character of the functions the largest technology platform firms now perform."); Salome Viljoen & Amanda Parsons, *Valuing Social Data*, COLUMB. L. REV. (2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4513235. For examples in encouraging innovation, *see, e.g.*, W. Nicholson Price II, *Grants*, 34 BERKELEY TECH. L.J. 1 (2019); Arti K. Rai, Rachel Sachs & W. Nicholson Price II, *Cryptic Patent Reform Through the Inflation Reduction Act*, HARV. J.L. & TECH. (forthcoming 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4402378.

[105] Julie E. Cohen, Oligarchy, State, and Cryptopia, 94 FORDHAM L. REV. (forthcoming)("For high tech ventures that succeed, however, and especially for the dominant tech platform firms, the dual-class ownership structure has thrown a wrench into conventional understandings of corporate governance. The traditional bargain—increased scale in exchange for increased accountability—no longer holds. Dominant tech platform companies seem to make more than the usual number of questionable decisions, engaging in some behaviors that any competent counsel would flag as clearly illegal and others that are, to put it politely, inexplicable from a business standpoint.").

communities might be implicated when people deploy technology at scale. New problems might arise at scale, or we might some assumptions we had once held about the nature of the deployment. Finally, when technologies exist at scale, some legal, social, market-driven, or design-based solutions might become available or be taken off the table. Lawmakers should take scale more seriously and, in doing so, could better respond to the challenges of automated tools.