Vanderbilt Journal of Entertainment & Technology Law

Volume 18
Issue 1 *Issue 1 - Fall 2015*

Article 3

2015

The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles

Meg L. Jones

Recommended Citation

Meg L. Jones, The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles, 18 *Vanderbilt Journal of Entertainment and Technology Law* 77 (2020)
Available at: https://scholarship.law.vanderbilt.edu/jetlaw/vol18/iss1/3

The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles

Meg Leta Jones[*]

ABSTRACT

Rapid developments in sensors, computing, and robotics, including power, kinetics, control, telecommunication, and artificial intelligence have presented opportunities to further integrate sophisticated automation across society. With these opportunities come questions about the ability of current laws and policies to protect important social values new technologies may threaten. As sophisticated automation moves beyond the cages of factories and cockpits, the need for a legal approach suitable to guide an increasingly automated future becomes more pressing. This Article analyzes examples of legal approaches to automation thus far by legislative, administrative, judicial, state, and international bodies. The case studies reveal an interesting irony: while automation regulation is intended to protect and promote human values, by focusing on the capabilities of the automation, this approach results in less protection of human values. The irony is similar to those pointed out by Lisanne Bainbridge in 1983, when she described how designing automation to improve the life of the operator using an automation-centered approach actually made the operator's life worse and more difficult. The ironies that result from automation-centered legal approaches are a product of the neglect of the sociotechnical nature of automation: the relationships between man and machine are situated and interdependent, humans will always be in the loop, and reactive policies ignore the need for general guidance for ethical and accountable automation design and implementation. Like system engineers three decades ago, policymakers must adjust the focus of

[*] Meg Leta (Ambrose) Jones, J.D., Ph.D. is an Assistant Professor of Communication, Culture & Technology at Georgetown University. Many thanks to the We Robot community for their comments, critiques, and enthusiasm, most notably, Michael Froomkin, Woodrow Hartzog, Ryan Calo, Margo Kaminski, and Kate Darling.

legal treatment of automation to recognize the interdependence of man and machine to avoid the ironies of automation law and meet the goals of ethical integration. The Article proposes that the existing models utilized for safe and actual implementation for automated system design be supplemented with principles to guide ethical and sociotechnical legal approaches to automation.

TABLE OF CONTENTS

I. INTRODUCTION

In 1988, USS Vincennes military personnel shot down a passenger jet carrying 290 civilians because their automated radar system, which had been designed to detect Soviet bombers, initially identified the plane as an enemy and none of the crew was willing to challenge the system's determination.[1] The tragic event, and the many that followed, have made us question our reliance on machines.

In 2005, two American amateur chess players beat a supercomputer named Hydra and several teams of grandmasters in an online chess tournament,[2] taking home the $10,000 prize.[3] The "freestyle" tournament allowed anyone to compete alone, in teams, or with computers. The humans and machine teams dominated the supercomputers operating the same brute number-crunching strategies in place since the 1970s. While the amateurs were far less skilled than the grandmasters at chess strategy, they were far more skilled with their computers. "Weak human + machine + better process was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process."[4] Pairing a human with a machine can significantly increase desired performance beyond that which could be achieved by man or machine separately. The line between achieving new feats and catastrophic losses must be toed carefully, but so goes the reality of technological innovation—technology is neither good nor bad. Nor is it neutral. This is the first of Melvin Kranzberg's "Six Laws of Technology," or "a series of truisms . . . deriving from a longtime immersion in the study of the development of technology and its interactions with sociocultural change."[5] The second law is that invention is the mother of necessity, and the third is that technology comes in big and small packages.[6] The fourth is that non-technical factors take precedence in technology-policy decisions, and the fifth is that "all history is relevant, but the history of technology is the most relevant."[7] His sixth law is technology is human-centric.[8] Each of these truths is either difficult to remember or difficult to realize in everyday practice. It is arduous not to succumb to technological

[1] David Evans, *Vincennes: A Case Study,* 119:8 Proceedings 49 (1993).

[2] *Dark Horse ZackS Wins Freestyle Chess Tournament,* ChessBase Chess News (June 19, 2005), http://en.chessbase.com/post/dark-horse-zacks-wins-freestyle-che-tournament [http://perma.cc/EX92-VQJY].

[3] *Freestyle Tournament for $20,000,* ChessBase Chess News (May 9, 2005), http://en.chessbase.com/Home/TabId/211/PostId/4002379/freestyle-tournament-for-20-000.aspx [http://perma.cc/C7QK-HVYV].

[4] Garry Kasparov, *The Chess Master and the Computer,* N.Y. Rev. of Books (Feb. 11, 2010), http://www.nybooks.com/articles/archives/2010/feb/11/the-chess-master-and-the-computer/?pagination=false [http://perma.cc/RBA8-F328].

[5] Melvin Kranzberg, *Technology and History: "Kranzberg's Laws,"* 27:3 TECHNOLOGY AND CULTURE 544, 544 (1986).

[6] *Id.* at 548.

[7] *Id.* at 549.

[8] *Id.* at 557.

determinism—thinking that "technology is the prime factor in shaping our lifestyles, values, institutions, and other elements of our society"[9]—as we are faced with the "need" to continually adapt to new forms of communicating with friends and family, upgrade our organizations and skillsets to be competitive, and act upon values never collaboratively established. The big, small, and connected packages create complex ecologies that are difficult to navigate or evaluate, but complex, international problems seem solvable by the optimistic and computationally minded. In the face of extraordinary and overwhelming technological integration, we must remind ourselves of Marshall McLuhan's wise words: "There is absolutely no inevitability as long as there is a willingness to contemplate what is happening."[10] Technology law scholars have taken different approaches to contemplate what is happening. Many, if not most, take a problem-based approach to sociotechnical issues. They identify a new sociotechnical harm, breakdown and describe the relevant technology, describe the shortcomings of existing law in addressing the new harm, and propose changes.[11] Others bravely engage in debate over exceptionalism and the need to overhaul policy due to new technology.[12] A few have investigated "the pacing problem"—law's inability to keep up with accelerating technological change.[13] Still others look back in history to find corollaries and lessons from the past—no easy task when the topic is technology.[14] This Article takes a different perspective, borrowing a bit from each approach. Instead of drawing lines that distinguish the many technology innovations entering the world from "big data"[15] analytics to sophisticated robots, it ties them together under the umbrella of automation. Although it may have once made sense to focus on information technologies that offered a virtual existence or stationary robots isolated on factory floors, the issues that arise from these technologies are merging as

[9] *Id.* at 545.

[10] MARSHALL MCLUHAN & QUINTEN FIORE, THE MEDIUM IS THE MASSAGE: AN INVENTORY OF EFFECTS 25 (1996).

[11] *See* A. Michael Froomkin & Zak Colangelo, *Self-Defense Against Robots and Drones*, 48 CONN. L. REV. 1 (forthcoming 2015).

[12] *See, e.g.*, Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. (forthcoming 2015).

[13] *See, e.g.*, THE GROWING GAP BETWEEN EMERGING TECHNOLOGIES AND LEGAL-ETHICAL OVERSIGHT: THE PACING PROBLEM 19 (Gary E. Marchant, Braden R. Allenby & Joseph R. Herkert eds., 2011) [hereinafter GROWING GAP].

[14] *See, e.g.*, TIM WU, The Master Switch 19 (2010).

[15] "Big data" is most often defined by its three Vs: velocity, variety, and volume. Sometimes a fourth V, veracity, is included to describe the uncertainty associated with big data. The term refers to the collection and use of very large amounts and types of information that are produced and flow in at high rates to be analyzed and shared for various purposes.

robots and screens attain intelligence and gain mobility. This Article makes no comment on exceptionalism or the "newness" of new technology; instead, it takes a step back and frames emerging technologies in such a way to allow for both reflection and prediction. By categorizing modern digital man-machine systems as automation, one can look back, as we have always automated life, and inform the future, wherein man-machine systems will change—but will ultimately remain man-machine systems.

Through the broad lens of automation, reflection reveals the law has not been particularly good at toeing the line between achieving new feats and causing catastrophic losses. The law has ignored the delicate interdependence between man and machine, resulting in ineffective protection of specified value. Five case studies of legal approaches to automation from across various contexts expose an irony. When presented with an automation-related problem, law and policy responses have been to preserve or protect an explicit value by simply inserting or removing a human from the loop, which actually ends up backfiring. The value stated by policymakers is more vulnerable than before legal intervention occurs or the intended goals of the policy are left further out of reach. Most automation used today already has a well-established place in the world and has developed ethical and legal treatment. The challenge is how to approach emerging digital man-machine systems that carry with them so much uncertainty and so much promise.

Sheila Jasanoff has called for a reexamination of human control over technological systems in light of uncertainty and unpredictability.[16] Critical of American theorists that view technological failings as avoidable error, Jasanoff finds the work of German sociologist Ulrich Beck more encompassing.[17] Beck's thesis of "reflexive modernization" argues that risk is an inherent part of a technically intensive society and describes risk as part of the modern human condition, not cold, rationally calculated probabilities.[18] Jasanoff argues for the development of a set of "technologies of humility" or:

> methods, or better yet institutionalized habits of thought, that try to come to grips with the ragged fringes of human understanding[—]the unknown, the uncertain, the ambiguous, and the uncontrollable. Acknowledging the limits of prediction and control, technologies of humility confront 'head-on' the normative implications of our lack of perfect foresight. They call for different expert capabilities and different forms of

[16] Sheila Jasanoff, *Technologies of Humility: Citizen Participation in Governing Science*, 41:3 MINERVA 223, 223–24 (2003).

[17] *Id.* at 224.

[18] *See e.g.*, ULRICH BECK, RISK SOCIETY: TOWARDS A NEW MODERNITY 22 (1992).

> engagement between experts, decision-makers, and the public than were considered needful in the governance structures of high modernity.[19]

Seeking to democratize the development and governance of technology, Jasanoff provides a framework to promote more meaningful interaction between corporate parties, policymakers, technical experts, and the public.[20] In order to achieve rich, democratized deliberations, participation should focus on framing the issues broadly enough to encompass more than cost-benefit analyses. This participation should consider vulnerabilities of ordinary citizens that do more than reduce those people to groups and populations based on shared categorical characteristics, distribution of consequences not only with an ethical conversation at the beginning but throughout the development and deployment of technology, and learning from the various interpretations of technological integrations through collective reflection and assessment of alternative explanations.[21] By treating technology incrementally and broadening the scope of consideration, this Article adjusts the framing of the issue and attempts to provide a policy approach that is inclusive and adaptive.

Before an emerging technology has sufficiently taken hold in society, it is difficult to know what it is capable of and how it will be used. While its technical capabilities may be known, others—as well as users—may quickly adapt those capabilities in the market. What innovative or nefarious uses the technology will facilitate are also hard to imagine. These two unknowns make governance of emerging technology quite challenging and lead to a pacing problem, in which technological innovations outpace ethical and legal developments intended to direct design and use.[22] Traditional prescriptive regulations can be too restrictive to allow for the flexibility required in the design and implementation of human-automation systems and responsible situated use.[23] Policy innovations, like delegation and self-governance, have developed to address a quickly evolving technological landscape.[24] These innovations have left many frustrated, arguing the lack of protection and guidance is a detrimental and deterministic approach to innovation. This Article argues that any legal treatment must not mess with a well-formed loop if the irony presented in the case studies is to be avoided.

[19] Jasanoff, *supra* note 16, at 227.

[20] *Id.* at 238–42.

[21] *Id.*

[22] GROWING GAP, *supra* note 13.

[23] *Infra* Section IV.

[24] Kenneth A. Bamberger, *Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State*, 56 DUKE L.J. 377, 386–87 (2006).

Instead, a "policy knot"[25] should be formed that utilizes existing "good" design principles, enriches underdeveloped areas, and identifies and fills important holes in the creation, integration, and implementation of automation. Applying the policy-knot approach to automation (specifically, automated decision making and self-driving cars) within the framework of Jasanoff's technologies of humility, a set of seven principles, the Fair Automation Practices Principles (FAPPs), is derived. Using existing principles from automation design, human-robot automation, and information policy, the FAPPs state that automation design and use should involve: (1) informed risk assessment, (2) transparent processes, (3) error detection and correction, (4) consideration of sensitive situations, (5) diversity and discrimination testing, (6) man and machine reallocation comparisons, and (7) an inventory of the predictable and unpredictable.

Roboticist Illah Nourbakhsh, at the Carnegie Mellon Robotics Institute, made a fairly big request of the law in his book *Robot Futures*:

> Instead of reacting case by case to new loopholes in law discovered by ever more ingenious machines, our system of jurisprudence must proactively gather the expertise and wherewithal to predict our robot future, debate the most critical issues of safety, accountability, equality, and quality of life, and create a viable legal framework for this century. Not only would such an exercise provide guide rails for future robot engineers and businesses, it would also catalyze a public awareness that we are entering an uncharted space but are girding ourselves with knowledge and moral authority to make sense of our future.[26]

This Article will attempt to contribute some guide rails. It provides both a broad description of automation that allows many overlapping technologies to be discussed simultaneously and an expansive overview of automation design. The Article then makes three novel contributions to an already-novel subject area. The first contribution is a reflection on legal approaches to automation in the United States; the Article identifies an existing and ill-conceived approach to governing automation that focuses on the capabilities of the day's automation. The second contribution is a legal approach to emerging technologies that recognizes the complex relationship between man and machine: policy knots. Using the policy knot approach, this Article includes a preliminary series of Fair Automation Practices Principles based on existing automation design principles, missing guidance, and recently identified issues.

[25] Steven J. Jackson, Tarleton Gillespie & Sandy Payette, *The Policy Knot: Re-integrating Policy, Practice and Design in CSCW Studies of Social Computing*, in Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (2014); *infra*, Section V.

[26] ILLAH REZA NOURBAKHSH, Robot Futures 117 (2013).

II. AUTOMATION IN THE DIGITAL AGE

By focusing on man-machine systems, the category of emerging technology covered in this Article is exceptionally broad, encompassing the many automated tools used in everyday life as well as those sophisticated, networked, and "intelligent" technologies and spaces on the horizon. The subject matter is narrow enough to exclude three of the other GRINN[27] technology categories: genetic engineering, neuroscience, and nanotechnology.

Broadly, automation includes all the ways computers and machines help people perform tasks more quickly, accurately, and efficiently. The term "automation" refers to: (1) the mechanization and integration of the sensing of environmental variables through artificial sensors, (2) data processing and decision making by computers, and (3) mechanical action by devices that apply forces on the environment or information action through communication to people of information processed.[28] The term encompasses open-loop operations[29] and closed-loop control,[30] as well as intelligent systems.[31] Leading automation designer Thomas Sheridan explains the transition across these concepts:

> Computers have continued to become smaller, faster, more powerful and cheaper. Automation has moved from open-loop mechanization of the industrial revolution, then to simple closed-loop linear control, then to non-linear and adaptive control, and recently to a mix of crisp and fuzzy rule-based decision, neural nets and genetic algorithms and other mechanisms that truly recognize patterns and learn.[32]

Older automation was not mobile, held minimal purpose-specific sensors, and operated with limited processing power, but ubiquitous computing means ubiquitous automation of many functions of many tasks.[33] Harvard business professor James R. Bright was correct in 1958 when he said:

[27] GRINN is an acronym used to refer to genetic engineering, robotics, information technology, neuroscience, and nanotechnology.

[28] THOMAS B. SHERIDAN, HUMANS AND AUTOMATION: SYSTEM DESIGN AND RESEARCH ISSUES 9–10 (2002).

[29] Open loop controls have no measurement of system output or feedback.

[30] In a closed loop control system, the output is monitored and fed back to a control to make adjustments.

[31] Intelligent systems can be defined as autonomous systems with intelligence or achieving intelligent behavior through computation. ROBERT J. SCHALKOFF, INTELLIGENT SYSTEMS: PRINCIPLES, PARADIGMS, AND PRAGMATICS 1 (2009).

[32] Thomas B. Sheridan, *Function Allocation: Algorithm, Alchemy or Apostasy?*, 52 INT'L J. HUMAN-COMPUT. STUDIES 203, 205 (2000).

[33] Raja Parasuraman and Christopher D. Wickens, *Humans: Still Vital After All These Years of Automation*, 50 HUMAN FACTORS 511, 511–12 (2008).

> [Automation] has been used as a technological rallying cry, a manufacturing goal, an engineering challenge, an advertising slogan, a labor campaign banner, and as the symbol of ominous technological progress. . . . Automation simply means something significantly more automatic than previously existed in that plan, industry, or location.[34]

Although automation is continual, the nature of automation today, from its ubiquity to its intelligence to its import, has taken on a new set of characteristics worthy of evaluation and reflection. Automation is never really old or new but is getting another close look in light of new technological advances. Nicholas Carr has recently written a book dedicated to the subject of automation—"about the use of computers and software to do things we used to do ourselves."[35] Today's automation is characterized by data collected through sensors and ever-advancing algorithms processing that data. This development has a powerful impact on the world. Digital automation has crept into every facet of life. We have automated decisions about whether to delete emails, who to date, who to hire, what movies to watch, what search terms to enter, what to eat, where to drive, where to shop, when to sleep, when to send birthday greetings, and how to go through our days.[36] It is not just seemingly mundane daily tasks that we have augmented with computational support. Human resources departments, government agencies, school districts, parole boards, and Medicaid administrators rely on big data to make determinations about individuals and resources. Digital automation tells us what is important and whether we are important.

Digital automation utilizes elegant algorithms to process piles and piles of data to some end. The algorithm itself has recently come under scrutiny as a powerful force that can dictate interests and actions. As part of a conference entitled "Governing Algorithms" held at New York University in 2013, Tarleton Gillespie defined algorithms as "encoded procedures for transforming input data into a desired output, based on specified calculations."[37] Gillespie argued that algorithms produce and certify knowledge in a world of ubiquitous computing. Frank Pasquale's book *The Black Box Society* details the way in which these deliberately concealed mathematical processes shape our reputations, knowledge inquiries, and financial existence.

[34] NICHOLAS CARR, THE GLASS CAGE: AUTOMATION AND US 3435 (2014) (CITING JAMES R. BRIGHT, AUTOMATION AND MANAGEMENT 4–5 (1958)).

[35] *Id.* at 1.

[36] *Id.*; *see e.g.*, FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 34–35 (2015); *see also* CHRISTOPHER STEINER, AUTOMATE THIS (2012).

[37] Tarleton Gillespie, *The Relevance of Algorithms*, *in* MEDIA TECHNOLOGIES (Tarleton Gillespie, Pablo Boczkowski & Kirsten Foot eds., 2014), http://governingalgorithms.org/wp-content/uploads/2013/05/1-paper-gillespie.pdf [http://perma.cc/4UDK-77G9].

For example, job candidates are coded and scored based on everything from their application materials to their online networks.[38] Once hired, emails are monitored for insights into productivity and teamwork; even smiles can be "datafied" and digitized for algorithmic purposes.[39] Google's algorithms have revolutionized inquiry in the twenty-first century. It is Google's algorithms (which represent the new "Coke recipe" of trade secret examples)[40] that initially gave the company so much value. A Google search gives an inquirer results not only based on how the words she entered match words in particular webpage content but also on who she is relative to who others are, where she has been, and where she will be. Banks have always used numbers and prediction to make financial decisions, but every aspect of banking today is automated. Paychecks are automatically deposited, accounts are monitored for fraud, credit is extended without speaking to anyone, and investments move in less than seconds. The algorithm is a vital piece (but only a piece) of the digital automation process. In fact, much of big data progress has been made possible due to big algorithms.[41]

Professionals are incorporating digital automation to make work more efficient and precise. Pilots are the reference of choice, the implementation of autopilot coming under fire whenever planes crash.[42] Doctors have been the focus a similar conversation in light of expert systems like IBM's Watson (now being used to diagnose and

[38] Pasquale, *supra* note 36, at 34.

[39] *Id.*

[40] *See* Bruce Horovitz, "Buzz Surrounds Relocation of Coke's Secret Formula," USA Today (Dec. 8, 2011), http://usatoday30.usatoday.com/money/industries/food/story/2011-12-08/secret-formulas-as-pr/51751328/1 [http://perma.cc/PAM2-TZ5Z]; National Institute of Standards and Technology, Trade Secrets Protection in the U.S., http://www.nist.gov/mep/upload/marinaslides.pdf [http://perma.cc/5S7M-A32Q] (Coke recipe and Google's PageRank included on the "Examples of Trade Secrets" slide).

[41] Jonathan Shaw, *Why 'Big Data' is a Big Deal*, HARV. MAG. (Mar.–Apr. 2014), http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal [http://perma.cc/63HK-VUFW].

[42] *Final Air France Crash Report says Pilots Failed to React Swiftly*, CNN (July 5, 2012), http://www.cnn.com/2012/07/05/world/europe/france-air-crash-report/ [http://perma.cc/P4H5-RCTT]; *'Ghost Flight' Horror Crash Blamed on Pilots*, DAILY MAIL (Oct. 10, 2006), http://www.dailymail.co.uk/news/article-409703/Ghost-flight-horror-crash-blamed-pilots.html [http://perma.cc/C3F9-XRD9];Jacob Kastrenakes, *Asiana Airlines Crash in San Francisco Blamed on Overuse of Autopilot*, THE VERGE (June 24, 2014), http://www.theverge.com/2014/6/24/5838072/asiana-airlines-flight-214-crash-autopilot-issues-at-fault-ntsb-finds [http://perma.cc/83QE-DSYP]; Calum MacLeod, *Authorities: Flight 370 on Autopilot When it Crashed*, USA TODAY (June 26, 2014), http://www.usatoday.com/story/news/world/2014/06/26/mh370-search-shifts-south/11392315/ [http://perma.cc/TVV5-Y7SS].

treat patients)[43] and surgery robots like da Vinci.[44] Other professions that have high risks, albeit not life-threatening risks, and recently incorporated significant amounts of digital automation include lawyers and traders. High frequency trading has reshaped Wall Street, transforming the skills and know-how of stock traders.[45] As one analyst bluntly explained, "All [traders] do today is hit buttons on computer screens."[46] Legal practice has long been digitized, dominated by the WestLaw and LexisNexis legal databases, but the practice of law has recently gotten "smarter." E-discovery software performs critical document review in search of evidence, synthesizes material, and details connections between events and people. Lex Machina and other software developments that predict outcomes and strategies for specific cases[47] may soon present the courtroom equivalent of Deep Blue.[48]

Streets and airspaces are full of automated systems as well and will soon hold more automated systems—both in quantity and quality. Drones were a hot Christmas gift in 2014—so much so that the FAA issued a statement asking recipients to fly with care.[49] Already in use by law enforcement, these increasingly flyer-friendly devices will soon be taking to the air to deliver packages and capture news, as well as to monitor everything from crowds to crops. As drones are paced to take the air, self-driving vehicles are poised to take the streets. Even moving by foot or public transportation has been transformed by personal devices carried in our pockets and networked buses, trains, and ferries. Finding and moving ourselves around the world

[43] Jonathan Cohn, *The Robot Will See You Now*, THE ATLANTIC (Mar. 20, 2013), http://www.theatlantic.com/magazine/archive/2013/03/the-robot-will-see-you-now/309216/ [http://perma.cc/FH9M-U4B6].

[44] Roni Caryn Rabin, *New Concerns on Robotic Surgeries*, N.Y. TIMES, (Sept. 9, 2013), http://well.blogs.nytimes.com/2013/09/09/new-concerns-on-robotic-surgeries/ [http://perma.cc/ZXE2-ZSHH].

[45] CARR, *supra* note 34, at 115–16.

[46] CARR, *supra* note 34, at 115 (citing Max Raskin and Ilan Kolet, *Wall Street Jobs Plunge as Profits Soar*, BLOOMBERG NEWS (Apr. 23, 2013, 11:01 PM), http://www.bloomberg.com/news/2013-04-24/wall-street-jobs-plunge-as-profits-soar-chart-of-the-day.html [http://perma.cc/ZKA5-G3ZM].

[47] *Id.* at 116.

[48] Legendary chess player Garry Kasparov played IBM chess program Deep Blue in a set of high profile challenges, famously winning the first and losing the second. Rich McCormick, *How a Computer Error Helped Deep Blue Beat Humanity's Best Chess Player*, THE VERGE, (Oct. 24, 2014), http://www.theverge.com/2014/10/24/7056493/how-a-computer-error-helped-deep-blue-beat-humanitys-best-chess-player [http://perma.cc/2SQ4-KB5N].

[49] Craig Whitlock, *Drones for Christmas Worry the FAA*, WASH. POST, (Dec. 22, 2014), http://www.washingtonpost.com/world/national-security/drones-for-christmas-worries-the-faa/2014/12/22/b4f0bd2a-8a02-11e4-a085-34e9b9f09a58_story.html [http://perma.cc/N49X-7CDC].

efficiently is possible anytime and anywhere thanks to digital automation.

III. IRONIES OF AUTOMATION

There are downsides to automation that we choose to endure because of the perceived benefits. There are also unintended consequences that were not foreseen or considered when the automation was engineered. And even when automation is intended to make a specific thing better, it can make that very thing worse. Before delving into legal treatment of automation, it is important to understand not only the definition of and potential for automation, but also its complex man-machine nature—the nature that leads to ironies. A number of engineers, scholars, and commentators have made strides in uncovering this nature and are highlighted in this Section.

A. The Dark Side of Automation

"If computers' abilities are expanding so quickly and if people, by comparison, seem slow, clumsy, and error-prone, why not build immaculately self-contained systems that perform flawlessly without any human oversight or intervention?"[50] The short answer is that automation can be flawed, break, and have widespread detrimental effects. There are downsides to supplementing tasks and processes with mechanical or computational automation. Carr's book details a number of studies on the impact of automation, many of which conclude "sharp tools, dull minds."[51] Shifting mental and physical tasks, memory, and analysis has been the practice of humans since writing on cave walls, but our brains, bodies, and expectations are altered by the shift. Does anyone know how to Shepardize[52] a case in print anymore? When was the last time you drove a manual transmission car? These tasks were always supported by automation to some extent, but it is important to recognize that relying on automation can make us less capable generally.

[50] Nicholas Carr, *All Can be Lost: The Risk of Putting Our Knowledge in the Hands of Machines*, THE ATLANTIC, (Oct. 23, 2013), http://www.theatlantic.com/magazine/archive/2013/11/the-great-forgetting/309516/ [http://perma.cc/9REE-H7QF].

[51] CARR, *supra* note 34, at 78 (citing Vivek Halder, *Sharp Tools, Dull Minds*, THIS IS THE BLOG OF VIVEK HALDAR, http://blog.vivekhaldar.com/post/66660163006/sharp-tools-dull-minds [http://perma.cc/CN5T-TZL5]).

[52] "Shepardize" refers to the process of consulting Shepard's Citation Service to see if a case has been overturned, reaffirmed, questioned, or cited by later cases. *Shepard's Citations Service*, LEXISNEXIS, http://www.lexisnexis.com/en-us/products/shepards.page [http://perma.cc/4FFZ-G5Z2].

Reliance on automation may have a general dulling effect on our minds and bodies, but automation can also have significant and inherent flaws, producing more immediate and significant harms. In his book *To Forgive Design*, Henry Petroski explains that technologies break because innovations are commissioned, funded, designed, built, and maintained by humans and that we overestimate the reliability and capabilities of technology.[53]

> We humans first conceive of and design even the most autonomous systems, and we inadvertently invest our human limitations in them. . . . [T]hose people whom we call inventors, designers, and engineers set out to achieve what they perceive to be a good end and to do it with as much care and dedication as they are capable of mustering. The creators, maintainers, and operators of technology are by and large capable and committed individuals and teams who above all else want their creations and plans and charges to succeed. That they sometimes fail is but testimony to the humanness of the people involved.[54]

In her 2008 article *Technological Due Process*, Danielle Citron outlined the problems in converting government policy to code and automating agency decisions like distributing welfare benefits or excluding individuals from air travel.[55] The programmers' bias was represented in a Colorado public benefit system that required workers to enter whether a potential welfare recipient was a "beggar" and the federal Parent Locator Service that identified individuals as "dead-beat" parents.[56] Even something as seemingly neutral as maps and location are fraught with subjectivity and choices—and always have been. Maps play a large role in how we understand the world, but maps have always been about power, plagued with mistakes, and the product of human choices.[57] In spite of this subjectivity and lack of reliability, we may rely on digital automation using GPS to get us where we need to go effectively and efficiently, even when it sends us into lakes[58] or off-road.[59] As with all technological innovation, we must be both thoughtful about our choices to utilize automation and aware of its limitations.

[53] *See generally* HENRY PETROSKI, TO FORGIVE DESIGN (2012).

[54] *Id.* at 23–24.

[55] *See* Danielle Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1252 (2008).

[56] *See id.* at 1257, 1280.

[57] Aleks Krotoski, *Digital Humans: Maps*, BBC RADIO 4 (Nov. 10, 2014), http://www.bbc.co.uk/programmes/b04nrmgs [http://perma.cc/89HQ-UA7Y].

[58] *See* Jesus Diaz, *Man Drowns After GPS Guides Him Into a Lake*, GIZMODO (Oct. 4, 2010, 7:13 PM), http://gizmodo.com/5655527/man-drowns-after-gps-guides-him-into-a-lake [http://perma.cc/CVA5-SXTY].

[59] *See* Casey Chan, *This is What Happens When the GPS is Wrong*, GIZMODO (Oct. 2, 2010, 7:00 PM), http://gizmodo.com/5654044/this-is-what-happens-when-the-gps-is-wrong [http://perma.cc/AW93-E27W].

B. The Human in the Loop

In 1983, Ruth Schwartz Cowan's *More Work for Mother* traced the way in which technology shifted the burden of domestic labor from men and boys to mothers and wives over the last three centuries.[60] "Advances" in the industrialization of household technologies to make domestic work processes easier did not actually, according to Cowan, diminish the amount of work women had to perform.[61] Cast iron stoves liberated men from having to chop wood, and municipal water meant men did not have to haul water.[62] Advances in technology were made with the intent of lightening the load of housework, but domestic tasks were left entirely to women, whose workload was changed, but not lightened. Edward Tenner collected and synthesized innovations like those reflected upon by Cowan—innovations that created what he calls revenge effects, those with unforeseen and unpleasant consequences.[63] Instead of solving problems, or in addition to solving them, problems are often simply spread out in space and time, creating chronic issues.[64] For example, carelessness and accidents are more common due to new safety devices, and tougher insects and diseases came about from better pesticides.[65] A relevant chapter of Tenner's work includes his claim that our computational improvements, meant to significantly reduce the strain of physical and mental tasks, are complex and actually require a great deal of vigilance.[66] This is because our technologies are so often complex, imperfect, and frequently break. We have to remember to save files on numerous devices in various places and replace batteries in smoke detectors. Overlooking the "humanness" of the human in the loop rarely serves that human well.

In 1983, Bainbridge succinctly described the ironies of automation. The automation designer, a human, automates what she can under the theory that the human is unreliable and inefficient.[67] This is the first irony, of course, because as a human, the designer is unreliable and inefficient. She delegates the easy tasks of the automation operator, a human, to an automated process, making the

[60] *See generally* RUTH SCHWARTZ COWAN, MORE WORK FOR MOTHER: THE IRONIES OF HOUSEHOLD TECHNOLOGY FROM THE OPEN HEARTH TO THE MICROWAVE (1985).

[61] *See id.* at 12.

[62] *See id.*

[63] EDWARD TENNER, WHY THINGS BITE BACK: TECHNOLOGY AND THE REVENGE OF UNINTENDED CONSEQUENCES (1997).

[64] *Id.*

[65] *Id.*

[66] *Id.*

[67] Lisanne Bainbridge, *Ironies of Automation*, 19:6 AUTOMATICA 775 (1983).

difficult aspects more difficult. The human in the loop is left with hard and unpleasant tasks, those that could not be automated, and automation errors and failures. The second irony is then that the automation designer intends to make the life of the operator easier and better, but by focusing on automation capabilities, the designer makes the operator's life more difficult.

These ironies result from relegating the human to a monitor and a safeguard, a responsibility that even the most motivated human will have problems maintaining vigilance toward. Rare, abnormal conditions are difficult to detect when inappropriate deference and trust of the machine (automation bias)[68] builds in human operators interacting with a well-functioning system. Additionally, a decline in the perception of environmental elements and system functioning (situational awareness)[69] occurs in the operator. When inevitable errors occur, the operator's inappropriate reliance upon the automation (complacency)[70] and lack of sharp skills (skill degradation)[71] result in a decreased ability to perform when needed. These human-automation interaction concepts are discussed in greater detail below, but the true irony is that the most successful automation systems—those that fail and cause need for manual intervention on the rarest occasions—require the greatest investment in operator training. In short, the more advanced and reliable the automation, the more important the human operator must be.

Our quest for more, better, and faster should also be adaptive, diverse, and reflective, according to Tenner, who argues that we must recognize revenge effects and act on them early.[72] Petroski echoes this sentiment, explaining: "Successful change comes not from emulating success and trying to better it, but from learning from and anticipating failure, whether actually experienced or hypothetically imagined."[73] The next Section analyzes the way in which law may or may not support recognition of these larger social goals and the human in the loop.

[68] Kathleen L. Mosier, Linda J. Skitka, Susan Heers, & Mark Burdick, *Automation Bias: Decision Making and Performance in High-Tech Cockpits*, 8:1 INT'L. J. AVIATION PSYCHOLOGY 47, 47 (1998); *infra* Section V.C.

[69] Mica R. Endsley, *Automation and Situation Awareness, in* Automation and Human Performance: Theory and Application 163, 163–65 (Raja Parasuraman & Mustapha Mouloua, eds.,1996); *infra* Section V.C.

[70] Raja Parasurman, Robert Molloy, & Indramani L. Singh, *Performance Consequences of Automation-Induced "Complacency,"* 3:1 INT'L. J. AVIATION PSYCHOLOGY 1 (1993); *infra* Section V.C.

[71] Earl L.Wiener & Renwick E. Curry, *Flight-Deck Automation: Promises and Problems*, 23:10 ERGONOMICS 995 (1980).

[72] Tenner, *supra* note 63, at 8, 107.

[73] Petroski, *supra* note 53, at 329.

IV. THE IRONY OF AUTOMATION LAW

This Article examines automation in particular because it covers an array of sociotechnical systems that are both old and broad enough to gain insight from the past. Most importantly, it is not possible to look back on prior robot or artificial intelligence law, but policy approaches to automation continue to be relevant through pressing policy efforts to manage social concerns surrounding the datafied, algorithmic, and robotics "revolutions." In this Section, the results of investigating five case studies are presented. They reveal an irony of automation law, named for its resemblance to the ironies of automation.

The case studies below are old enough to give some sense of the effectiveness of the law. Accordingly, they do not include new or proposed automation regulation (such as those applicable to domestic commercial drones or automated trading in financial markets) but are intended to inform current and future regulatory debates. Each involves an overwhelmingly complex area of law, social context, and technological innovation and is only touched upon briefly. While the chosen cases may not reflect a general policy trend toward automation, the cases do reveal an approach to automation that should be avoided: an automation-centered approach. Following the extraction of this approach from the different examples is a discussion of reasons that lead to the flawed outcomes from an automation-centered approach and a proposal for a more suitable, sociotechnical policy approach to automation.

A. Out of the Loop

When accidents happen or bias or abuse occurs, a mechanical fix is a tempting solution. Removing the human, this line of thinking goes, will remove the subjectivity, the errors, the inexactness, and the carelessness. This result is neither possible nor desirable and approaches automation as if it has had no negative consequences.

1. Legislating Railroad Safety

While developments in robotics are certainly driving regulatory conversations, Congress was regulating automated mechanisms as far back as 1893. In that year, Congress passed the Safety Appliance Act, which required railroads to place automatic couplers on all freight cars over a period of five years.[74] Implemented in 1904, the Act was

[74] *See* Safety Appliance Act of 1893, 27 Stat. 531 (repealed 1994).

intended to reduce the staggering injuries and deaths suffered by railroad workers. In 1894, one in 428 employees was killed, and one in thirty-three was injured, totaling 25,245 employees killed or injured. After compliance with the automatic couplers and brakes was established, one in 357 employees died, and one in nineteen was injured.[75] The increase was a dramatic blow to those that felt railways represented an important progress for the United States and that safety could be achieved for this innovation. Policymakers saw humans being injured by a task that could be automated; therefore, they automated it. However, they failed to recognize how humans were interacting with the cars and each other to achieve objectives and would need to do so with the automated additions.

The Accident Reports Act was passed by President Taft in 1910 to better evaluate the effectiveness of railway safety measures.[76] Additional issues, including the identification of safety hazards and defects, were addressed in the Safety Appliance Act of 1910, which required standards for equipment, practices, and inspection.[77] Safety First programs were then initiated by Chicago and North Western Railway in 1910.[78] By 1918, all Class I railroads were required by legislation to adopt similar programs.[79] Statistics in employee injuries and fatalities began to improve after the initial increase, dropping by 75 percent from 1920 to 1940.[80] After a more comprehensive approach was adopted to address railroad employee safety and companies started taking an active role in decreasing injuries, the human in the loop became a pivotal part of the regulatory equation.

2. Adjudicating Warrantless Searches

Other laws, regulations, and rights that do not mention automation explicitly are interpreted as regulating the human in the loop by the judicial system. Whether a human is required to observe or receive information disclosed by an individual so that the individual loses her expectation of privacy (and associated rights) is an important aspect in debates surrounding Fourth Amendment privacy rights.[81]

[75] S. W. Usselman, *The Lure of Technology and the Appeal of Order: Railroad Safety Regulation in Nineteenth Century America*, 21 BUS. & ECON. HISTORY 290 (1992).

[76] *See* Accident Reports Act of 1910, Pub. L. No. 62-165 (1910) (codified as 49 U.S.C. §§ 20901–03 (2015)).

[77] *See* Safety Appliance Act of 1910, 36 Stat. 298 (repealed 1994).

[78] IAN SAVAGE, THE ECONOMICS OF RAILROAD SAFETY 23 (1998).

[79] *Id.*

[80] *Id.*

[81] *See, e.g.*, Kevin S. Bankston & Amie Stephanovich, *When Robot Eyes Are Watching You: The Law & Policy of Automated Communications Surveillance*, 2014 WE ROBOT CONF., http://robots.law.miami.edu/2014/wp-

No "search" by government agents necessarily occurs until information is exposed to a human being. In other words, a human is required to be in the loop for a search to have been performed, meaning a machine alone cannot violate one's right to privacy.[82] In *United States v. Karo*, the Court explained that a defendant's acceptance of a container with a hidden homing beacon did not invade his privacy, but the monitoring of the information by the agents later was an invasion.[83] The Court explained:

> [W]e have never held that potential, as opposed to actual, invasions of privacy constitute searches for purposes of the Fourth Amendment. A holding to that effect would mean that a policeman walking down the street carrying a parabolic microphone capable of picking up conversations in nearby homes would be engaging in a search even if the microphone were not turned on. It is the exploitation of technological advances that implicates the Fourth Amendment, not their mere existence.[84]

Additionally, the Supreme Court held in *Kyllo v. United States* that the use of thermal imaging to detect infrared heat waves was a search, not because the Fourth Amendment extended protection to heat waves outside a home, but because the use of thermal imaging technology allowed human agents to infer activities within the home.[85]

In Fourth Amendment jurisprudence, the line is drawn between man and machine: the machine is relied upon as less invasive and protecting dignity—the opposite determination established for robocalls, as discussed below. By focusing on the capabilities of a fully automated information system in the 1980s, the Court determined that a human must be in the loop for a reasonable expectation of privacy to be invaded, but today's rampant, fully-automated data collection schemes have left citizens vulnerable to incredibly granular, widespread, and systematic invasions. In June 2013, a National Security Agency (NSA) program called PRISM was brought to the public's attention revealing the government collection of metadata

content/uploads/2014/07/Bankston_Stepanovich_We_Robot.pdf [http://perma.cc/FTW8-5LD5]; *see* Orin Kerr, *Applying the Fourth Amendment to the Internet: A General Approach*, 62 STAN. L. REV. 1005, 1009–12, 1032–38 (2010); Orin Kerr, *Searches and Seizure in a Digital World*, 119 HARV. L. REV. 531 (2005); Matthew J. Tokson, *Automation and the Fourth Amendment*, 96 IOWA L. REV. 581 (2011).

[82] *See* Kerr, *Searches and Seizure in a Digital World*, *supra* note 81, at 554; Note, *Data Mining, Dog Sniffs, and the Fourth Amendment*, 128 HARV. L. REV. 691 (2014); *see also* Florida v. Harris, 133 S. Ct. 1050, 1058 (2013); Kyllo v. United States, 533 U.S. 27 (2001); United States v. Karo, 468 U.S. 705 (1984); Jay Stanley, *Computers vs. Humans: What Constitutes A Privacy Invasion?* ACLU (July 2, 2012), https://www.aclu.org/blog/computers-vs-humans-what-constitutes-privacy-invasion [http://perma.cc/AK37-RYJQ].

[83] United States v. Karo, 468 U.S. 705, 712 (1984).

[84] *Id.*

[85] Kyllo v. United States, 533 U.S. 27, 28 (2001).

through companies like Verizon, Google, and Facebook.[86] Ruling that the bulk collection of American telephone metadata is unconstitutional, Judge Leon, writing for the US District Court for the District of Columbia, characterized the "collect now and query later" form of surveillance in the following way:

> I cannot imagine a more "indiscriminate" and "arbitrary" invasion than this systematic and high-tech collection and retention of personal data on virtually every single citizen for purposes of querying and analyzing it without prior judicial approval. . . . Surely, such a program infringes on "that degree of privacy" that the founders enshrined in the Fourth Amendment.[87]

There are a number of legal issues related to the PRISM program, including the treatment of foreign versus domestic communications and the difference between pen registers and metadata, but the distinction between man and machine searches in Fourth Amendment jurisprudence has certainly played a role in the development of such programs. In fact, the NSA defines the "acquisition" of information for the purposes of a wiretap interception (a violation of the federal wiretapping statute) as, "the collection by NSA or the FBI through electronic means of a non-public communication to which it is not an intended party."[88] But the agency defines "collection" as information that "has been received for use by an employee of a Department of Defense intelligence component in the course of his official duties. . . . [D]ata acquired by electronic means is 'collected' only when it has been processed into intelligible form."[89] In other words, the NSA is operating under the assumption that the collection and processing of communications is not an issue until a human has it in human-readable format.

It is still unclear whether a machine "alone" can invade one's privacy; this line of inquiry has not been settled, and a number of decisions and commentaries point in various directions. For instance, Kevin Bankston and Amie Stepanovich argue that requiring a human to perform specific tasks that trigger Fourth Amendment issues is both unsupported by (some) case law and leads to far, far less privacy

[86] Steven Nelson, *Nine Companies Tied to PRISM, Obama Will be Smacked With Class-Action Lawsuit Wednesday*, U.S. NEWS & WORLD REP.: NEWSGRAM (June 11, 2013, 6:23 PM), http://www.usnews.com/news/newsgram/articles/2013/06/11/nine-companies-tied-to-prism-obama-will-be-smacked-with-class-action-lawsuit-wednesday [http://perma.cc/D454-5Z38].

[87] Klayman v. Obama, 957 F. Supp. 2d 1, 42 (D.D.C. 2013) *vacated and remanded*, No 14-5004, 2015 WL 5058403 (D.C. Cir. Aug. 28, 2015).

[88] NAT'L SEC. AGENCY, MINIMIZATION PROCEDURES USED BY THE NAT'L SEC. AGENCY IN CONNECTION WITH ACQUISITION OF FOREIGN INTELLIGENCE INFORMATION PURSUANT TO SECTION 702 OF THE FOREIGN INTELLIGENCE SURVEILLANCE ACT OF 1978, as amended, § 2(a) (Oct. 31, 2011).

[89] *Id.*

protection.[90] They, however, are in the minority. Richard Posner has argued that only a human can raise constitutional privacy issues.[91] Similarly, Bruce Boyden argues that no "interception" occurs under the Fourth Amendment without human review,[92] and Matthew Tokson argues that automation "alone" cannot violate the Fourth Amendment.[93] By ignoring the man-machine nature of technology, interpretations of laws intended to protect against government invasions of privacy have unintentionally encouraged automated systems to keep the human far from the loop and to operate outside black boxes, which has allowed for widespread automated surveillance and will later be considered as poor design.

3. State Enforcement for Traffic Safety

While states pass laws and produce judicial opinions related to automation, recent controversies surrounding red light cameras have drawn attention to the use of automated enforcement of traffic violations and the laws that authorize this enforcement. There is wide variation among states. For instance, a number of states allow for statewide use of automated enforcement without an officer present (almost all have slightly lower penalties for violations enforced through automation than traditional methods).[94] On the other hand, photo enforcement is prohibited in a number of states, including Nevada, which only allows for the use of the imaging equipment when it is in the hands of an officer or installed in a law enforcement vehicle or facility.[95] Traffic laws are intended to promote safety. Speed limits prohibit drivers from legally driving at speeds known to significantly increase accident numbers and severity. Red lights organize drivers in high traffic zones to prevent collisions. In theory, the enforcement of both of these functions could be fully automated, but prohibiting the use of automated enforcement is as popular as installation.[96] The

[90] *See* Bankston & Stephanovich, *supra* note 81.

[91] *See* Richard A. Posner, *Privacy, Surveillance, and the Law*, 75 U. CHI. L. REV. 245, 254 (2008).

[92] *See* Bruce E. Boyden, *Can a Computer Intercept Your Email?*, 34 CARDOZO L. REV. 669, 673 (2012).

[93] *See* Matthew Tokson, *Automation and the Fourth Amendment*, 96 IOWA L. REV. 581, 613 (2011).

[94] *See Speed and Red Light Camera Laws*, GOVERNORS HIGHWAY SAFETY ASSOCIATION (Oct. 2015), http://www.ghsa.org/html/stateinfo/laws/auto_enforce.html [http://perma.cc/MCE9-8R3D].

[95] *Id.*

[96] Maggie Clark, *Red-Light Cameras Generate Revenue, Controversy*, USA TODAY (Oct. 15, 2013), http://www.usatoday.com/story/news/nation/2013/10/15/stateline-red-light-cameras/2986577/ [http://perma.cc/W8DK-BJG5].

problem is that while cameras reduce red-light running violations, they do not necessarily make intersections safer. In fact, there is mounting evidence that red light cameras have made many intersections more dangerous because human drivers brake differently at these intersections resulting in more rear collisions.[97] Additionally, while automation of enforcement is intended to be accurate, equal, consistent, and particularly suitable for determinable legal conclusions like traffic violations, the *Do Robots Dream of Electric Laws? An Experiment in the Law as Algorithm* project presented at We Robot 2013 reveals significant variation in the number and types of citations issued.[98] We must be critical of technology, especially when introduced to enforce laws, and be aware of the ever-present human in the loop.

B. In the Loop

Other times, machines appear to be the source of the problem. In order to quickly deal with the issue, the law has simply banned the lack of a human or required their involvement. Neither approach effectively solves the problem or protects the stated interests.

1. Regulating Invasive Robocalls

Congress has amended the Telephone Consumer Protection Act (TCPA) twice, and the Federal Communication Commission (FCC) has made numerous changes to implementing the law since it passed in 1991.[99] In the late 1980s, robocalls came under regulatory scrutiny because the automation was considered more invasive than human callers because the rate at which the robocalls could invade the home—and later the pockets and purses of individuals—was much more efficient.[100] An initial spike in complaints prompted the National Do Not Call Registry maintained by the Federal Trade Commission (FTC), which makes no distinction between human and

[97] Carl Bialik, *Seeing Red*, WALL ST. J. (Feb. 1, 2013), http://blogs.wsj.com/numbers/seeing-red-1208/ [http://perma.cc/6J5J-554S].

[98] Lisa Shay, Woodrow Hartzog, John Nelson & Gregory Contri, *Do Robots Dream of Electric Laws? An Experiment in the Law as Algorithm*, 2013 WE ROBOT CONF., http://conferences.law.stanford.edu/werobot/wp-content/uploads/sites/29/2013/04/Shay-et-al_Lisa.pdf [http://perma.cc/5JGL-VKB7].

[99] Telephone Consumer Protection Act of 1991, 47 U.S.C. § 227 (2012).

[100] S. W. Waller, D. Heidtke & J. Stewart, *The Telephone Consumer Protection Act of 1991: Adapting Consumer Protection to Changing Technology* (Loy. U. Chi. Sch. of Law, Research Paper No. 2013-016, 2013), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2327266 [http://perma.cc/49XM-NRN9].

automated calls and today has more than 221 million numbers on it.[101] The FCC responded with regulations that prohibit autodialing[102] and artificial or prerecorded messages except in limited circumstances and now require prior express written consent before telemarketing companies may use either technology to reach customers.[103] An interesting distinction remains for political calls, which are outside the FTC's purview: there are no restrictions on manually dialed political calls to landlines or cell phones, but robocalls (autodialed calls or artificial voice messages) to mobile numbers are prohibited without prior express consent.[104]

Judge Easterbrook explained the justification for regulatory variation between human and robot callers: "A human being who called [a] Cell Number would realize that [the] Customer was no longer the subscriber. But predictive dialers lack human intelligence and, like the buckets enchanted by the Sorcerer's Apprentice, continue until stopped by their true master."[105] The problem is that calls continue to come in when they are unwanted, not that they are a human or artificial voice. Telemarketing robots have become almost indistinguishable from human callers,[106] and reaching voters with automated support or fully automated systems is one way that candidates with smaller bank accounts can promote their message and candidacy.[107] By creating this distinction instead of enforcing recipient choice, regulators limit the benefits of political calls without protecting citizens from unwanted privacy invasions.

2. Right to Fair Decision Making in Europe

Europeans have managed to maintain a legal distance from high levels of automation, though the effectiveness of requiring a

[101] *The Do Not Call Registry*, U.S. FED. TRADE COMM'N, http://www.ftc.gov/news-events/media-resources/do-not-call-registry [http://perma.cc/CH6K-4SAS].

[102] Autodialing is defined as "equipment which has the capacity to store or produce telephone numbers to be called using a random or sequential number generator and to dial such numbers." The FCC has emphasized that this covers equipment that has the "*capacity* to dial numbers without human intervention whether or not the numbers called actually are randomly or sequentially generated or come from calling lists." U.S. Fed. Comm. Comm'n, Enforcement Advisory No. 2012-06 (Sept. 11, 2012), https://www.fcc.gov/document/political-campaigns-restrictions-autodialed-prerecorded-calls [https://perma.cc/2BF8-95JF].

[103] 47 C.F.R. § 64.1200.

[104] U.S. Fed. Comm. Comm'n, *supra* note 102.

[105] Soppet v. Enhanced Recovery Co. LLC, 679 F.3d 637, 639 (7th Cir. 2012).

[106] George Dvorsky, *Robots So Realistic They Can Deny They're Bots*, DISCOVERY NEWS (Dec. 12, 2013) http://news.discovery.com/tech/robotics/robots-so-realistic-they-can-deny-theyre-bots-131212.htm [http://perma.cc/524Y-UCSY].

[107] Jason C. Miller, *Regulating Robocalls: Are Automated Calls the Sound of, or a Threat to, Democracy?*, 16 MICH. TELECOMM. & TECH. L. REV. 213, 215 (2009).

human in the loop to meet certain social goals is questionable. The European Data Protection Directive of 1995 includes the right of every person "not to be subject to a decision which produces legal effects concerning him or significantly affects him or her and which is based solely on automated processing of data intended to evaluate certain personal aspects related to him, such as his performance at work, creditworthiness, reliability, conduct, etc."[108] Significantly weakened by exceptions, the essence of the right ensures that individuals have a right to a human in the loop for any decision that produces legal or significant effects. Leaving open the option for an individual to go outside the automated system that processes everyone else prevents the equalizing purpose of such systems and allows for beneficial treatment to be granted to those who have historically received it. The right is certainly less disruptive than an all-out ban on automated decision making, but it still has done little to protect against the different types of bias or errors that derive from both humans and automation.

Many of the concerns about automated decision making have since been incorporated into regulations related to the expansive concept of "profiling," defined as "any form of automated processing intended to evaluate certain personal aspects relating to [a] natural person or to analyse or predict in particular that natural person's performance at work, economic situation, location, health, personal preferences, reliability, or behaviour"[109] in the Data Protection Regulation, which is set to update the Data Protection Directive. For instance, the main provision in the Regulation is in Article 20.[110] It was previously entitled "Measures based on profiling," which suggests that it refers to decision making based on profiles, but the title was changed to "Profiling" by the Committee on Civil Liberties, Justice, and Home Affairs (LIBE) through its amendments.[111] The method of utilizing a human to protect against harms caused by automation was

[108] Directive 95/46/EC, of the European Parliament and of the Council of 24 Oct. 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, art. 15(1), 1995 O.J. (L 281), 31 et. seq.

[109] *Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individual with Regard toe the Processing of Personal Data and on the Free Movement of Such Data*, at Art. 4(3a), COM (2012) 11 (Oct. 7, 2013) http://www.europarl.europa.eu/meetdocs/2009_2014/documents/libe/dv/comp_am_art_0129/ comp_am_art_01-29en.pdf [http://perma.cc/WYU9-UBP5] [hereinafter LIBE Committee] (the compromise amendments on Arts. 1–29).

[110] *Proposal for a Regulation of the European Parliament of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data* (Data Protection Regulation) at Art. 20, COM (2012) 11 final (Jan. 25, 2012).

[111] LIBE Committee, *supra* note 109, at Art. 20.

reinforced through the LIBE amendments, which retained the following language: "In any case, such processing should be subject to suitable safeguards, including specific information of the data subject and the *right to obtain human assessment* (previously *human intervention*) and that such measure should not concern a child."[112]

When the government required automatic coupling to protect railroad workers, it ignored the way humans interact with the automation, resulting in even more deaths and injuries. Robocalls are heavily regulated by the FCC, but today's automated telemarketers are sophisticated to a level that offers the same recipient action as a human with less invasive treatment. Courts have determined that a machine cannot violate privacy. However, automated tracking, surveillance, and processing reveals more about us today to more organizations than any human could possibly discover. States have automated enforcement of traffic violations in order to improve safety statistics without considering the way in which humans would interact with the technology, resulting in more accidents.[113] Bans on automated decision making in Europe have protected against neither human nor machine error. Referring to problems with the regulation of big data analytics, Cynthia Dwork and Deirdre Mulligan explain: "Viewing the problem as one of machine versus man misses the point. The key lies in thinking about how best to manage the risks to the values at stake in a sociotechnical system."[114]

V. POLICY KNOTS

This Article is not intended to provide specific regulatory proposals or judicial frameworks but to assess legal approaches broadly and suggest ways in which the law may better account for the man-machine nature of automation and support optimized man and machine systems that meet the goals of law. While relying on the false distinction between man and machine is not an effective governance approach, one that recognizes the interdependence within the automated system will still need to offer constraints, provide guidance, and establish accountability. It is important to understand how automation has changed to resolve its ironic outcomes and locate an available position in existing practices for the law. This exercise

[112] *Id.* at Art. 20, Recital 58.

[113] For example, California permits red light cameras to issue citations to registered owners or identifiable drives by taking images of tags and the driver. In Nevada, speed and red light cameras are both prohibited with narrow exceptions. *See Speed and Red Light Camera Laws*, *supra* note 94.

[114] Cynthia Dwork & Deirdre K. Mulligan, *It's Not Privacy, And It's Not Fair*, 66 STAN. L. REV. ONLINE 35, 38 (2013).

allows law to work with, not against, technological design and use what is referred to as "policy knots" in this Section. And so, this Section will expand on why and how utilizing existing automation models can create an effective sociotechnical legal approach to emerging technology.

A. Automation Policy Knots

Policy knots recognize the relationship between policy, design, and practice without being too stringent or too lenient. Coined and articulated by a team at Cornell University as a way of reflecting upon certain phenomenon like Girls Around Me[115] and Google Buzz,[116] the policy knot is a concept that accounts for the way emerging computing "practices and design impact and are impacted by structures and processes in the realm of policy."[117] The policy knot concept is informative to the governance of emerging technology, because it highlights the importance of each element.

When any of the three elements—policy, design, and practice— are out of balance or ignore one another, policies intended to protect specified values are ineffective, unresolved, or ironic. This Article argues that the policy knot can be utilized as a policy tool for addressing emerging technologies. By actively balancing the three elements of policy, design, and practice, more input and guidance may be exerted while maintaining necessary flexibility in design and accounting for technological practices.

A policy knot reinforces and complements existing human-centered design methods for actual and safe use that recognize the sociotechnical nature of technology. In other words, it reinforces automation policy that is in balance with design and practice. More than seeking to balance design processes with social concern, policy knots are actively tied into and proactively help to shape design processes.

[115] GIRLS AROUND ME, http://girlsaround.me/ [http://perma.cc/GWY9-XWY9] (a mobile application that shows users women nearby, using location-based mobile services, like Foursquare, Google Maps, and GPS, and provides information on them based on available Facebook account information).

[116] Google Buzz was a social networking, microblogging, and messaging platform launched in 2010 and discontinued in 2011. *See* Drew Olanoff, *Reminder: Google Buzz is Still Dead, Your Data Will be Moved to Drive, and They Thank You for Using it*, TECHCRUNCH (May 25, 2013), http://techcrunch.com/2013/05/25/reminder-google-buzz-is-still-dead-your-data-will-be-moved-to-drive-and-they-thank-you-for-using-it/ [http://perma.cc/B9BN-VFS4].

[117] Steven J. Jackson, Tarleton Gillespie & Sandy Payette, *The Policy Knot: Re-integrating Policy, Practice and Design in CSCW Studies of Social Computing*, PROCEEDINGS OF CSCW'14, Feb. 15–19, 2014, at 1.

Turning to resolving the man-versus-machine divide in this overlooked subset of law, the challenge for the legal field is to somehow address the complex sociotechnical nature of automation that is situated in practice and designed for optimal human-automation interaction. In an early draft of his *Proxity-Driven Liability* article, Bryant Walker Smith explained that human factors intend for product use to be legal, safe, and intuitive:[118]

> Tensions among the three key design targets suggest particular structural failures. A mismatch between legality and safety implies that law as written is inefficient because it is either too permissive or too restrictive. A mismatch between safety and actuality suggests that users are either uninformed or irrational. And a mismatch between actuality and legality suggests that law is either underenforced or obsolete.[119]

The above examples are mismatches because they focus on the capabilities of the automation, which conflicts with both safe design and actual use, resulting in unprotected values. The task for law is to bring legal treatment of automation in line with responsible automation design and implementation in practice—to tie a policy knot. There is no general legal framework for all automation—it is introduced by government and private entities, in commercial and public service, and across industries and communities for all kinds of purposes. However, when the law addresses automation, based on the above examples, focusing on the capabilities of the automation can be counter-productive.

Just as automation-centered design leads to the ironies of automation, legal treatment that focuses on automation has a similar effect. This may be because the law drives the design in an automation-centered direction even if the designer has intended a human-centered approach. In many ways this trend suggests that the dichotomy between man and machine is a false one, but at a minimum, it is not reliable or stable enough to draft policy around. Without this dichotomy, we must rethink the regulatory approach to automation to meet stated goals and protect established values.

Command-and-control regulations that seek to produce specific outcomes with universal rules prescribing particular conduct or technology are often ill-suited to such complex goals.

> Specific rules often cannot reflect the large number of variables involved in achieving multifaceted regulatory goals, such as reducing the types of risk produced by a combination of factors. . . . They thus direct behavior toward compliance with an incomplete set of detailed provisions that may frustrate, rather than further, the

[118] Bryant Walker Smith, *Human Factors in Robotic Torts*, 2013 WE ROBOT CONF., http://conferences.law.stanford.edu/werobot/wpcontent/uploads/sites/29/2013/04/HumanFactorsRoboticTorts_BryantWalkerSmith.pdf [http://perma.cc/9CSV-NJYS].

[119] *Id.* at 4.

> broader regulatory goal in any particular circumstance. The problem is compounded when regulated entities are heterogeneous, and contexts are varied.[120]

Although less "all or nothing" than the above case studies, requiring policymakers to regulate specific levels of autonomy for different functions is simply not a realistic means of effective governance. Command and control is a too "hands on" form of governance for automation.

Automation use is context-specific and situational, and regulation can, as the above examples suggest, frustrate, rather than further, broad regulatory goals. Thus, it may be a good candidate for a "more 'incomplete' regulatory instrument"[121] that takes a more "hands off" approach to regulation and simply asks for limited harms to occur or goals to be met. But performance or goal-based regulations that identify specific outcomes, leaving the means up to the regulatory party, are ineffective when "desired performance is difficult to identity in advance or assess contemporaneously"—the focus shifts from punishment to prevention.[122] Relatively recently, assessing and preventing risk in complex spaces has been delegated based on the understanding that the regulators are too far removed to comprehend and effectively direct internal workings.[123] The delegation to the regulated party seeks to take advantage of the expertise and judgment within the regulated organization to reduce complex risk by not only mitigating the risks, but also by defining and monitoring it.[124]

For a legal approach to automation that exerts external control while leaving room for necessary flexibility and responsibility, "shared responsibility" and institutional conditions that support enhanced ethical decision making may serve as important governance goals.[125] Mark Coekelbergh explains that external controls in the form of regulating engineers have come in two relevant varieties: prescriptions which require engineers to follow codes and standards, offering a great deal of certainty with little autonomy, and goal setting, which offers more flexibility by encouraging the designer to focus on risk and justification of choices but less certainty.[126] Coekelbergh suggests that this ethical responsibility may not be desired and argues that external constraints limit the moral

[120] Kenneth A. Bamberger, *Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State*, 56 DUKE L.J. 377, 386–87 (2006).

[121] *Id.* at 389–90.

[122] *Id.*

[123] *Id.*

[124] *Id.* at 390.

[125] Mark Coeckelbergh, *Regulation or Responsibility? Autonomy, Moral Imagination, and Engineering*, 31 SCI., TECH., & HUMAN VALUES 237, 245 (2006).

[126] *Id.* at 240–41.

imagination of engineers,[127] but Shilton suggests that ethical constraints, such as privacy, can be welcome additions to the design process.[128] While this Article calls for an approach that aligns somewhat with a goal-setting form of governance and somewhat with a delegation approach, it seeks nonetheless to support enhanced ethical decision making by proposing that a set of principles appropriately situated will offer some sense of certainty, while preserving flexibility and establishing accountability.

B. Man versus Machine Design

Overpowering design and practice of sociotechnical systems is likely the result of neglect of these issues on the part of policymakers. In order to find the right complimentary role for law in the automation policy knot, automation design and practice are detailed in the following subsections. The treatment of human-automation system models included here is one that has directed the field and serves as a foundational marker. The treatment should not be considered exhaustive nor does it include in any detail models for adaptive autonomy, co-robotics, or scalable autonomy. An evolution of the field is presented to show the way in which the legal approach must adapt to new forms of design and practice.

Not only is requiring a human in or out of the loop too simplistic a response to the threat to human values because of her permanent role in the loop, it can also be an ineffective form of accountability and create safety issues. Realizing that humans will always be in the loop, a body of research has developed to understand how the human in the loop should be accounted for to preserve or optimize performance of the system. In 1951, the Fitts List sparked an entire body of research focusing on function allocation—those functions that humans should perform and those that machines (today computers and sometimes robots) should perform.[129] Fitts et al. intended to "search for a general answer to the problem of dividing responsibility between men and machines."[130] The Fitts List is a list of eleven statements (also called MABA-MABA: men are better at, machines are better at) that categorize whether the man or machine performs a function better than the other.

[127] *Id.* at 252.

[128] Katie Shilton, *Value Levers: Building Ethics into Design*, 38:3 SCI., TECH., & HUMAN VALUES 374, 383–84 (2012).

[129] J.C.F. de Winter & D. Dodou, *Why the Fitts List has Persisted Throughout the History of Function Allocation*, 16 COGNITION, TECH. & WORK, 1 (2014).

[130] *Id.* at 3, 6.

Figure 1. The original Fitts List (MABA-MABA List), 1951.[131]

Humans Excel In	Machines Excel In
Ability to detect a small amount of visual or acoustic energy	Ability to respond quickly to control signals and to apply great force smoothly and precisely
Ability to perceive patterns of light or sound	Ability to perform repetitive, routine tasks
Ability to improvise and use flexible procedures	Ability to store information briefly and then to erase it completely
Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time	Ability to reason deductively, including computational ability
Ability to reason inductively	Ability to handle highly complex operations, i.e. to do many different things at once
Ability to exercise judgment	

Figure 2. Department of Defense adaptation, 1987.[132]

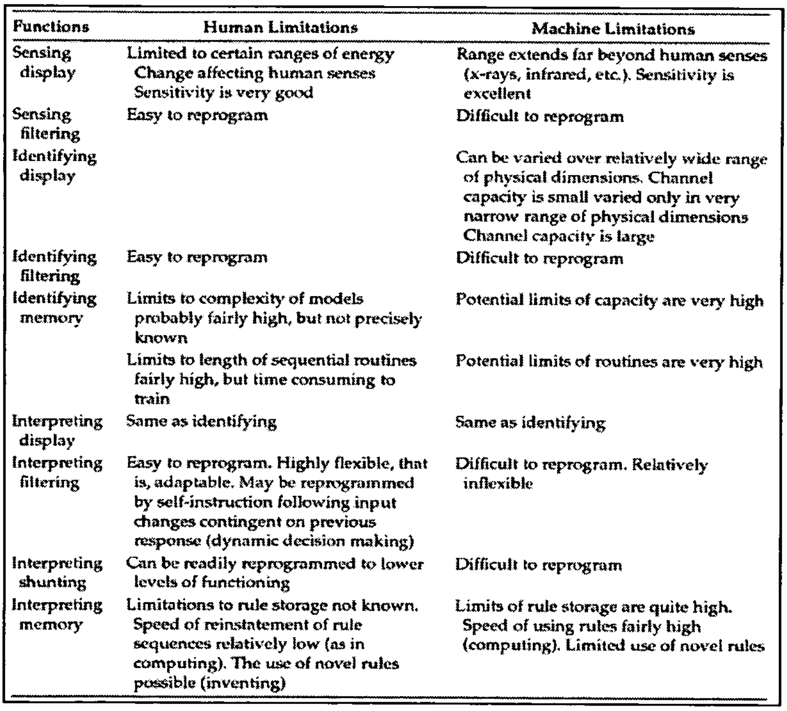
Humans Excel In	Machines Excel In
Detection of certain forms of very long energy levels	Monitoring (both men and machines)
Sensitivity to an extremely wide variety of stimuli	Performing routine, repetitive, or very precise operations
Perceiving patterns and making generalizations about them	Responding very quickly to control signals
Store large amounts of information for long periods—and recall relevant facts at appropriate moment	Storing and recalling large amounts of information in short time periods
Ability to exercise judgment where events cannot be completely defined	Performing complex and rapid computation with high accuracy
Improving and adopting flexible procedures	Sensitivity to stimuli beyond the range of human sensitivity (e.g., infrared, radio waves)
Reacting to unexpected low-probability events	Doing many different things at the same time
Applying originality in closing problems	Exerting large amounts of force smoothly and precisely

[131] PAUL M. FITTS, HUMAN ENGINEERING FOR AN EFFECTIVE AIR NAVIGATION AND TRAFFIC CONTROL SYSTEM (1951).

[132] U.S. DEP'T OF DEF., HUMAN ENGINEERING PROCEDURES GUIDE, MIL-HDBK–763, 93 (1987).

Profiting from experience and altering course of action	Insensitivity to extraneous factors
Performing fine manipulation, especially where misalignment appears unexpectedly	Repeating operations very rapidly, continuously, and precisely
Continuing to perform when overloaded	Operating in environments that are hostile to man or beyond human tolerance
Reasoning inductively	Deductive processes

Figure 3. Robert Gagne limitation-based adaptation, 1962.[133]



Functions	Human Limitations	Machine Limitations
Sensing display	Limited to certain ranges of energy Change affecting human senses Sensitivity is very good	Range extends far beyond human senses (x-rays, infrared, etc.). Sensitivity is excellent
Sensing filtering	Easy to reprogram	Difficult to reprogram
Identifying display		Can be varied over relatively wide range of physical dimensions. Channel capacity is small varied only in very narrow range of physical dimensions Channel capacity is large
Identifying filtering	Easy to reprogram	Difficult to reprogram
Identifying memory	Limits to complexity of models probably fairly high, but not precisely known	Potential limits of capacity are very high
	Limits to length of sequential routines fairly high, but time consuming to train	Potential limits of routines are very high
Interpreting display	Same as identifying	Same as identifying
Interpreting filtering	Easy to reprogram. Highly flexible, that is, adaptable. May be reprogrammed by self-instruction following input changes contingent on previous response (dynamic decision making)	Difficult to reprogram. Relatively inflexible
Interpreting shunting	Can be readily reprogrammed to lower levels of functioning	Difficult to reprogram
Interpreting memory	Limitations to rule storage not known. Speed of reinstatement of rule sequences relatively low (as in computing). The use of novel rules possible (inventing)	Limits of rule storage are quite high. Speed of using rules fairly high (computing). Limited use of novel rules

While consistently referenced in function-allocation research, the Fitts List has been heavily criticized as an intrinsically flawed descriptive list, little more than a useful starting point, insufficient, outdated, static, and incapable of acknowledging the organizational context and complementary nature of humans and machines.[134] Like many efforts of initial thinking on automation, the list focuses on

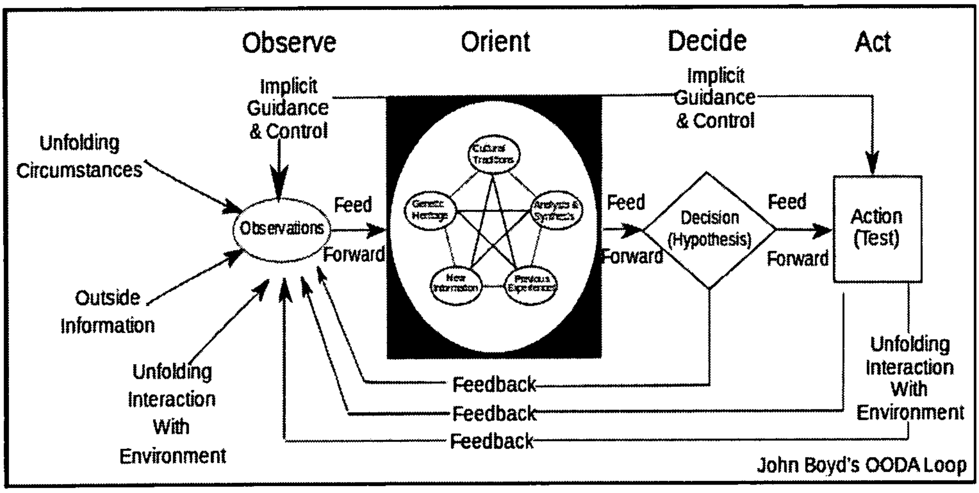
[133] R. M. Gagne, *Human Functions in Systems, in* Psychological Principles in System Development (R. M. Gagne ed., 1962).

[134] *Id.* at 1.

automation capabilities and could not serve as a sufficient framework for moving forward with automation.

Around the same time, John Boyd developed a model of human decision making to inform dogfighting tactics for military pilots: Observe, Orient, Decide, Act (OODA).[135] Boyd discerned that the best systems were not the best planes or the best pilots, but the best systems that could move through the OODA model the quickest and most effectively.[136] This groundbreaking model has played a part in the continued elite performance of U.S. pilots.[137]

Figure 4. John Boyd's "Observe, Orient, Decide, Act" model long relied upon for aviation systems.[138]



The four stage information processing model used by Thomas Sheridan, Raja Parasuraman, and Christopher Wickens for automation is comparable: information acquisition, information analysis, decision selection, and action implementation.[139] Automation of information acquisition deals with input data. It may include

[135] ROBERT CORAM, Boyd: The Fighter Pilot Who Changed the Art of War, 1 (2004); DANIEL FORD, A Vision So Noble: John Boyd, the OODA Loop, and America's War on Terror, 1 (2010); FRANS P. B. OSINGA, Science, Strategy and War: The Strategic Theory of John Boyd, 1 (2006).

[136] CORAM, *supra* note 135; FORD, *supra* note 135, at 28; OSINGA, *supra* note 135, at 176.

[137] CORAM, *supra* note 135; FORD, *supra* note 135; OSINGA, *supra* note 135.

[138] OSINGA, *supra* note 135, at 231.

[139] Raja Parasuraman, Thomas B. Sheridan & Christopher D. Wickens, *A Model for Types and Levels of Human Interaction with Automation*, 30:3 IEEE TRANSACTIONS ON SYS. MAN AND CYBERNETICS PART A: SYS. & HUMANS 286, 290 (2000).

highlighting to bring attention to potential problem information or filtering to bring certain information to the person's attention.[140] The information analysis phase involves the manipulation of retrieved and processed information in working memory.[141] Algorithms can be applied to incoming data to produce predictions, and automated information managers can provide context-dependent summaries of data to human operators. The decision and action selection phase involves decision making based on cognitive processing.[142] Examples of the decision phase include conditional logic used in expert systems to present a decision if particular conditions exist. This phase may require value assumptions about different possible outcomes of the decision. At the action implementation phase, a response selection consistent with choice is made. Automated actions may include an agent that executes certain tasks automatically in a contextually appropriate fashion[143] (i.e., photocopiers sort, collation, stapling, and other similar actions can have different levels of automation, leaving certain tasks to the human). Of course, these levels are not linear; they are coordinated and overlap into "perception-action" cycles.[144]

Consider the elevator, an oft-cited example of the removal of the human from the loop.[145] Elevator operators were necessary when the automation was first introduced to ease public uncertainty about the innovation.[146] However, the human is still very much a part of the loop. A human must press a button to initiate the process, and all information processing and decisions are made by the human. The automation is only of the action implementation. The elevator must be designed with two goals: that the buttons make sense to humans and that a human can fix the problem when an error occurs. Many of our daily tasks have the opposite allocation. We rely on the automation of information acquisition and analysis of weather applications to decide whether to walk or take the bus. Our emails are automated to highlight, filter, and organize messages to support decisions about which information to focus and take action on.

These four stages have been overlaid with various levels of autonomy to further model automation options. Thomas Sheridan and

[140] *Id.* at 288, 290.

[141] *Id.*

[142] *Id.*

[143] *Id.* at 289.

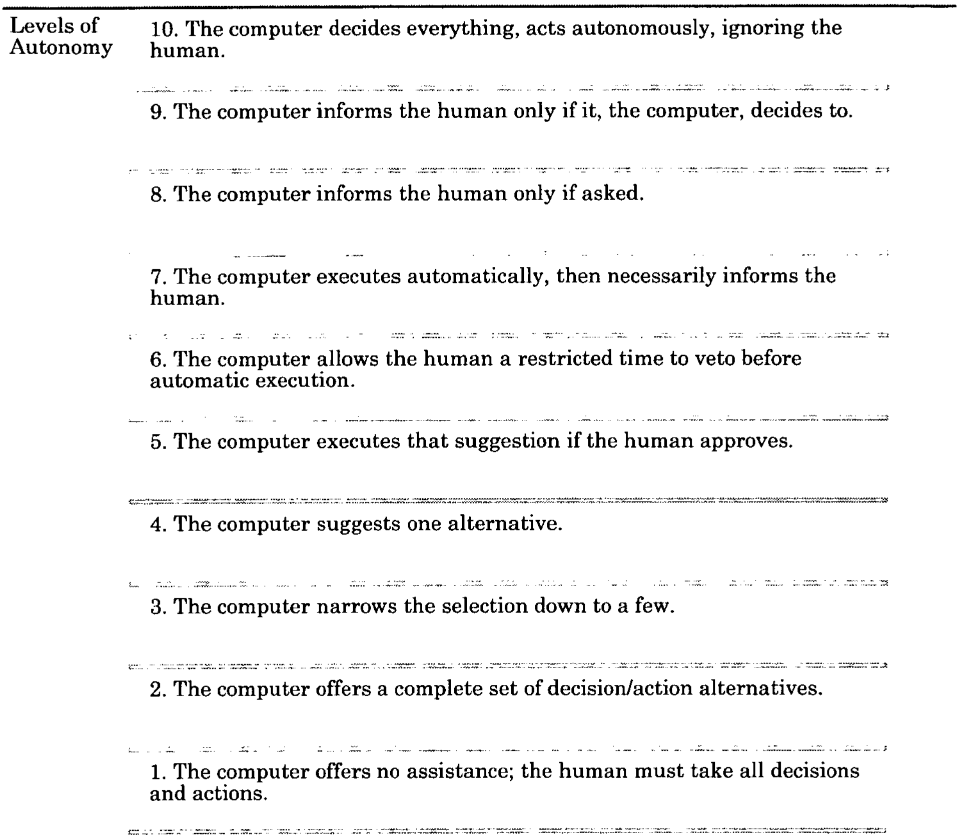
[144] *Id.* at 289–90.

[145] P.W. SINGER, Wired for War 126 (2009).

[146] *Id.*

William Verplank are frequently credited with pioneering the concept of levels of autonomy which are condensed to the following: [147]

Figure 5. Levels of autonomy to be applied to the four-stage model.[148]



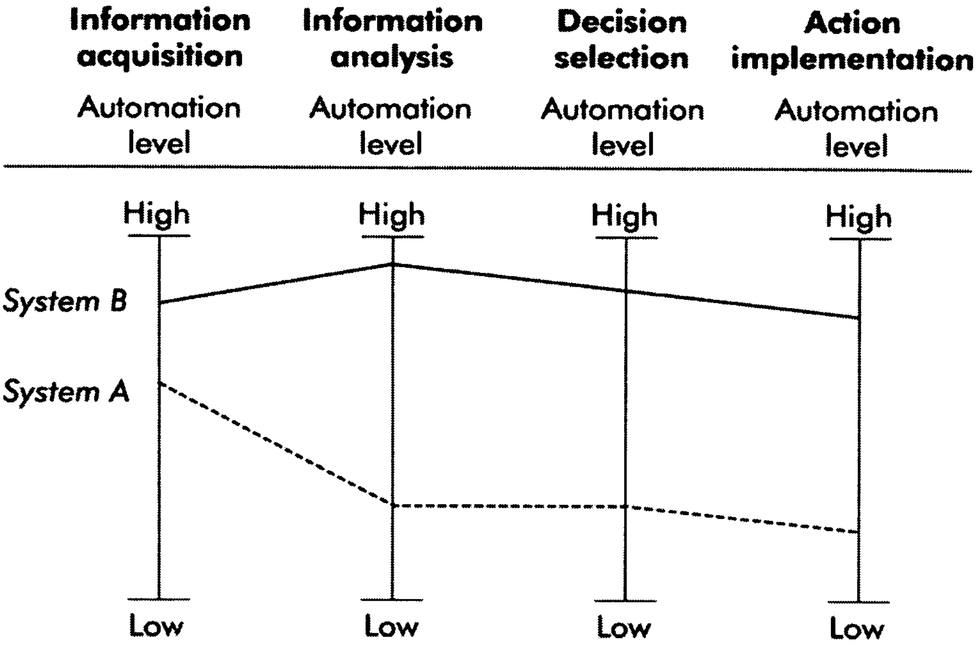
Levels of Autonomy	10. The computer decides everything, acts autonomously, ignoring the human.
	9. The computer informs the human only if it, the computer, decides to.
	8. The computer informs the human only if asked.
	7. The computer executes automatically, then necessarily informs the human.
	6. The computer allows the human a restricted time to veto before automatic execution.
	5. The computer executes that suggestion if the human approves.
	4. The computer suggests one alternative.
	3. The computer narrows the selection down to a few.
	2. The computer offers a complete set of decision/action alternatives.
	1. The computer offers no assistance; the human must take all decisions and actions.

[147] Thomas. B Sheridan and William L. Verplank, *Human and Computer Control of Undersea Teleoperators*, MIT MAN-MACH. SYS. LAB REPORT 26 (1978); *see* THOMAS B. Sheridan, Telerobotics, Automation & Human Supervisory Control (1992).

[148] *See id.*

Automation can be applied to the four classes of functions to differing degrees:

Figure 6. Example of two systems with different levels of automation across functions.[149]



Breaking processes into four phases and assigning each a level of automation appropriate for maintaining optimal system functioning has been vital to resolving the ironies of automation outlined by Bainbridge.

C. Man and Machine Design

The MABA-MABA dichotomous approach continued until work on "human-centered" design began to penetrate a number of fields in the 1980s.[150] An evaluation for automating functions to a certain level by looking at impact on human operator performance, automation reliability,[151] and costs[152] was in place in 2000 and has been relied

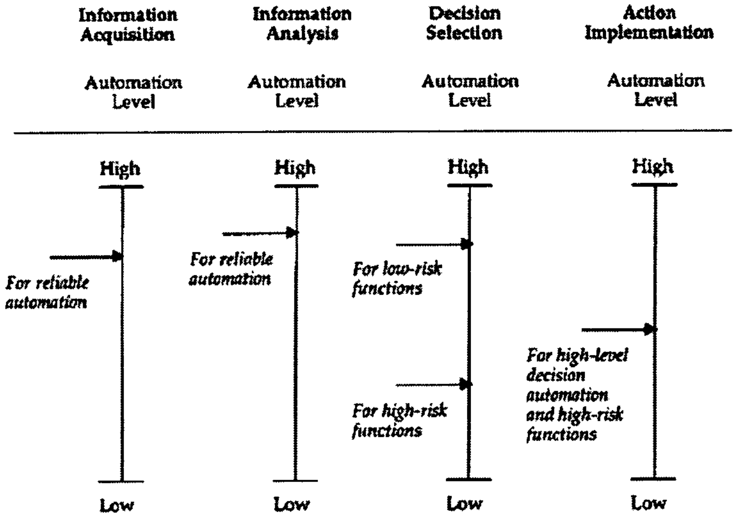
[149] Parasuraman et al., *supra* note 139, at 289.

[150] Thomas B. Sheridan, *Human Centered Automation: Oxymoron or Common Sense?*, 1 Sys., Man & Cybernetics 823 (1995).

[151] Parasuraman et al., *supra* note 139, at 291.

upon to provide general guidance. Charts like the following can help guide responsible implementation of automation taking into account the overall goals of the system and the humans surrounding and in the loop.

Figure 7. Example of recommended levels for Air Traffic Control systems after evaluation of human performance consequences, automation reliability, and costs of actions.[153]



In relation to manual operations, if research shows that both human and system performance are enhanced by automation at levels of five but degrade above level seven, then the reliability and social costs of automation should be considered within the bounds of a five-to seven automation design. From these evaluation exercises, guidance like the following can be drafted:

> For rigid tasks that require no flexibility in decision making and with low probability of system failure, full automation often provides the best solution. However, in systems like those that deal with decision making in dynamic environments with many external and changing constraints, higher levels of automation are not advisable because of the risks and the inability of an automated decision aid to be perfectly reliable.[154]

[152] *Id.*

[153] *Id.* at 294.

[154] Mary L. Cummings, *Automation and Accountability in Decision Support System Interface Design*, 32:1 J. TECH. STUDIES 23, 24 (2006).

These types of conclusions will continue to come out of human factors and systems engineering research, establishing expectations for designers and implementers.

In short, automation changes the nature of the errors that occur by reducing human error but not the probability of system error in general. Automation leads to the deterioration of human operator skill, which needs to be more sophisticated to deal with novel and unique situations. Automation may increase operator workload, complacency, and situational awareness resulting in a decline of safety and performance. Automation reliability leads to over or under trust of the system. Operators may commit misuse, abuse, or disuse of an automation system due to any number of the above factors. However, as these systems are integrated into social settings beyond factories, flight routes, and power plants, additional factors must be considered. These social costs are where policy can play a more active role.

D. Man and Machine Ethics

One interesting void in the automation-design evolution is a lack of attention paid explicitly to value-centered or value-sensitive design. This Article includes only a small sample of the extraordinary work that has focused on making automation safe and effective by acknowledging the human element of these systems, but little work has been done on integrating ethical considerations.[155] That being said, larger concerns are part of the evaluation process and a space to add input and structure.

Ethical determinations, beyond effective performance, in man-machine systems are a portion of the conversation that seems to have lost steam. In 1954, Norbert Wiener expressed general principles for the automatic future he envisioned.[156] He toyed with many ethical possibilities but was unambiguous about his feelings on trusting machines to make critical decisions as a substitute for a human:

[155] There is a large body of work on robot ethics and moral artificial intelligence. *See, e.g.*, RONALD ARKIN, GOVERNING LETHAL BEHAVIOR IN AUTONOMOUS ROBOTS (2009); PATRICK LIN, KEITH ABNEY, & GEORGE BEKEY, Robot Ethics: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS (2011); WENDELL WALLACH & COLIN ALLEN, MORAL MACHINES: TEACHING ROBOTS RIGHT FROM WRONG (2010). Additionally, there is work available on ethical impacts of automation. *See, e.g.*, SRINIVASAN RAMASWAMY & HERMANT JOSHI, *Automation Ethics*, *in* HAND OF AUTOMATION 809 (Shimon Y. Nof, ed., 2009); NORBERT WIENER, The Human Use of Human Beings: Cybernetics and Society (1954); Cummings, *supra* note 154.

[156] WIENER, *supra* note 155, at 16.

> [A human should] not leap in where angels fear to tread, unless he is prepared to accept the punishment of the fallen angels. Neither will he calmly transfer to the machine made in his own image the responsibility for his choice of good and evil, without continuing to accept a full responsibility for that choice.[157]

He suggested ethical principles to be coded into systems for decision making machines, likely still expecting machines would not decide the heaviest of decisions:

> Any machine constructed for the purpose of making decisions, if it does not possess the power of learning, will be completely literal-minded. Woe to us if we let it decide our conduct, unless we have previously examined the laws of its action, and know fully that its conduct will be carried out on principles acceptable to us![158]

However, he continues, if our machine can alter its code in such a way that alters the ethical restraints:

> On the other hand, the machine . . . which can learn and can make decisions on the basis of its learning, will in no way be obliged to make such decisions as we should have made, or will be acceptable to us. For the man who is not aware of this, to throw the problem of his responsibility on the machine, whether it can learn or not, is to cast his responsibility to the winds, and to find it coming back seated on the whirlwind.[159]

Imagining an automatic society, Wiener argued that humans should maintain ultimate responsibility for critical decisions, program non-learning automation with ethical code, and realize that learning automation will not necessarily make human-like decisions. This type of automatic society is easier said than done, but the concepts (and associated methods) are still not incorporated into the design process. There is, however, a placeholder of sorts.

In a model designed by Parasuraman, Sheridan, and Wickens, the resulting system's impact on the operator's performance is considered to establish initial types and levels of automation.[160] Then, risk is assessed as part of the "secondary evaluation." These are separated into "Automation Reliability" and "Costs of Decision/Action Outcomes."

"Although it would be nice if constructed systems functioned well forever, they do not."[161] Reliability is defined as "the probability that an item will operate adequately for a specified period of time in its intended application."[162] While machine reliability and human reliability can be analyzed and combined to predict overall

[157] *Id.* at 184.

[158] *Id.* at 185.

[159] *Id.*

[160] Parasuraman et al., *supra* note 139, at 290.

[161] ROBERT W. PROCTOR & TRISHA VAN ZANDT, HUMAN FACTORS IN SIMPLE AND COMPLEX SYSTEMS 65 (2nd ed. 2008).

[162] KYUNG S. PARK, Human Reliability: Analysis, Prediction, and Prevention of Human Errors 149 (1987).

performance of the system,[163] automation reliability will dramatically influence the actual use of the system because of its tremendous impact on human trust.[164]

> Trust in automation is limited to the degree that evidence from an operator's past experience does or doesn't provide adequate warrant for predicting how the machine will behave in novel situations. If adequate trust and mistrust signatures for every situation were always available, we could remedy this problem[—]but such expectations are unrealistic.[165]

Understanding trust is important to moving beyond rigid levels of autonomy designations to adaptive autonomy and supervisory control imitations to collaborative models,[166] but for the purposes of understanding how a system will be assessed beyond its limited scope (performance of the human operator-machine system), costs of action are more relevant.

Assuming errors will occur and accounting for the reliability of system performance, the way in which errors are managed by the system to avoid costs-of-action outcomes will determine whether automation levels need to be adjusted. Costs-of-action outcomes speak directly to risk. Risk is defined generically as the costs of an error multiplied by the probability of the errors.[167] High levels of automation are not recommended for systems where costs of errors are dramatic, such as the loss of human life, because when errors occur in highly automated systems, it is difficult for a human to step in to resolve the problem.

Zero-risk impacts can still exist even with complete automation failure, and these situations are good candidates for high-level automation throughout the phases. High levels of automation for decisions may also be justified when there is insufficient time for a human operator to respond and take appropriate action or if the human operator is not required to intervene or mange the system in the event of automation failure. For instance, high levels of automated decisions are set in place at nuclear power plants so that control rods automatically drop into the core under emergency

[163] PROCTOR & VAN ZANDT, *supra* note 161, at 59–79.

[164] John D. Lee & Neville Moray, *Trust, Self-Confidence, and Operator's Adaptation to Automation*, 40 INT'L J. HUMAN-COMPUT. STUDIES 153 (1994); John D. Lee & Katrina A. See, *Trust in Automation: Designing for Appropriate Reliance*, 46 HUMAN FACTORS 50 (2004); Raja Parasuraman, Thomas B. Sheridan & Christopher D. Wickens, *Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs*, 2:2 J. OF COGNITIVE ENG'G & DECISION MAKING 142 (2008).

[165] Robert R. Hoffman et al., *Trust in Automation*, 28 INTELLIGENT SYS. 84, 84–85 (2013).

[166] *Id.* at 84.

[167] Parasuraman et al., *supra* note 139, at 292.

circumstances because the operator cannot reliably respond in time to avoid a potentially catastrophic accident. An anesthesiologist is in a similarly high-risk situation but utilizes lower levels of automation at each stage to maintain familiarity with the system as it works because, under abnormal circumstances, she may need to intervene and take control.

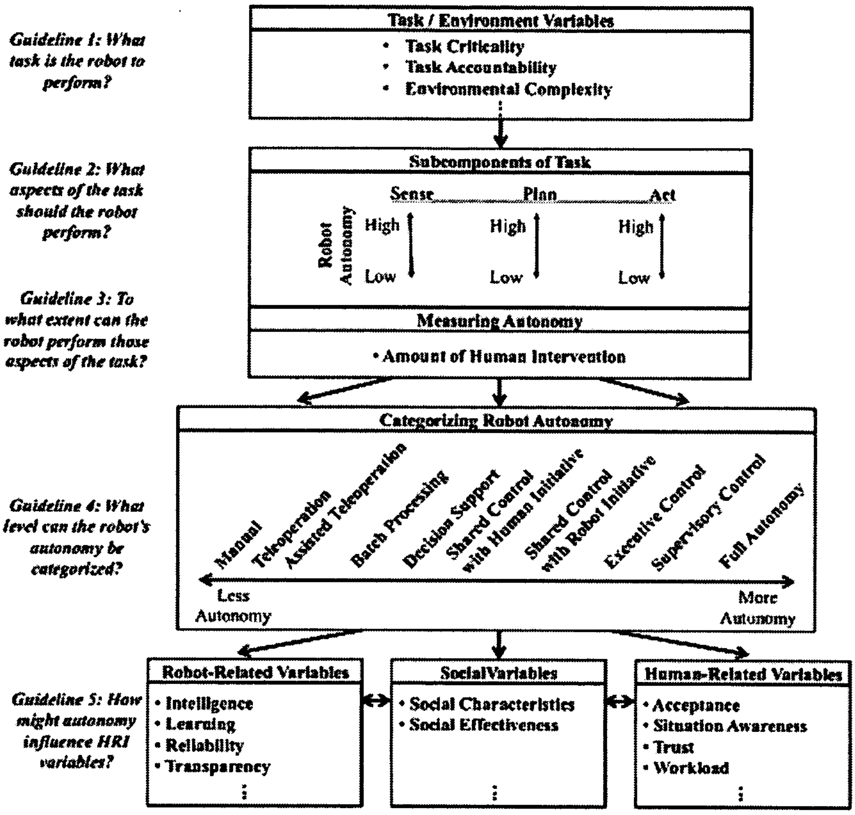
If the costs that occur when the actions are incorrect or inappropriate are high, automation may still be allowable or advisable depending on how human involvement will protect against the risk. When a human is never expected to take control, full automation may be justified.

> Full automation requires highly reliable error handling capabilities and the ability to deal effectively and quickly with a potentially large number of anomalous situations. In addition to requiring the technical capability to deal with all types of known errors, full automation without human monitoring also assumes the ability to handle unforeseen faults and events. This requirement currently strains the ability of most intelligent fault-management systems.[168]

The system's reliability is assessed and levels of automation are assigned based on the system functioning improperly—because it will.

[168] *Id.* at 292 n.3.

Figure 8. Beer, Fisk, and Rogers embrace fluctuating levels of autonomy in various environments, tasks, and interactions for deterring emerging "smart" automation.[169]


Guideline 1: What task is the robot to perform?
Task / Environment Variables
• Task Criticality
• Task Accountability
• Environmental Complexity
Guideline 2: What aspects of the task should the robot perform?
Subcomponents of Task
Sense
Plan
Act
Robot Autonomy
High
Low
High
Low
High
Low
Guideline 3: To what extent can the robot perform those aspects of the task?
Measuring Autonomy
• Amount of Human Intervention
Guideline 4: What level can the robot's autonomy be categorized?
Categorizing Robot Autonomy
Manual
Teleoperation
Assisted Teleoperation
Batch Processing
Decision Support
Shared Control with Human Initiative
Shared Control with Robot Initiative
Executive Control
Supervisory Control
Full Autonomy
Less Autonomy
More Autonomy
Guideline 5: How might autonomy influence HRI variables?
Robot-Related Variables
• Intelligence
• Learning
• Reliability
• Transparency
SocialVariables
• Social Characteristics
• Social Effectiveness
Human-Related Variables
• Acceptance
• Situation Awareness
• Trust
• Workload

Human factors engineers at the University of Southern California have utilized automation models to develop an updated framework that accounts for the mobility, various environmental contexts, and social aspects of emerging robotics and in doing so, have moved elements of the secondary evaluation to the beginning of the automation assessment.[170] Task criticality and accountability set the stage in Guideline 1 for how automation and autonomy develop in human-robot teams but still consider the human-centered aspects of

[169] Jenay M. Beer, Arthur D. Fisk & Wendy A. Rogers, *Toward a Framework for Levels of Robot Autonomy in the Human-Robot Interaction*, 3:2 J. HUMAN-ROBOT INTERACTION 74, 92 (2014).

[170] *Id.* at 91–92.

automation as well as the unpredictability of advanced artificial intelligence.[171]

The "secondary evaluation" exercise could be enriched to serve as what Katie Shilton calls a value lever, "practices that open new conversations about social values and encourage consensus around those values as design criteria."[172] Reflective design,[173] critical technical practice,[174] participatory design,[175] value-sensitive design,[176] values in design,[177] engineering ethics,[178] and other related fields offer methods for bringing bias to the surface and challenging existing design assumptions that may serve to enhance this evaluation stage, but more active involvement from law and policy would provide much-needed direction for identifying, interpreting, and resolving larger social concerns. This involvement will require designers and implementers of automation to expand a narrow view of evaluation[179] and for those in legal roles to resist the urge to solve complex issues with a human or a machine or to draw convenient lines between the two.

Policy is in sync with automation design paradigms from the 1950s and 1960s that have long evolved to recognize the sociotechnical nature of human-automation interaction. As Dave Woods stated, "automation that is strong, silent, and hard to direct is not a team player."[180] For automation that is more dispersed amongst sectors and further integrated into everyday life, automation needs to be a good team player, but also must be part of a team that reflects larger social concerns. There is room for policy principles that support the design

[171] *Id.*

[172] Shilton, *supra* note 128, at 374.

[173] Phoebe Sengers, Kirsten Boehner, Shay David & Joseph Kaye, *Reflective Design*, 2005 Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility 49.

[174] Philip E. Agre, *Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI*, *in* SOCIAL SCIENCE, TECHNICAL SYSTEMS AND COOPERATIVE WORK: BEYOND THE GREAT DIVIDE (Geoffrey C. Bowker, Les Gasser, Susan Leigh Star & Bill Turner eds., 1997).

[175] Peter M. Asaro, *Transforming Society By Transforming Technology: The Science and Politics of Participatory Design*, 10 ACCT. MGMT. & INFO. TECH. 257, 285 (2000).

[176] Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, *in* Human Values and the Design of Computer Technology (Batya Friedman ed., 1997).

[177] Cory Knobel & Geoffrey C. Bowker, *Values in Design*, 54 COMM. ACM 26, 26–28 (2011).

[178] Peter-Paul Verbeek, *Materializing Morality: Design Ethics and Technological Mediation*, 31:3 SCI. TECH. & HUM. VALUES 361, 362–63 (2006).

[179] *See generally* Michael Davis, *Explaining Wrongdoing*, 20 J. SOC. PHIL. 74 (1989).

[180] Charles E. Billings, "Issues Concerning Human-Centered Intelligent Systems: What's 'Human Centered' and What's the Problem?", NSF Workshop on Human-Centered Systems: Information, Interactivity, and Intelligence Talk (Feb. 17–19, 1997), http://www.ifp.illinois.edu/nsfhcs/talks/billings.html [http://perma.cc/UGB8-XDTG].

process instead of fighting it. Methods for promoting ethical discourse and reflection need substantive supplementation, as well as structures for accountability and enforcement. These sociotechnical principles can lay the foundation for sociotechnical regulations and implementation.

VI. FAIR AUTOMATION PRACTICES PRINCIPLES

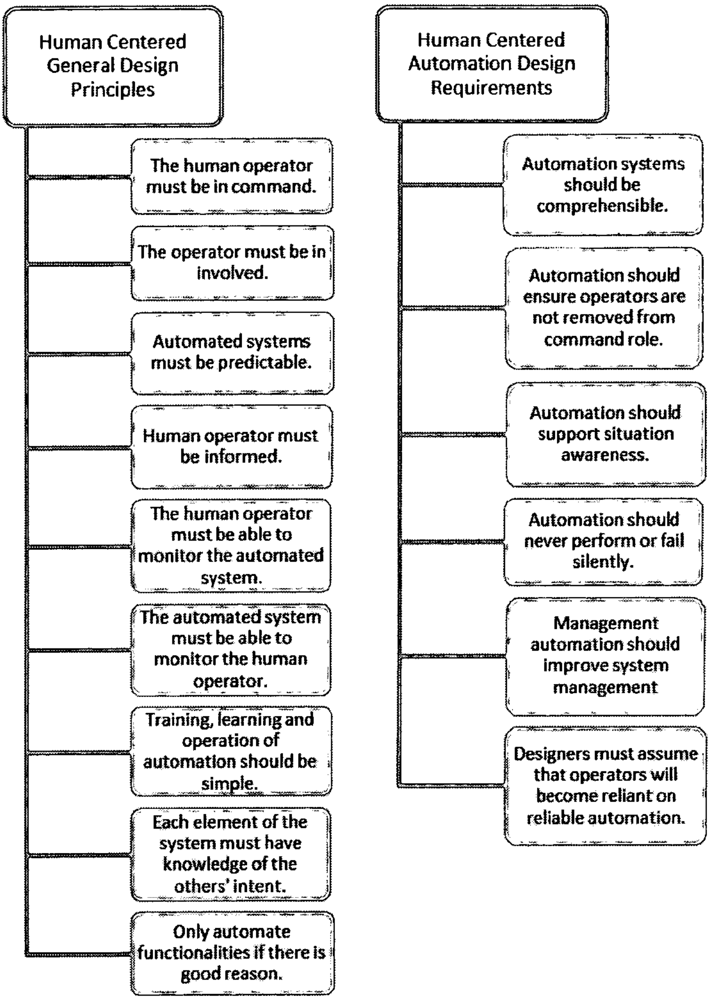
A need for an innovative approach to the increasingly complex nature of emerging technology does not necessarily demand a set of principles. For instance, law professor Gary Marchant has pointed out that expedited rulemakings, self-regulation, issue-specific legislation, special courts, sunset clauses, periodic reviews, independent review boards, adaptive management, and principle-based regulation are all possible and innovative governance options.[181] Principles are best suited to these tasks because of the broad framing chosen. Automation is not a specialized category of technology deserving specific legislation or courts. Indeed, all of the innovative models proposed in the field of study could be implemented for more precise automated technologies in particular contexts if given the proper foundational principles. Less dangerous than disrupting the loop and more proactive than simply asking automation firms to perform risk assessments or self-govern, utilizing existing principles developed by man and machine automation designers can provide much-needed guidance and a platform for more contributions from technology law scholars, ethicists, policymakers, and the public.

Depending on the potential societal issues presented by the automation, flexible oversight can be more or less stringent, but such an approach would hold designers and implementers of automation accountable for their decisions surrounding threats to human values like safety, privacy, dignity, and transparency by giving them the necessary flexibility to make those determinations themselves based on the actual and intended use of the automation. Without the development of these principles, prescriptive regulations will be too near-sighted; reactive, goal-setting governance will remain industry specific; and delegation will be too lenient and overbroad. General principles for automation have been established that could serve as a standard for accountability, particularly if they were updated and edited with additional social concerns and risks relevant to large-scale integration of automation and robotics. Charles Billing's principles for

[181] Gary Marchant, *The Growing Gap Between Emerging Technologies and the Law*, *in* GROWING GAP, *supra* note 13, at 28–30.

human-centered aircraft automation are offered here by way of example:

Figure 9. Charles E. Billings's principles for aircraft automation, 1991.[182]


Human Centered General Design Principles
The human operator must be in command.
The operator must be in involved.
Automated systems must be predictable.
Human operator must be informed.
The human operator must be able to monitor the automated system.
The automated system must be able to monitor the human operator.
Training, learning and operation of automation should be simple.
Each element of the system must have knowledge of the others' intent.
Only automate functionalities if there is good reason.
Human Centered Automation Design Requirements
Automation systems should be comprehensible.
Automation should ensure operators are not removed from command role.
Automation should support situation awareness.
Automation should never perform or fail silently.
Management automation should improve system management
Designers must assume that operators will become reliant on reliable automation.

These principles were outlined for aircraft automation, but similar principles could be drafted for different fields or for more general purposes. The principles focus on safe, functional, and optimized man-machine performance, but do not cover everything.

[182] Charles E. Billings, *Human-Centered Aircraft Automation: A Concept and Guidelines*, 103885 NASA TECHNICAL MEMORANDUM 81–88 (1991), ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19910022821.pdf [http://perma.cc/53RX-X6AA].

For instance, the principles do not have a principle or value lever for ensuring genders are considered in the design. Some problems with the secondary evaluation outlined above in Figure 9, which include automation reliability and costs of automation outcomes, are that the costs of automation outcomes are assessed at the end of the analysis, the limited scope of risk assessment, and the lack of guidance on ethical constraints. These are all areas where people skilled in developing and providing such guidance should focus.

Inspired by the Fair Information Practice Principles (FIPPs),[183] a series of guiding principles representing acceptable actions taken toward personal data (which include providing notice, gaining consent, allowing access and participation to information, keeping up security measures, and offering redress), the Fair Automation Practice Principles (FAPPs) can lay a foundation upon which to build inclusive design practices, strong partnerships, specific laws, tailored regulations, and enriched jurisprudence. Similarly to FIPPs, laws can be built upon these principles and other forms of accountability. This initial governance step supports governing emerging technology and the associated struggle with foreseeability and hazy expectations. FIPPs, in some form or another, have been incorporated into nearly every information and privacy law at every level of legislation and regulation.[184] Similarly to a privacy-by-design approach,[185] which is "not a specific technology or product but a systematic approach to designing any technology that embeds privacy into the underlying specifications or architecture,"[186] FAPPs could guide policymakers and be incorporated into the design and implementation process. Designers and implementers would then be accountable to these principles depending on the nature of the automation, the risks involved, and the relevant area of law.[187] Related and relevant

[183] Robert Gellman, *Fair Information Practices: A Basic History* 1, 12 (Aug. 3, 2014), http://bobgellman.com/rg-docs/rg-FIPShistory.pdf [http://perma.cc/7G2K-XTLC].

[184] COLIN J. BENNETT & CHARLES D. RAAB, *The Governance of Privacy: Policy Instruments* in GLOBAL PERSPECTIVE 19 (2003).

[185] *See* Ira S. Rubinstein, *Regulating Privacy by Design*, 26 BERKELEY TECH. L.J. 1409 (2011); Ann Cavoukian, *Privacy by Design: The 7 Foundational Principles*, INFO. & PRIVACY COMM'N OF ONT. (Jan. 2011), http://www.ipc.on.ca/images/resources/7foundationalprinciples.pdf [http://perma.cc/ZXW7-RTGK].

[186] Rubinstein, *supra* note 185, at 1411–12.

[187] Human factors experts have been utilized to retrospectively assess design and implementation of products and systems since the 1970s but have been given varied receptions in courtrooms across the country. *See* James M. Doyle, *Applying Lawyers' Expertise to Scientific Experts: Some Thoughts About Trial Court Analysis of the Prejudicial Effects of Admitting and Excluding Expert Scientific Testimony*, 25 WM. & MARY L. REV. 619, 654 (1984); Frank D. Fowler, *Railroad Litigation and the Human Factors Expert: Why the Plaintiff Missed the Train*, 4 AM. J.

principles and codes of ethics have been developed elsewhere. FIPPs were drafted in the 1970s and have only recently begun to wane under the weight of technological change.[188] A code of ethics for big data was recently drafted by Neil Richards and Jonathan King,[189] and human-robot interaction ethics were developed by roboticist Laurel Riek and philosopher Don Howard.[190] Borrowing significantly, synthesizing where possible, and adapting slightly, the following are seven principles to tie into automation that complement existing design practices, account for actual use, and add necessary expertise. An important difference in the nature of each FAPP is that, unlike the FIPPs that direct what information practices must happen (e.g., data must not be altered until its original purpose is met), the automation principles are purposefully difficult to answer alone; they are intended to be collectively deliberated and developed by designers, managers, users, investors, policymakers, ethicists, and lawyers. In an effort to achieve technologies of humility that are inclusive and adaptive, the FAPPs are policy knots intended to be move as technology, use, and values adjust. Each is described and then discussed in terms of two complex forms of automation: big data automated decision making and self-driving cars. Big data has presented new challenges because it has allowed for new tasks to be digitally automated and others to be created. Defined by its size, speed, and variety, big data represents ways of discovering knowledge, solving problems, and improving inefficiencies.[191] The challenges that come with utilizing big data to automate decision making relate to its complexity, secrecy, and unpredictability. Self-driving cars are a product of big data in many ways. New vehicles have the capacity to navigate and maneuver modern streets because a great deal of information can be fed into and processed by the machine. The emerging new role of the human operator, previously as driver, is not well understood and has challenged established notions of reliance, accountability, and safety. Both big data decision making and self-driving cars create slightly

TRIAL ADVOC. 621, 624–29 (1981); Douglas R. Richmond, *Human Factors Experts in Personal Injury Litigation*, 46 ARK. L. REV. 333, 339–40 (1993).

[188] Fred Cate, *The Failure of Fair Information Practice Principles*, *in* Consumer Protection in the Age of the "Information Economy" 343, 343–44 (Jane K. Win ed., 2006).

[189] Neil Richards & Jonathan King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393, 408–26 (2014).

[190] Laurel D. Riek & Don Howard, *A Code of Ethics for the Human-Robot Interaction Profession*, 2014 WE ROBOT CONF., http://robots.law.miami.edu/2014/wp-content/uploads/2014/03/a-code-of-ethics-for-the-human-robot-interaction-profession-riek-howard.pdf [http://perma.cc/CBK8-49Q2].

[191] *See generally* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK (2014).

different man and machine systems. The human in the loop still exists as interactor and intervener, but in different ways. Both systems create new and different risks, while limiting others. Their complexities require innovative interfaces to effectively account for the humans in the loop—to maintain situational awareness and combat automation bias, for instance. Both big data and self-driving cars may exclude individuals or groups or have large-scale discriminatory effects, and both may be brought into (or create) sensitive situations related to intimate scenarios or involving private information. By incorporating existing design frameworks and identifying places where support is welcome or necessary, a set of FAPPs ties a policy knot that will allow for a more flexible and effective approach to emerging man-machine systems.

A. Risk

Automated systems should not be deployed without an assessment of risks to the human in the loop or humans impacted by the loop. The law's role in establishing what risks should be assessed, how they may be satisfactorily assessed, and enforcing this assessment is vital to all emerging technology and to tying a strong policy knot. Identifying harms and understanding benefits is incredibly challenging but must not be left solely to technology companies, innovators, or developers. The social costs are listed late in the older automation frameworks, but little guidance on how to actually assess those social costs is provided. Larger social costs are not built into the robotics framework above, but the operator is emphasized. In either case, more support from law, policy, ethics, and social sciences is needed. Risk in automated systems should be mindful of Jasanoff's discussion and corresponding criticism that industry and government provide tools for risk assessment, cost-benefit analysis, and predictive modeling even in highly uncertain contexts, but these tools are limited.[192] They are focused on the short-term, knowable risks and are less capable of accounting for ambiguities and long-term harms.[193] Additionally, risk portrayed as mathematic rationality can preempt political conversations by excluding those that cannot meet the demands of the dominant discourse and is necessarily limited by initial framing constraints.[194]

[192] Jasanoff, *supra* note 16, at 238–39.

[193] *Id.*

[194] *Id.*

There are a number of concerns surrounding the use of big data analytics including computational errors, misclassification, poor design, and unfair and negative impacts of application—to name a few.[195] Because big data occurs on the screen, the risks presented are different than some of the other forms of automation in this Article, but they are risks nonetheless. Pasquale reveals that by automating financial decisions based on health records, for example, a disenfranchised underclass can be created.[196] The problem, of course, is that while many systems will go through human-computer interaction (HCI) user testing (an uncertain number will not) to assess how the practice, the design process, does not include a greater risk assessment associated with use of the automation. Even models that explicitly call for risk assessment will struggle because, even in 2015, we are still trying to develop "a broader doctrine, one capable of assessing the various types of predictive inferences that shape life opportunities and of weighing their benefits against their harms," explains Dennis Hirsch.[197] Hirsch goes on to argue that the FTC can and should use its "unfairness authority" to "distinguish predictive practices that are in bounds, from those that are not."[198] Scholars must continue to work on developing risk assessment for harms that derive on the screen, as well as a future without them.

Self-driving cars are a major innovation following a long series of automation innovations from the electric starter, automatic transmission, power steering, and anti-lock braking to electronic stability control, cruise control, and automatic parking. Self-driving cars promise substantial safety benefits, traffic efficiencies, and energy conservation, but there are many complex parts that need to work properly together in conjunction with the human (operator or intervener) and various environmental contexts in order to achieve potential benefits. The safety risks are somewhat obvious—cars are big, dangerous machines—and so self-driving cars have been road tested. In April 2014, Google announced their vehicles had logged 700,000 autonomous miles and, in May 2015, unveiled a vehicle with

[195] *See generally* Kord Davis, Ethics of Big Data: Balancing Risk and Innovation (2012); Viktor Mayer-Schönberger & Kenneth Cukier, Big Data: A Revolution That Will Transform How We Live, Work, and Think (2014); Privacy, Big Data, and the Public Good (Julia Lane, Victoria Stodden, Stefan Bender, & Helen Nissenbaum eds., 2014).

[196] PASQUALE, *supra* note 36.

[197] Dennis D. Hirsh, *That's Unfair! Or is it? Big Data, Discrimination and the FTC's Unfairness Authority,* 103 KY. L.J. 345, 347 (2015).

[198] *Id.* at 5.

no steering wheel or pedals.[199] While these systems will not likely have the kinks worked out for another few years, the risks of them working properly or improperly, being innovated upon or abused, and used by the "good guys" or the "bad guys" has to be built into the structure of autonomy chosen for the man-machine systems that will once again change transportation.

B. Transparency

Automated systems should be comprehensible and support situational awareness through effective transparency. Unless there are no risks involved in the practice of an automated system, black boxes are bad design. This will be perceived as an unpopular, anti-innovation principle, but when an operator does not know what a system is doing because of the opaqueness of its design, then error recognition, intervention, and resolution are timely and costly, if not impossible. Pressing forward with black box designs (and the individual harms and social fallout that follow) will likely slow the integration of sophisticated automation and intelligent systems into society.

Big data automated decision making and self-driving cars seem to have different risks associated with them today, but self-driving cars depend on information gathering, processing, and decision analysis just as big data-driven automated decisions do. When discussing big data automated decision making, scholars like Citron and Pasquale have focused on the screen, featuring databases that sort out those that qualify for loans or welfare benefits.[200] While a self-driving car may cause physical and immediate harm, financial determinations can have significant negative impacts as well, and both must offer a level of transparency that allows for the human in the loop to optimize the system. It must be considered unacceptable to build a system to assess users, patients, students, clients, or applicants and then to abandon it. A human must understand what is happening and why.

Self-driving cars will need to be transparent enough for operators and interveners to maintain situational awareness and understand how to act if and when necessary, taking into account skill degradation and mental workload effects. Particularly because the

[199] Gail Sulivan, *Google's New Driverless Car Has No Brakes or Steering Wheel*, WASH. POST, (May 28, 2015), http://www.washingtonpost.com/news/morning-mix/wp/2014/05/28/googles-new-driverless-car-has-no-brakes-or-steering-wheel/ [http://perma.cc/2BEZ-2HR7].

[200] Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 6 (2014).

risks associated with driving are incredibly tangible and high, this amount of black boxing must be limited. Citron and Pasquale have argued for transparency in automated scoring systems that would provide accountability. This would require access to datasets, source code, programmer notes describing the variables, and correlations—anything required to "be able to meaningfully assess systems whose predictions change pursuant to AI logic."[201] These justifications are incredibly important, but stretch this principle of transparency, which focuses on the operator knowing what is happening with the system and why. However, transparency for the purposes of accountability is made possible by transparent systems generally.

C. Errors and Limitations

Automated system failures should not be surprising, silent, or irresolvable. "[S]ystems should never be permitted to fail in silence[;]"[202] impacts from errors occurring in digital automation may be immediate and buried in a pile of algorithms. Situational awareness, mental workload, skill degradation, and automation bias must be considered when designing error detection and considering limitations. This is particularly true when errors that occur infrequently and are associated with high risks or limitations of the system require only frequent—but skilled—interaction. Alarms must be set appropriately for the situational use of digital automation with the right level of sensitivity for the risks involved. Errors must not only be flagged but also resolvable, which relates to the need for transparency. Sometimes a system will work as intended but create unintended results, at which point intervention must be possible and effective. Even well functioning systems will have limitations, negatively impacting human interactors by creating inappropriate levels of trust or by decreasing situational awareness. These limitations may change as environments, tasks, or circumstances change. Human users should be aware of their own limitations as operators, interactors, or interveners, as well as the limitations of an optimized and well functioning man and machine system. Citron's work on public benefit systems reveals a large number of errors occurring without any good way to alert operators or resolve issues in a timely manner.[203] Systems may rely on humans to detect errors but

[201] *Id.* at 25.

[202] Billings, *supra* note 182, at 85.

[203] *See* Citron, *supra* note 55, at 1279.

must provide the tools to do so. The system must be designed with failure in mind. Depending on the nature of the errors, resolutions should be timely and fit the harm caused by the error. Self-driving cars will need to be designed in such a way that errors are sensitively set so that human operators will trust that something is wrong, instead of ignoring or disabling the alarm. Human factors engineers are experienced with interface design within the vehicle,[204] but the limitations of self-driving vehicles may provide a new challenge.[205] It is important for operators and implementers of automation to be aware of and understand the limitations of a system. For example, self-driving cars are not yet good at navigating leaf-covered roads, and data categories are often simplistic when assigning gender—an operator should be aware of this so that errors can be better assessed when they occur or go unnoticed by the machine.

D. Diversity and Discrimination

Automated systems should reflect on biases and choices during design and test for discriminatory impacts and diverse users. The lack of diversity amongst engineers and computer scientists has been well documented. Designers should consider their own backgrounds and physical features when reflecting on choices. Automation that creates barriers based on race, gender, or ability may want to be reconsidered. Systems should be tested for discriminatory impact. Big data's disparate impact has recently been the topic of many symposia and an FTC workshop.[206] Solon Barocas and Andrew Selbst explain:

> Sorting and selecting for the best or most profitable candidates means generating a model with winners and losers. If data miners are not careful, that sorting might create disproportionately adverse results concentrated within historically disadvantaged groups in ways that look a lot like discrimination. . . . Approached without care, datamining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision-makers, or simply reflect the widespread biases that persist in society.[207]

[204] *See generally* Human Factors in Automotive Design (Society of Automotive Engineers eds., 2001).

[205] Although one not unanticipated, see P.A. Hancock & Raja Parasuraman, *Human Factors and Safety in the Design of Intelligent Vehicle-Highway Systems (IVHS)*, 23 J. OF SAFETY RES. 181, 183–84 (1992), these limitations will be different than those in previous decades as users have gained some sense of how to interact with "smart" devices and environments.

[206] *See e.g.*, Latanya Sweeney, Chief Technologist, Fed. Trade Comm'n, Remarks at Big Data: A Tool for Inclusion or Exclusion? (Sept. 15, 2014).

[207] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. (forthcoming 2016), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899 [http://perma.cc/9EAU-XV77].

Using the Street Bump application as an example, Kate Crawford explains how the city of Boston must be careful in using the data collected from accelerometers in smart phones that detect when potholes are driven over because there is an uneven distribution of smartphones and application users across populations.[208] From collection and categories to analysis and use, the widespread use of big data automated decision making raises significant discrimination concerns. To illustrate, Barocas and Selbst find Title VII liability for disparate treatment and disparate impact ill-suited to the task of regulating discrimination in big data analytics used for employment.[209] Citron and Pasquale have proposed giving the FTC access to scoring systems that unfairly harm consumers so that the agency's technologists could test the systems for bias, arbitrariness, and unfair mischaracterizations.[210] Finding significant constitutional and political barriers to implementing needed reform, Barocas and Selbst suggest that non-legal solutions may be appropriate.[211] Incorporating Shilton's value levers related to discriminatory decisions and outcomes may have the most impact until reform is possible.

Self-driving cars will need to consider what biases are built into designs as well. Algorithmic decisions manage the system and will make difficult choices in what to hit when a vehicular accident is inevitable. Situations and circumstances unfamiliar to the designers may present themselves. The physical layout of the vehicle and the interfaces will need to account for the diversity of drivers, including various genders, ages, abilities, and backgrounds. Users will adapt to these systems, but whose driving experience will they adapt to? Automation designers should reflect on such questions.

E. Sensitive Situations

Automated systems should account for sensitive situations and information preferences of the humans in the loop. Sensitive situations, like those that deal with sensitive information, private places, or vulnerable populations should be assessed with an appropriate level of care and expertise. Of course, this principle is tightly tied to the previous principle of diversity and discrimination.

[208] Kate Crawford, *The Hidden Biases in Big Data*, HARV. BUS. REV. (Apr. 1, 2013), https://hbr.org/2013/04/the-hidden-biases-in-big-data/ [http://perma.cc/J3RB-HXLE].

[209] *See* Barocas & Selbst, *supra* note 207, at 46.

[210] Citron & Pasquale, *supra* note 200, at 25.

[211] *See* Barocas & Selbst, *supra* note 207, at 61.

It is important that automation is designed with foresight and diverse perspectives.

Big data conversations have previously been dominated by issues related to sensitive situations. Most notoriously, Target devised a system for identifying and targeting women early in pregnancy so as to capture their buying power from baby shower to first birthday. In doing so, the company mailed a teenage high school student living at home coupons for baby products.[212] The teenager was in fact pregnant, but Target had divulged this information to her parents, and its practices of "knowing" a woman was pregnant before she told friends and family were widely considered invasive.[213] Humans devised this system and were involved in its implementation, but no special attention was paid to the sensitive nature of using buying habits to suggest health conditions, informing others about pregnancy, or the age or cultural situations of those being targeted. More recently, Facebook manipulated what content appeared to 689,003 randomly selected users—positive or negative posts—to study how emotions spread across social media.[214] Although users consent to these kinds of tests when they agree to the site's terms of services by creating an account, it was troubling to many who were concerned about the impact of fiddling with users' emotional states without explicit notice.[215] "I wonder if Facebook KILLED anyone with their emotion manipulation stunt. At their scale and with depressed people out there, it's possible," privacy activist Lauren Weinstein posted on Twitter.[216] An outcry resulted because the study revealed the power platform operators have over users, that the company acted with limited foresight, and that preferences were not necessarily taken into account.[217] Personal information that is collected at the information acquisition stage, sorted through in the analysis phase, perhaps isolated in the decision phase, and acted upon in the action stage

[212] *See* Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES, (Feb. 16, 2012), http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all &_r=1 [http://perma.cc/N9QD-HC5M].

[213] *Id.*

[214] Vindu Goel, *Facebook Tinkers with Users' Emotions in News Feed Experiment, Stirring Outcry*, N.Y. TIMES, (June 29, 2014), http://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html [http://perma.cc/7DZU-LEFG].

[215] *Id.*

[216] Lauren Weinstein (@LaurenWeinstein), TWITTER (June 28, 2014, 5:55 PM), https://twitter.com/laurenweinstein/status/483051171255312384 [http://perma.cc/H6W3-GL62].

[217] Kashmir Hill, *After the Freak-Out Over Facebook's Emotion Manipulation Study, What Happens Now?*, FORBES (July 10, 2014), http://www.forbes.com/sites/kashmirhill/2014/07/10/after-the-freak-out-over-facebooks-emotion-manipulation-study-what-happens-now [http://perma.cc/BX9X-5DMS].

should not be overlooked and will be more relevant with increasing automation; a high level of consciousness must be built into an automation system's design and implementation, particularly when sensitive populations, information, or situations may be involved.

Self-driving vehicles will also be put in sensitive situations. Users may say or do things in vehicles or while driving that they may consider private and wish not to share with particular parties such as the government or an insurance company.[218] Of course, a number of features have already been added to cars that collect and track information, such as hands-free calling systems[219] and devices that allow for personalized insurance rates.[220] However, as more sophisticated automation is integrated into more environments and contexts, it may be increasingly difficult for users to provide consent or participate in information collection and use. Signaling information preferences, such as "do not map my face" or "do not identify me," is difficult in an Internet of Things[221] environment that provides many sources of collection with few screens available for members of the public to convey their preferences.[222] Information collected about others while driving may include fellow drivers and cars, pedestrians, cyclists, and activity on private and public property. How to collect, analyze, use, and share this information will need to be considered.

F. Man-Machine Comparison

An automated system's design and implementation should locate the human in the loop and reassess the system's impact on the human and larger social values. Instead of dividing up aspects of a task by talents of man versus talents of machine, this principle informs those in the development and implementation of automation to consider changes by allocating aspects of automation a certain way.

[218] Timothy B. Lee, *Self-driving Cars Are a Privacy Nightmare. And It's Totally Worth It*, WASH. POST, (May 21, 2013), http://www.washingtonpost.com/blogs/wonkblog/wp/2013/05/21/self-driving-cars-are-a-privacy-nightmare-and-its-totally-worth-it/ [http://perma.cc/A27P-JFDB].

[219] *See, e.g.*, ONSTAR, https://www.onstar.com/us/en/services/connections.html [http://perma.cc/JE4M-MPCL] (describing OnStar's hands free calling).

[220] *See, e.g.*, PROGRESSIVE, http://www.progressive.com/auto/snapshot-common-questions/ [http://perma.cc/P4XP-T6PJ] (describing Progressive's Snapshot, which tracks driving habits to "personalize your insurance rate based on your actual driving").

[221] "Internet of Things" refers to a category of objects newly fashioned with sensors and connected to the internet and each other. Bill Wasik, In the Programmable World, All Our Objects Will Act as One, Wired (May 14, 2013), http://www.wired.com/2013/05/internet-of-things-2/ [http://perma.cc/ZW5P-YP48].

[222] Meg Leta Jones, *Privacy Without Screens & The Internet of Other People's Things*, IDAHO L. REV. (forthcoming 2015).

We are often quick to point out the limitations of new forms of automation, but we must also consider the downsides of not utilizing the automation and relying more heavily on human engagement. This conception begins by locating and tracking the human in the loop in a more expansive way, considering an imperfect human in a loop as well as those humans tangentially impacted. We must develop a way to assess and think about this principle in a way that effectively locates and accounts for the ever-present humans in the loop. A discussion of what critical decisions are to be made by humans (and why) and how to limit automation bias and moral buffers in those instances would be an incredible contribution to the guidance of automation. Much of digital automation is incorporated incrementally. Big data automated decision making was not previously performed by a human using no tools at his disposal. Slight changes in function allocation or levels of autonomy must be reflected upon to assess changes in the system and their associated impacts. Bankston and Stepanovich perform this principle beautifully in their We Robot 2014 article, *When Robot Eyes are Watching You: The Law & Policy of Automated Communications Surveillance.*[223] The authors work through two scenarios, one where a set of computers is analyzing email communication and flagging it for further analysis and one where a human is doing the exact same tasks. Relying on cases that recognize the interdependence of man and machine, Bankston and Stepanovich find that invasions occur at the point of collection, not human interpretation.[224] For instance, in 1976, the Fifth Circuit established an important doctrine interpreting "interception" under the wiretap statute in *United States v Turk.*[225] When the police seized an audio recording of a telephone conversation from a vehicle, the question became whether the subsequent listening by human officers qualified as an interception of the call. The court explained that it did not—that the interception had occurred when the device collected the information, not when the humans heard the information.[226] Explaining further:

> In a forest devoid of living listeners, a tree falls. Is there a sound? The answer is yes, if an active tape recorder is present, and the sound might be thought of as "aurally acquired" at (almost) the instant the action causing it occurred. For § 2510(4) purposes, the recorder can be the agent of the ear.[227]

A previous case, *Smith v. Maryland,* dominates Fourth Amendment jurisprudence because it provides foundation for

[223] Bankston & Stephanovich, *supra* note 81.

[224] *See id.*

[225] United States v. Turk, 526 F.2d 654, 658–59 (5th Cir. 1976).

[226] *Id.*

[227] *Id.* at 658, n.2.

interpreting third-party doctrine in modern contexts. In *Smith*, the Supreme Court found no meaningful distinction between man and machine for the purposes of protecting privacy.[228] Instead, the Court found that monitoring non-content dialing information for twenty-four hours did not violate any expectations of privacy (and so was not an unconstitutional search), because it had previously held that individuals could hold no expectation of privacy for this information as it was knowingly disclosed to a third party (the telephone company):

> When he used his phone, petitioner voluntarily conveyed numerical information to the telephone company and "exposed" that information to its equipment in the ordinary course of business. In so doing, petitioner assumed the risk that the company would reveal to police the numbers he dialed. The switching equipment that processed those numbers is merely the modern counterpart of the operator who, in an earlier day, personally completed calls for the subscriber. Petitioner concedes that if he had placed his calls through an operator, he could claim no legitimate expectation of privacy. We are not inclined to hold that a different constitutional result is required because the telephone company has decided to automate.[229]

These cases represent man and machine legal approaches, because they consider the technology as an agent of humans or a part of a human system. In undertaking this comparison, Bankston and Stepanovich argue that expectations of privacy from invasions by the government are more vulnerable when analytics are not considered as a part of the search or seizure process because their scale is massive, even if they arguably lead to less embarrassment or self-consciousness surrounding specific personal information.[230]

Self-driving cars will need to make similar assessments, accounting for the human driver or operator who will have to interact or intervene in some situations whether to maneuver the vehicle through uninterpretable terrain, redirect a vehicle from a distance, or repair a machine that is failing at certain tasks. Courts will have to assess whether the human has been properly considered in the design of error detection systems and warnings. Shifting roles in man-machine systems require a reflection on how responsibilities shift as well and whether values are protected through that shift.

G. Predictability

Automated systems should be initially and continually inventoried for predictable and unpredictable behavior. One of the most challenging aspects of emerging technology is its unpredictability, but the novelty and value of emerging technology

[228] Smith v. Maryland, 442 U.S. 735 (1979).

[229] *Id.* at 744–45.

[230] Bankston & Stephanovich, *supra* note 81, at 1.

today is its adaptability and unpredictability. These aspects of modern automation will no doubt require a legal overhaul, but no one is in a better position than designers and implementers to assess a technology's type and level of unpredictability. Automation is predictable. A Roomba is not going to fly—its design prevents such action. A drone is not going to drive down the sidewalk. Security analytics are not going to start talking and giving people bad driving directions. Considering unpredictability supports mindful integration into society.

Judge Curtis E.A. Karnow described the types of technology most challenging for traditional tort liability to handle as those that operate using machine learning, such as "genetic algorithms, neural nets, or other sorts of feedback loops which generate unpredictable behavior."[231] He goes on:

> That is, these robots are given instructions as to ultimate goals and determine for themselves the means of accomplishing these goals. The means are not predictable by either the operator-owner or by the original programmers. Rather, the software teaches itself by running experiments or making other sorts of real or virtual attempts at a solution, corrects for error and approximates a result which it then implements.[232]

But everything and everyone is limited by design; we may beat the odds, anomalies will arise, and the unimaginable will raise eyebrows. However, if we are paying attention, we will be less shocked by technological outcomes. By inventorying predictable behavior, a range of possible adaptations, and tracking systems in the lab and in situated contexts (which requires transparency, risk assessment, and error detection), it will be more difficult to be caught off guard. A predictability inventory is an exercise that will help guide designers, the public, and policymakers. This is a principle that will gain refinement and precision with practice, but should be undertaken in order to analytically consider unpredictability.

A key feature of big data is that by processing massive amounts of data using elegant algorithms, new and unexpected insights are produced. Automated decision making based on unexpected insights can present problems when not designed properly or sensitively, but predictability is laced throughout what is considered an unpredictable innovation. The context in which the system operates provides some level of predictability. For instance, the Target study could reveal that woman who have recently become pregnant crave and buy alcohol even though risks increase when

[231] Curtis E.A. Karnow, *The Application of Traditional Tort Theory to Embodied Machine Intelligence*, 2013 WE ROBOT CONF., http://conferences.law.stanford.edu/werobot/wp-content/uploads/sites/29/2013/03/Karnow_Curtis.pdf [http://perma.cc/62PC-D5KD].

[232] *Id.*

alcohol consumption during pregnancy occurs.[233] Although surprising, this is not unpredictable. Target is looking for insights into what is consumed during early pregnancy. If Target decided to send coupons for diapers and alcohol to pregnant women, without anyone noticing or error detection catching the problem, it would be because the designers, operators, or implementers did not consider what was possible given the level of autonomy and function allocation between human and machine. Target's use of big data automated decision making is not likely to predict a number of scenarios or perform a number of tasks. Similarly, self-driving cars will behave unpredictably based on machine learning and operator peculiarities while remaining tied into the system and providing transparency, error detection, and other human-centered design and implementation decisions that will limit nasty surprises.

Each of the legal case studies above involved a harm perceived as one technology either caused or could resolve. None of them involved assessing the risks of a change in the loop. Certainly railroad safety and traffic safety would have been well served by situated-use testing to support the risks involved in adjusting the level of autonomy in the particular man-machine system. The EU data protection regulations do not consider whether forcing a human in the loop increases or decreases the likelihood of an error in a given situation.[234] These efforts are all well intentioned, but the man versus machine approach has not provided the right perspective to understand or complement the man-machine systems. The FAPPs provide a tool for approaching each, but do not specifically address any one problem. The set of principles is likely incomplete and certainly underdeveloped, but even guide rails need to be fine-tuned.[235]

VII. CONCLUSION

"It all just seemed so good the way we had it, back before everything became automatic," sings Miranda Lambert.[236] Nostalgia has always been a popular theme in country music but should be handled with care when contemplating technological change. Indeed, nostalgia ignores the new computational environment and the very different expectations humans situate themselves within.

[233] *See* discussion *supra* Section VI.E.

[234] *See supra* Section IV.B.

[235] Jarvis D. Michie & Bronstad E. Maurice, *Highway Guardrails: Safety Feature or Roadside Hazard?*, TRANSP. RESEARCH RECORD 1468 (1994).

[236] MIRANDA LAMBERT, *Automatic*, *on* PLATINUM (RCA Records 2014).

The Defense Science Board was right when it explained that defining levels of autonomy "deflects focus from the fact that all autonomous systems are joint human-machine cognitive systems, thus resulting in brittle design" and that "all systems are supervised by humans to some degree, and the best capabilities result from the coordination and collaboration of humans and machines."[237] Man and machine systems are the best systems but present the hardest problems. They are hard to design well and responsibly, hard to predict, and hard to govern. The ironies of automation have resulted because of a man-versus-machine approach to automation and the irony of automation law result for the same reason. A human will always be in the loop, at a minimum, as interactor or intervener in digital automation. By focusing on the humans in the loop, automation law may be able to resolve its own ironies.

This requires recognition of the fact that law can break a well-formed loop of the need for flexibility in man-machine design and implementation and an understanding of design and implementation practices. Moving a human in or out of the loop is both ineffective (because a human is always in the loop) and dangerous (because her existing role in the loop may have been the key to keeping the system reliable). Creating policy knots by aligning legal treatment with frameworks for responsible design, locating moments where legal expertise would serve the process, and identifying holes where important societal concerns are overlooked will help the law account for the flexibility necessarily involved in man-machine systems and adaptive to new innovations.

As we attempt to grapple with the new, identify and forecast innovation, and anticipate governance issues, it is important to acknowledge the incremental nature of innovation and tie where we are to where we are trying to go. This broad perspective provides an analysis that links a number of overlapping issues under a single, expansive, unifying concept but does not blind us to exceptional innovations that require a long, hard look at assumptions.

[237] DEF. SCI. BD., U.S. DEP'T OF DEF. TASK FORCE REPORT: THE ROLE OF AUTONOMY IN DOD SYSTEMS 23–24 (July 2012), http://www.fas.org/irp/agency/dod/dsb/autonomy.pdf [http://perma.cc/7S5Q-KSQP].