REGULATING THE LOOP: IRONIES OF AUTOMATION LAW

MEG LETA AMBROSE, JD, PHD
GEORGETOWN UNIVERSITY, COMMUNICATION, CULTURE & TECHNOLOGY

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 cock pits, the need for a legal approach suitable to guide an increasingly automated future becomes more pressing.

This paper 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 socio-technical 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 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 for automated system design and principles currently utilized for safe and actual implementation be added to for ethical and socio-technical legal approach to automation.

I. INTRODUCTION

In 1988, U.S.S. Vincennes military personnel shot down a passenger jet carrying 290 civilians, because their automated radar system, which had been designed to detect Soviet bombers, 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 have 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, and/or with computers. The humans + machine teams dominated the supercomputers operating the same brute number crunching strategies in place since the 70s. 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 loss must be toed carefully.

The human-automation relationship is not simple. "There will always be a human in the loop," is a statement made repeatedly by those seeking to quell public fears about military use of weaponized intelligent automated systems.[5] However, the loop is an ill-defined parameter[6] and in certain capacities, the human may not benefit, but harm, system performance.[7] Advancements in sensors, information processing, memory, statistics, and a number of other areas have increased

[1] Lieutenant Colonel 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/home/TabId/211/PostId/4002461.
[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.
[4] Garry Kasparov, "The Chess Master and the Computer," NEW YORK REVIEW BOOKS (Feb. 11, 2010), http://www.nybooks.com/articles/archives/2010/feb/11/the-chess-master-and-the-computer/?pagination=false.
[5] P.W. Singer, Wired for War, 123-124 (2009).
[6] William C. Marra and Sonia K. McNeil, *Understanding "The Loop": Regulating the Next Generation of War Machines*, 36:3 HARVARD JOURNAL OF LAW & PUBLIC POLICY 1139 (2012).
[7] Raja Parasuraman, Thomas B. Sheridan, & Christopher D. Wickens, *A Model for Types and Levels of Human Interaction with Automation*, SYSTEMS, MAN, AND CYBERNETICS 286, 291 (May 2000). Raja Parasuraman, *Humans and Automation: Use, Misuse, Disuse, Abuse*, 39:2 HUMAN FACTORS 230 (1997); Lorrie Cranor, *A Framework for Reasoning About the Human in the Loop*, UPSEC'08 PROCEEDINGS OF THE 1ST CONFERENCE ON USABILITY, PSYCHOLOGY, AND SECURITY (2008); Lisanne Bainbridge, *Ironies of Automation*, 19:6 AUTOMATICA 775 (1983).

the adoption of automated systems[8] into nearly every aspect of society.[9] Although research on human-robotic teams,[10] human-robot interaction,[11] and human factors in systems engineering[12] is growing, it is not often utilized to make well-informed technology policy related to automated systems.

Law and policy institutions have had to confront automated systems in somewhat limited instances, resulting in a small precedence across a range of contexts. Although these scattered instances have not yet been addressed comprehensively, the themes from these policy decisions will without intention create a foundation inevitably relied upon as automated systems penetrate society. Reflection on early policy treatment of the human in the loop can provide an assessment of legal approaches to help escort ethical and thoughtful incorporation of automation moving forward.

This paper will first present five case studies from different areas of law and policy including railroad legislation, robocall regulations, fourth amendment search decisions, state-level automated traffic enforcement, and European data protection regulations. The select cases offer a breadth of examples covering legislative, judicial, and administrative perspectives at the state, national, and international level and are used to discover themes in regulation related to the human in the loop. They reveal that legal approaches focused on the capabilities of the automation may backfire, what I call the irony of automation law for its resemblances to the

[8] "Automated systems," for the purpose of this article, are machine processes that have "the capacity to operate without outside intervention." This includes systems along a spectrum of intelligence levels and learning capabilities. O. Grant Clark et al., MIND AND AUTONOMY IN ENGINEERED BIOSYSTEMS, 12 ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE 389 (1999), available at http://dx.doi.org/10.1016/S0952-1976(99)00010-X.

[9] Christopher Steiner, AUTOMATE THIS: HOW ALGORITHMS CAME TO RULE OUR WORLD (2012); Evgeny Morozov, TO SAVE EVERYTHING, CLICK HERE: THE FOLLY OF TECHNOLOGICAL SOLUTIONISM (2013).

[10] *See e.g.*, Julie L. Marble, et al., *Evaluation of Supervisory vs. Peer-Peer Interaction with Human-Robot Teams*, PROCEEDINGS OF THE 37TH ANNUAL HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES (2004); Laura Hiatt, Anthony Harrison, & Greg Trafton, *Accommodating Human Variability in Human-Robot Teams Through Theory of Mind*, 3 IJCAI'11 PROCEEDINGS OF THE TWENTY-SECOND INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE 2066 (2011); Matthew Johnson, et al., *Beyond Cooperative Robotics: The Central Role of Interdependence in Coactive Design*, 26:3 INTELLIGENT SYSTEMS 81 (May-June, 2011).

[11] *See e.g.*, Goodrich & Schultz, *Human-Robot Interaction: A Survey*, 1:3 FOUNDATIONS AND TRENDS IN HUMAN-COMPUTER INTERACTION 203 (2007); Sarah Kiesler, et al., *Fostering Common Ground in Human-Robot Interaction*, IEEE INT. WORKSHOP ON ROBOT AND HUMAN INTERACTION COMMUNICATION 729 (2005);

[12] *See e.g.*, M.L. Cummings, *Human-Automation Collaboration in Complex Multivariate Resource Allocation Decision Support Systems*, 4:2 INTERNATIONAL JOURNAL OF HUMAN COMPUTER STUDIES 616 (2010); John D. Lee and Katrina A. See, *Trust in Automation: Designing for Appropriate Reliance*, 46:1 HUMAN FACTORS 50 (Spring 2004); Robert W. Proctor, HUMAN FACTORS IN SIMPLE AND COMPLEX SYSTEMS (2nd ed.) (2008); Kim Vicente, THE HUMAN FACTOR: REVOLUTIONIZING THE WAY PEOPLE LIVE WITH TECHNOLOGY (2004); Thomas B. Sheridan, HUMANS AND AUTOMATION: SYSTEMS DESIGN AND RESEARCH ISSUES (2002).

Ironies of Automation.[13] This irony is more fully investigated as a neglect of the socio-technical nature of automation. Finally, the article assesses these themes in light of human factor and systems engineering research suggesting that utilization of existing models may be incorporated into future policy efforts to govern automated systems.

II. THE LAW AND THE LOOP

At WeRobot 2012, Neil Richards and William Smart asked how should the law think about robots?[14] This article asks how *has* the law thought about robots? More precisely, what kind of conclusions can we draw about previous legal treatment and how should these approaches inform polices governing further integration of sophisticated automation across society?

Broadly, automation includes all the ways computers and machines help perform tasks more quickly, accurately, and efficiently - for people. The term automation refers to (a) the mechanization and integration of the sensing of environmental variables through artificial sensors; (b) data processing and decision making by computers; and (c) mechanical action by devices that apply forces on the environment or information action through communication to people of information processed.[15] The term encompasses open-loop operations[16] and closed loop control,[17] as well as intelligent systems.[18]

> Computers have continued to become smaller, faster, more powerful and cheaper. Automation has moved from open-loop mechanization of 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.[19]

[13] Lisanne Bainbridge, *Ironies of Automation*, 19:6 AUTOMATICA 775 (1983).
[14] Neil Richards and William Smart, *How Should the Law Think About Robots?*, WEROBOT CONFERENCE (2012).
[15] Thomas B. Sheridan, HUMANS AND AUTOMATION: SYSTEM DESIGN AND RESEARCH ISSUES 9-10 (2002).
[16] Open loop controls have no measurement of system output or feedback.
[17] In a closed loop control system, the output is monitored and fed back to a control to make adjustments.
[18] 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).
[19] Thomas B. Sheridan, *Function Allocation: Algorithm, Alchemy or Apostasy?* 5:2 INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES 205 (2000).

Older automation was not itself mobile, held minimal purpose-specific sensors, and operated with limited processing power, but ubiquitous computing means ubiquitous automation of many functions of many tasks.[20]

While the focus of the paper is larger than Richard and Smart's focus of robots as non-biological autonomous agents, the expansion to automation across the spectrum of intelligence and levels of autonomy avoids anthropomorphic metaphors by establishing the conversation firmly within mechanical automation and adding levels of computational sophistication.

The case studies below are old enough to give some sense of the effectiveness of the law, and so 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, socio-technical policy approach to automation.

A. LEGISLATIVE

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.[21] Implemented in 1904, the Act was intended to reduce the staggering injuries and deaths surrounding railroad worker safety. In 1984, one in 428 employees were in killed and one in 33 were injured, totaling 25,245 employees. After compliance with the automatic couplers and brakes was established, one in 357 died and one in 19 were injured.[22] The increase was a dramatic blow to those that felt railways represented an important progress for the US and that safety could be achieved for this innovation. Policymakers saw humans being injured

[20] Raja Parasuraman and Christopher D. Wickens, *Humans: Still Vital After All These Years of Automation*, 50 HUMAN FACTORS 511 (2008).
[21] Safety Appliance Act of 1893 (27 Stat. 531, 45 U.S.C. § 1 *et seq*.).
[22] S. W. Usselman, *The Lure of Technology and the Appeal of Order: Railroad Safety Regulation in Nineteenth Century America*, 21 BUSINESS AND ECONOMIC HISTORY 290 (1992).

by a task that could be automated and so they automated it, but 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.[23] 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.[24] Safety First programs were then initiated by Chicago and North Western Railway in 1910, and by 1918, all Class I railroads were required by law to adopt similar programs.[25] Statistics in employee injuries and fatalities began to improve after the initial increase dropping by 75% from 1920 to 1940.[26] After a more comprehensive approach was taken to address railroad employee safety and companies took an active role in decreasing injuries, the human in the loop is still a pivotal part of the regulatory equation. Once worker safety stabilized, attention shifted to grade-crossing safety, an area where automation has played a significant role in decreasing train-vehicle collisions[27] but have negatively impacted driver understanding of appropriate action where only the standard mandated signs are in place.[28] Today the most hotly contested and relevant railroad law is the Rail Safety Improvement Act of 2008, which requires the costly installment of Positive Train Control technology[29] to eliminate the types of human errors responsible for the Metrolink commuter train derailment in Los Angeles that killed 25 and injured hundreds.[30]

---

[23] Accident Reports Act of 1910 (49 U.S.C. §§ 20901–20903).
[24] Safety Appliance Act of 1910 (36 Stat. 298, 45 U.S.C. § 11 *et seq*.).
[25] Ian Savage, THE ECONOMICS OF RAILROAD SAFETY 25-26 (1998).
[26] *Id.*
[27] "U.S. Railroad Safety Statistics and Trends," Association of American Railroads (2006) (includes table entitled "Grade Crossing Warning Device Upgrades Work Gates Cut the Accident Fatality Rates By 93%."). By law, railroads have must only sound the engine's horn, obey speed limits, and properly maintain tracks at non-gated crossings, but where automated gates have been put in place, the railroad company has paid for their installation and must also pay for their upkeep. It should also be mentioned that numerous factions contributed to the significant decline of railway crossing accidents over the last forty years, with two-fifths of the decrease being attributed to reduced drunk driving and improved emergency medical response and a fifth attributed to crossing warnings and gates. Shannon C. Mok and Ian Savage, *Why Has Safety Improved at Rail-Highway Grade Crossings?*, 25:4 Risk Analysis 867 (Aug. 2005).
[28] "Safety at passive grade crossings, Volume 1: Analysis", Report NTSB/SS-98/02, National Transportation Safety Board, Washington, DC, 1998; "Safety at passive grade crossings, Volume 1: Analysis", Report NTSB/SS-98/02, National Transportation Safety Board, Washington, DC, 1998.
[29] Rail Safety Improvement Act of 2008 (49 U.S.C. § 20156).
[30] Positive Train Control essentially creates a smart track system where a communication network connects trains, centralized dispatchers, and track signals. The system monitors speed, configuration and switches so the system

B. ADMINISTRATIVE

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.[31] In the late 1980s, robocalls came under regulatory scrutiny, because the automation was considered more invasive than human callers -the rate at which they could invade the home and later the pockets and purses of individuals was much more efficient.[32] 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 automated calls and today has more than 221 million numbers on it.[33] The FCC responded with regulations that prohibit autodialing[34] 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.[35] 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.[36]

Judge Easterbrook explained the justification for regulatory variation between human and robot callers, "A human being who called Cell Number would realize that 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."[37] 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

---

knows where the train is, how fast it's going and its potential for accidents and will assume control if the operator fails to when an issue arises.

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

[32] S. W. Waller, D. Heidtke, and J. Stewart, *The Telephone Consumer Protection Act of 1991: Adapting Consumer Protection to Changing Technology*, Loyola University Chicago School of Law Research Paper No. 2013-016 (Sept. 17, 2013), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2327266.

[33] "The Do not Call Registry," Federal Trade Commission news, http://www.ftc.gov/news-events/media-resources/do-not-call-registry.

[34] 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." FCC Enforment Advisory, Enforcement Advisory No. 2012-06 (Sept. 11, 2012), available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-12-1476A1.pdf.

[35] 47 CFR 64.1200.

[36] FCC Enforment Advisory, Enforcement Advisory No. 2012-06 (Sept. 11, 2012), available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-12-1476A1.pdf.

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

callers,[38] and reaching voters with automated support or fully automated systems is one way that candidates with smaller bank accounts can promote their message and candidate.[39] By creating this distinction instead of enforcing recipient choice, regulators limit the benefits of political calls without protecting citizens from unwanted privacy invasions.

C. JUDICIAL

As Judge Easterbrook's quote represents, courts have also commented on the human in the loop. Other laws 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. No "search" by government agents 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. In *United States v. Karo* (1984), the Court explained that Karo'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.[40]

Here, 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. 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, systematic invasions. In June, 2013, a National Security Agency program called Prism was brought to the public's attention revealing the government collection of metadata through companies like Verizon, Google and Facebook.[41] Ruling that the bulk collection of American telephone metadata is unconstitutional, Justice Leon, writing for the D.C.

[38] G. Dvorsky, "Robots So Realistic They Can Deny They're Bots," Discovery News (Dec. 12, 2013), available at http://news.discovery.com/tech/robotics/robots-so-realistic-they-can-deny-theyre-bots-131212.htm.
[39] J. C. Miller, *Regulating Robocalls: Are Automated Calls the Sound of, Or a Threat to, Democracy?*, 16 MICH. TELECOMM. TECH. L. REV. 213 (2009).
[40] *United States v. Karo*, 468 U.S. 705 (1984).
[41] Steven Nelson, "Nine Companies Tied to PRISM, Obama Will be Smacked With Class-Action Lawsuit Wednesday," US News (June 11, 2013), available at http://www.usnews.com/news/newsgram/articles/2013/06/11/nine-companies-tied-to-prism-obama-will-be-smacked-with-class-action-lawsuit-wednesday.

District Court, 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.[42]

There are a number of legal issues related to the Prism program including 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.

D. STATE

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).[43] 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 office or installed in a law enforcement vehicle or facility.[44]

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.[45] The 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

[42] Klayman v. Obama, Civil Action No. 13-0881, at 64, (D.D.C. filed Dec. 16, 2013).
[43] "Speed and Red Light Camera Laws," Governors Highway Safety Association, available at http://www.ghsa.org/html/stateinfo/laws/auto_enforce.html.
[44] *Id.*
[45] Maggie Clark, "Red Light Cameras Generate Revenue, Controversy," USA Today (Oct. 15, 2013).

cameras have made many intersections more dangerous because human drivers brake differently at these intersections resulting in more rear collisions.[46] Additionally, while automation of enforcement is intended to be accurate, equal, and consistent and particularly suitable for determinable legal conclusions like traffic violations, *Do Robots Dream of Electric Laws? An Experiment in the Law as Algorithm* project presented at WeRobot 2013 reveals significant variation in the number and types of citations issued.[47]

E. INTERNATIONAL

The European Data Protection Directive of 1995 includes a 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."[48] 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 that have historically received it. The right is certainly less disruptive than an all-out ban on automated decision-making, but still has done little to protect against the different types of bias or errors that derive from both humans and automation.

Much 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, or generate data about, aspects relating to natural persons or to analyse or predict a natural person's performance at work, economic situation, location, health, preferences, reliability, behaviour or personality"[49] in the Data Protection Regulation, which is set to update the Data Protection Directive. For instance, the main provision in the Regulation is

[46] Carl Bialik, "Seeing Red," Wall Street Journal, Feb. 1, 2013.
[47] Lisa Shay, Woodrow Hartzog, John Nelson, and Gregory Contri, *Do Robots Dream of Electric Laws? An Experiment in the Law as Algorith*, WeRobot 2013.
[48] Directive 95/46/EC of the European Parliament and of the Council of 24.10.1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (O.J. L 281, 23.11.1995, p. 31 et seq.), Article 15(1).
[49] LIBE Committee, Compromise Amendments on Articles 1-29, Art. 4(3a) (Oct. 7, 2013), available at http://www.europarl.europa.eu/meetdocs/2009_2014/documents/libe/dv/comp_am_art_01-29/comp_am_art_01-29en.pdf.

in Article 20.[50] 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 Committee) through its amendments.[51] The method of utilizing a human to protect against harms caused by automation was 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."[52]

Drawing lines between man and machine in data processing allows a prejudiced human to inject bias and a complacent human to ignore the bias of the automation. Like the above examples, data protection regulation has had to provide additional guidance and safeguards to support the effectiveness of regulating the human in or out of the loop to promote and protect human values. "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 socio-technical system."[53]

III. IRONIES OF AUTOMATION LAW

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.[54] This is ironic, of course, because as a human, the designer is unreliable and inefficient. She delegates the easy tasks of the automation operator (human) making the difficult aspects more difficult. The human in the loop is left with difficult 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, makes the operator's life more difficult and worse. 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 because rare abnormal conditions are

[50] 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), COM final (Jan. 25, 2012), Art. 20 (2012).
[51] LIBE Committee, *supra* note 49, Article 20, at 49.
[52] *Id*., Article 20 Recital 58, at 58.
[53] Cynthia Dwork and Deirdre K. Mulligan, "It's Not Privacy, And It's Not Fair," 66 Stan. L. Rev. Online, 35 (Sept. 3, 2013).
[54] Bainbridge, *supra* note 13.

difficult to detect as automation bias builds and situational awareness declines. When inevitable errors occur, the complacency and skill degradation of the operator results in a decreased ability to perform when needed. Therefore, the most successful automation systems, those that fail on the rarest occasions causing need for manual intervention, require the greatest investment in operator training. In short, the more advanced and reliable the automation, the more important the human operator must be.

Regulators attempting to protect certain human values by focusing on the technological capabilities of automation end up providing less protection that previously in place. In other words, requiring or prohibiting a human in the loop based on what can be automated works against the interests of the regulation, the irony of automation law. 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 automated telemarketers are sophisticated to a level that offers the same recipient action as a human with less invasive treatment. Courts determined that a machine cannot violate privacy because it cannot judge an individual, but today automated tracking, surveillance, and processing reveals more about us 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. Bans on automated decision making in Europe have protected against neither human nor machine error.

Requiring a human in or out of the loop results in this irony because ignores the socio-technical nature of automation, similarly to the way automation-centered design is often counter-productive. It may be (1) detrimental, as system engineering research has shown that proper human-automation design is nuanced; (2) ineffective and redundant, as a human will always be in the loop; and (3) reactive, as regulating specific loops as issues arise is no longer a desirable governance strategy. This automation centered legal treatment of the issues sets a dangerous precedence, just as it did for design.

A. HUMAN FACTORS

Automation is no more a solution to protecting human values than humans. Automation can lead to the deterioration of human operator skills, increase operator workload, complacency, and

situational awareness resulting in a decline of safety and performance.[55] Automation reliability may lead to over or under trust of the system by operators, who may commit misuse, abuse, or disuse of an automation system do to any number of the above factors.[56] Requiring a human in or forcing a human out of the loop does not necessarily improve a system. It may in fact be detrimental, but certainly restricts flexibility to achieve a safe and realistic system.

Automation has consequences on human operator performance, categorized as mental workload, situation awareness, complacency, and skill degradation.[57] The following are examples of how automation may impact the human operator and the human-automation system.

Human-automation interaction research suggests that automation can improve mental workload by organizing, prioritizing, summarizing, highlighting, filtering, transforming (e.g., data into visual), and filtering, but can *increase* mental workload if 'clumsy' creating difficult engagement and data entry.[58] "In general, the effect of automation on mental workload has been mirrored by the similarly mixed record of automation in improving human productivity and efficiency."[59] Situational awareness refers to "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."[60] Automating decision-making can degrade operator's situational awareness of the system because the human becomes less familiar with the changes controlled by an automated agent, often losing a clear 'picture' of the informational environment without active engagement.[61] This is likely when decision making is in the form of monitoring for intervention to prevent errors and incidents.[62] Complacency (over-trust or over-reliance of

[55] Sheridan, *supra* note 15.
[56] Raja Parasuraman, *Humans and Automation: Use, Misuse, Disuse, Abuse*, 39:2 HUMAN FACTORS 230 (1997).
[57] Sheridan, *supra* note 15; Raja Parasuraman, Thomas B. Sheridan, & Christopher D. Wickens, *Situational Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs*, 2:2 JOURNAL OF COGNITIVE ENGINEERING AND DESIGN MAKING 140 (Summer 2008).
[58] Raja Parasuraman, *Adaptive automation matched to human mental workload*, 355 NATO SCIENCE SERIES SUB SERIES I LIFE AND BEHAVIOURAL SCIENCES 177-193 (2003); Pamela S. Tsang and Glenn F. Wilson, *Mental Workload*, in Handbook of Human Factors and Ergonomics (2nd ed., Gavriel Salvendy (ed.)) 417-449 (1997); Earl L. Wiener, *Human Factors of Advanced Technology ('glass cockpit') Transport Aircraft*, NASA Technical Report 117528 (1989), available at http://human-factors.arc.nasa.gov/publications/HF_AdvTech_Aircraft.pdf.
[59] Parasuraman, et al., *supra* note 7 (citing Thomas K. Landauer, THE TROUBLE WITH COMPUTERS (1995)).
[60] Mica Endsley, *Situational Awareness*, in Handbook of Human Factors and Ergonomics (3rd ed., Gavriel Salvendy (ed.)) 528-542 (2006).
[61] Mica Endsley, *Level of Automation and Situation Awareness*, in AUTOMATION AND HUMAN PERFORMANCE: THEORY AND APPLICATIONS, Raja Parasuraman and Mustapha Mouloua (eds.) 163-181 (1996); David B. Kaber and Mica R. Endsley, *Out-of-the-Loop Performance Problems and the Use of Intermediate Levels of Automation for Improved Control System Functioning and Safety*, 37 HUMAN FACTORS 381 (1995).
[62] *Id.*

automation) by the human operator has been shown to occur in both information and decision automation.[63] The effect of inappropriate reliance develops in highly - but not perfectly - reliable automation systems. For instance, the operator is supposed to monitor the automation because it is not perfectly reliable but without engagement with the information sources the operator fails to detect the occasional times when automation does fail.[64] Complacency is more pronounced when multiple tasks beyond just monitoring are undertaken by the human,[65] and attention cueing - an alarm to guide attention - has been shown to lead operators to pay less attention,[66] neither of which has been overcome by training or instruction.[67] Skill degradation involves mental or physical skill decay that occurs with disuse.[68] Essentially automating can make us rusty and requires additional training and practice.

Each of the above costs to the operator and system is influenced by the reliability of the automation. For instance, automation benefits to workload and situation awareness do not hold when automation is unreliable.[69] The automation increases the amount of work on the operator who must double check the information processes which may be more difficult to uncover due to automating the process. If the operator is aware of the level of unreliability and has access to unfiltered data, information acquisition and analysis can still be automated at somewhat high levels.[70] On the other hand, highly reliable (but of course, not completely reliable) systems can lead to automation bias, when a person does not acknowledge or seek contradictory information

[63] John D. Lee and Katrina A. See, *Trust in Automation: Designing for Appropriate Reliance*, 46 HUMAN FACTORS 50 (2004); Raja Parasuraman, Robert Molloy, & Indramani L. Singh, *Performance Consequences o Automation-Induced "Complacency,"* 79 INTERNATIONAL JOURNAL OF PSYCHO-ANALYSIS 349 (1998); John D. Lee and Neville Moray, *Trust, Self-Confidence, and Operator's Adaptation to Automation*, 40 INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES 153 (1994); John. D. Lee and Neville Moray, *Trust, Control Strategies and Allocation Function in Human Machine Systems* 22 ERGONOMICS 671 (1992);
[64] *Id*.
[65] *Id*.
[66] Christopher D. Wickens, *Imperfect and Unreliable Automation and Its Implications for Attention Allocation, Information Access, and Situation Awareness*, Aviation Research Lab Technical Report ARL-00-10/NASA-00-2 (2000), available at http://www.aviation.illinois.edu/avimain/papers/research/pub_pdfs/techreports/00-10.pdf; Christopher D. Wicken, Rena Conejo, & Keith Gempler, *Unreliable Automated Attention Cueing for Air-Ground Targeting and Traffic Maneuvering*, in Proc. 34th Annual Human Factors and Ergonomics Society (1999).
[67] Raja Parasuraman & D. H. Manzey, *Complacency and Bias in Human Use of Automation: An Attentional Integration* 52:3 HUMAN FACTORS 381 (June 2010).
[68] Andrew M. Rose, *Acquisition and Retention of Skills*, in APPLICATION OF HUMAN PERFORMANCE MODELS TO SYSTEM DESIGN, Grant R. McMillian (ed.) 419-426 (1989); S. Baron, *Pilot Control*, in Human Factors in Aviation, Earl L. Weiner & D. C. Nagel (eds.) (1988); Jerry M. Childs and William D. Spears, *Flight-Skill Decay and Recurrent Training*, 62:1 PERCEPTUAL AND MOTOR SKILLS 235 (1986).
[69] Parasuraman, Sheridan, & Wickens, *supra* note 7 at 291.
[70] Wickens, Conejo, & Gempler, *supra* note 66.

when presented with a computer generated solution.[71] She accepts the information as correct or truthful. "Automation reliability is an important determinant of human use of automated systems because of its influence on human trust. Unreliability lowers operator trust and can therefore undermine potential system performance benefits of automation."[72]

The way that systems are supposed to work and the way they actually work when situated are different.[73] Forcing a human in or out of the loop does not account for the interdependence between man and machine in automated systems. By ignoring the actual ways in which human-automation interaction plays out, automation-centric law can be detrimental to the values it intends to protect.

B. HUMAN(S) ALWAYS IN THE LOOP

The loop is as flexible a term as automation, intelligence, and autonomy. "In fact, virtually any machine could be considered fully autonomous if we define the grain size of its task to be sufficiently small."[74] Consider two versions of the loop. If a human were required to be in a reliably functioning loop (sense, think, act), automation may be allowed for sense (e.g., aerial view of combatant territory), think (e.g., process data and suggest appropriate decision), but perhaps not act (e.g., release weapon). If a human were required to be in an unreliable, malfunctioning loop, she could serve as a supervisor or monitor to do any of the three (sense, think, act) when necessary. No system is perfectly reliable, so defining the loop as broadly as a 'less than perfectly reliable sense-think-act system' would result in a human in the loop always.[75]

Human involvement in a system is simplistically viewed as serving as a fail-safe or performing aspects of the system the machine cannot perform. These types of involvement are not trivial and have led many to assume a human will always *need* to be 'in the loop.' However, history has shown many instances where humans have been removed from systems that perform the previously human function (or eventually remove that function), operate at appropriately safe levels, and still allocate legal and social responsibility. History has shown the opposite: accidents caused by mismanaged systems leaving those involved confused about who should have been responsible.

[71] Parasuraman, *supra* note 56; Lee & See, *supra* note 63.
[72] Parasuraman, Sheridan, & Wickens, *supra* note 7 at 291.
[73] Lucy Suchman, HUMAN-MACHINE RECONFIGURATIONS: PLANS AND SITUATED ACTIONS (2006).
[74] Matthew Johnson, et al., *Beyond Cooperative Robotics: The Centrol Role of Interdependence in Coactive Design*, 26 INTELLIGENT SYSTEMS 81 (May/June, 2011).
[75] Raja Parasuraman and Christopher D. Wickens, *Humans: Still Vital After All These Years of Automation*, 50:3 HUMAN FACTORS, 511 (June 2008).

Design, maintenance and accountability will keep a human in the loop, broadly defined, but I find three other justifications for recognizing the human in the loop, even after machine operation status or high levels of autonomy have been reached across a range of social settings. These justifications are integration, optimization, and interaction.

First, humans will be required to integrate sophisticated automation and robots into their roles and society. "As the human operator in actual systems moves to successively higher levels of supervisory control, the single and most important task left to the human operator is that of setting the objective function – deciding and communicating to the computer what is good and what is bad."[76] Advances have not reached the point of fully-autonomous robots and arguably never will (debates about the degree and possibility of human autonomy rage on). The day when a robot will be useful and functional 'out of the box' without human intervention has not yet arrived, with a few arguable exceptions (e.g., the Roomba, a robot vacuum); in the meantime, humans spend a lot of time integrating, training, and socializing robots before releasing them into the world. Once there, robots are continually monitored, supervised, and subject to upkeep. Human interaction with the Roomba includes emptying the machine, cleaning missed areas, and moving down stairs. Some machines need more tending than others, but few will enter the world and reach a level of autonomy that will not involve a human guidance.

Second, there are extraordinary gains to be made by creating human-machine teams, as the open chess tournament victors exemplify. In order to reach optimization, human-machine systems should be understood as socio-technical systems that do not ignore the human or social contribution to automation and vice versa. The term socio-technical was coined by Trist while studying the curious failings of coal mines to increase productivity after major investments in increased mechanization.[77] Adapting practices and organization evolved the way in which worker and machine were able to achieve better outcomes together, than either had separately. This does not have to be a philosophy about the skills and capabilities of man and machine, although computational innovations in chess have certainly taken that tone. IBM's Watson on its own deserves the response Pablo Picasso gave: "But they are useless. They can only give you answers." Watson, now being used in medical research, paired with a physician that did not

[76] Thomas B. Sheridan, *Human Centered Automation: Oxymoron or Common Sense?*, 1 SYSTEMS, MAN AND CYBERNETICS 823 (1995).
[77] E.L., Trist, G.W. Higgin, H. Murray, & A.B. Pollock, ORGANIZATIONAL CHOICE: CAPABILITIES OF GROUPS AT THE COAL FACE UNDER CHANGING TECHNOLOGIES: THE LOSS, REDISCOVERY & TRANSFORMATION OF A WORK TRADITION (1963).

know its functions or capabilities would not even be that useful. As IBM Research recognizes, medical research service is optimized by creating an interactive system that is natural leading to "more informed and accurate decisions faster and… new insights from electronic medical records (EMR)."[78] Humans will remain in the loop for optimization. Kevin Kelly explains, "This is not a race against the machines. If we race against them, we lose. This is a race with the machines. You'll be paid in the future based on how well you work with robots."[79]

Finally, multipurpose automated machines and autonomous agents will interact with the human world whether through their designed functions or real-world environment.[80] Factory robots have been behind cages for years, but now are introduced into production lines and warehouse floors surrounded by humans going about their business. Similar to the way humans work with other humans (supported by automated and non-automated tools) to perform the numerous tasks and achieve the many goals in their daily lives, humans will work with robotic and intelligent systems to go about their daily lives - creating a loop in which they are necessarily a part. These loop actors are intertwined. For instance, Jeanette Blomberg and Julian Orr's work at the Xerox Palo Alto Research Park on users' perception of machine reliability was heavily impacted by the technicians called in when service was requested.[81] Advancements in automation actually involve more humans, not less. At a minimum, those situations most pressing for regulators will be interactive. This concept of the loop accounts for the human involvement in the loop in a more expansive way than has been previously discussed, but the loop cannot be defined without accounting for the human data that is input, the impacts from the output, or their involvement feedback, not to mention their role as operators, managers, or technicians. An expansion is necessary to incorporating ethical concerns that will need to be accounted for if optimization is going to be reached.

C. REACTIVE

[78] "IBM Research: WatsonPaths," IBM Research, http://www.research.ibm.com/cognitive-computing/watson/watsonpaths.shtml#fbid=o0dSo3XZl-a.
[79] Kevin Kelly, "Better Than Human: Why Robots Will – And Must – Take Our Jobs, WIRED (Dec. 24, 2012, available at http://www.wired.com/gadgetlab/2012/12/ff-robots-will-take-our-jobs/all/.
[80] Linda J. Skitka, Kathleen Mosier, and Mark D. Burdick, *Accountability and Automation Bias*, 52 INT. J. HUMAN-COMPUTER STUDIES 701 (2000).
[81] Jeanette Blomberg, *Social Interaction and Office Communication: Effects on User's Evaluation of New Technologie*, in TECHNOLOGY AND THE TRANSFORMATION OF WHITE COLLAR WORK, Robert E. Kraut (ed.) 195-210 (1987); Jeanette Blomberg, *The Variable Impact of Computer Technologies on the Organization of Work Activities*, COMPUTER-SUPPORTED COOPERATIVE WORK: A BOOK OF READINGS, Irene Greif (ed.) 771-782 (1988); Julian Orr, TALKING ABOUT MACHINES: AN ETHNOGRAPHY OF A MODERN JOB (1996).

The case studies above are quite reactive – reactive to deaths, annoyances, invasions, and uncertainty. As wide spread or severe social problems arose from an innovation or use of technology, a solution presented itself: draw a line between man and machine. Regulate them differently - regulate or prohibit one and not the other. A human or machine was quickly perceived as the solution or problem. Ubiquitous, increasingly advanced automation calls for general principles to inform designers and users of automation to promote smooth integration into society. However, just as systems engineers realized in the 1980's, building systems around the capabilities of the automation can have negative impacts on the operator, the system, and those affected by the system. When more and more objects, materials, and services are becoming "smart," it is no longer desirable to address issues of automation as if they were industry specific or can be isolated from human activity. The same questions that plague privacy researchers investigating big data and targeted services plague ethicists and social scientists concerned about robotics, the internet of things, and algorithmic living. These debates "are ultimately about values first, and about math and machines only second,"[82] but the values get lost in reactive policy that has not been organized into general principles.

The challenge for the legal field is to somehow address the complex socio-technical nature of automation. 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.[83]

> 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.[84]

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.

IV. A SOCIO-TECHNICAL FRAMEWORK

[82] Directive 95/46/EC of the European Parliament and of the Council of 24.10.1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (O.J. L 281, 23.11.1995, p. 31 et seq.), Article 15(1).
[83] Bryant Walker Smith, *Human Factors in Robotic Torts*, WEROBOT CONFERENCE (2013), available at http://conferences.law.stanford.edu/werobot/wp-content/uploads/sites/29/2013/04/HumanFactorsRoboticTorts_BryantWalkerSmith.pdf.
[84] *Id*., at 4.

The task for law is to bring legal treatment of automation in line with responsible automation design and implementation. There is no general legal framework for all automation - it is introduced by government and private entities, in commercial and public service, 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, and 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 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. One way to establish a human or value-centered legal approach that compliments existing design methods for actual and safe use that recognize the socio-technical nature of automation is to add explicit ethical concerns into the existing automated systems design model. In doing so, the governance model becomes similar to Fair Information Practices Principles and privacy by design efforts that focus on guiding ethical technical innovation and establishing accountability, which is popular but still developing. Modeling new uses of automation for government use of automation can be similarly guided and its appropriateness assessed. Because these models are already used in automation system design and implementation in many settlings, they offer an excellent tool for both flexibility and accountability. In fact, establishing accountability within human-automation systems has been an ongoing conversation in system design and an appropriate place for policy contributions.[85]

A. MAN VS. MACHINE DESIGN

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

[85] William C. Marra and Sonia K. McNeil, *Understanding "The Loop": Regulating the New Generation of War Machines*, 36 HARVARD JOURNAL OF LAW & PUBLIC POLICY 1139, 1181 (2013) (suggesting an accountability regime may be desirable if high levels of automation are sought); Missy Cummings, *Creating Moral Buffers in Weapon Control Interface Design*, 23:3 TECHNOLOGY AND SOCIETY MAGAZINE 28 (2004) (arguing that increased distance and interaction with highly autonomous weapons systems may create moral buffers that remove the decision-maker from the loop to the point that perceived accountability is significantly diminished); Ronald C. Arkin, *Accountable Autonomous Agents: The Next Level*, DARPA TOTAL INTELLIGENCE WORKSHOP (Feb., 2009), available at http://www.cc.gatech.edu/ai/robot-lab/online-publications/ArkinAccountable.pdf ("We need to now focus on a machine's ability to better address moral and ethical capabilities in context: i.e., among human beings and subject to the presence of social norms.").

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.[86] Fitts, et. al. intended to "search for a general answer to the problem of dividing responsibility between men and machines."[87] The Fitts List is a list of 11 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.

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	

Fig. 1. The original Fitts List (MABA-MABA List), 1951.[88]

[86] J.C.F. de Winter and D. Dodou, *Why the Fitts List has Persisted Throughout the History of Function Allocation*, 16 COGNITION, TECHNOLOGY & WORK, 1 (Feb., 2014).
[87] *Id*., at 6.
[88] Paul M. Fitts, *Human Engineering for an Effective Air Navigation and Traffic Control System* (1951).

Humans Excel In	Machines Excel In
Detection of certain forms of very low 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 judgement 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
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

Fig. 2. Department of Defense adaptation, 1987.[89]

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

Fig. 3. Robert Gagne limitation-based adaptation, 1962.[90]

[89] Department of Defense, *Human engineering procedures guide*, MIL-HDBK-763 (1987).

While consistently referenced in function allocation research, the Fitts List has been heavily criticized as 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.[91] Like many efforts of initial thinking on automation, the list focuses on 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.[92] 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.[93] This groundbreaking model has played a part in the continued elite performance of U.S. pilots.[94]

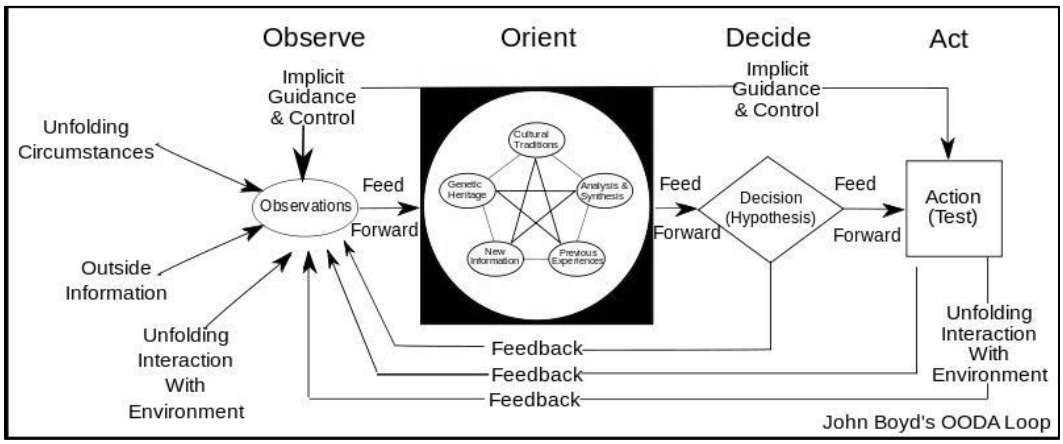

Fig. 4. John Boyd's "Observe, Orient, Decide, Act" model long relied upon for aviation systems.[95]

The four stage information processing model used by Thomas Sheridan, Parasuraman, and Christopher Wickens for automation is comparable: information acquisition, information analysis, decision selection and action implementation.[96]

Automation of information acquisition deals with input data. It may include highlighting to bring attention to potential problem information or filtering to bring certain information to the person's

[90] R. M. Gagne, *Human Functions in Systems, in Psychological Principles*, in SYSTEM DEVELOPMENT (R. M. Gagne ed., 1962).
[91] *Id*., at 1.
[92] Daniel Ford, A Vision So Noble: John Boyd, THE OODA LOOP, AND AMERICA'S WAR ON TERROR (2010); Frans P. B. Osinga, SCIENCE, STRATEGY AND WAR: THE STRATEGIC THEORY OF JOHN BOYD (2006); Robert Coram, BOYD: THE FIGHTER PILOT WHO CHANGED THE ART OF WAR (2004).
[93] *Id*.
[94] *Id*.
[95] *Id*.
[96] Parasuraman, Sheridan, & Wickens, *supra* note 7 at 290.

attention.[97] The information analysis phase involves the manipulation of retrieved and processed information in working memory.[98] 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.[99] Examples of this are 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 action may include an agent that executes certain tasks automatically in a contextually appropriate fashion[100] (i.e., photocopiers sort, collation, stapling, etc. can have different levels of automation, leaving certain tasks to the human). Of course, these are not linear; they are coordinated and overlap into "perception-action" cycles.[101]

Consider the elevator, an often cited example of the removal of the human from the loop.[102] Elevator operators were necessary when the automation was first introduced to ease public uncertainty about the innovation.[103] 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 so that the buttons make sense to humans and when an error occurs, a human can fix the problem. Many of our daily tasks have the opposite allocation. We rely on the automation of information acquisition and analysis of weather apps to decide whether to walk or take the bus. Our emails are automated to highlight, filter, and organize messages to support decisions about which 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 William Verplank are frequently credited with pioneering the concept of levels of autonomy which are condensed to the following: [104]

[97] *Id.*
[98] *Id.*
[99] *Id.*
[100] *Id.*
[101] *Id.*
[102] Singer, *supra* note 5, at 126, 131.
[103] *Id.*
[104] Thomas. B Sheridan and William L. Verplank, *Human and Computer Control of Undersea Teleoperators*, MIT Man-Machine Systems Lab Report (1978); Thomas B. Sheridan, Telerobotics, AUTOMATION AND HUMAN SUPERVISORY CONTROL (1992).

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

Fig. 5. Levels of autonomy to be applied to four-stage model.[105]

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

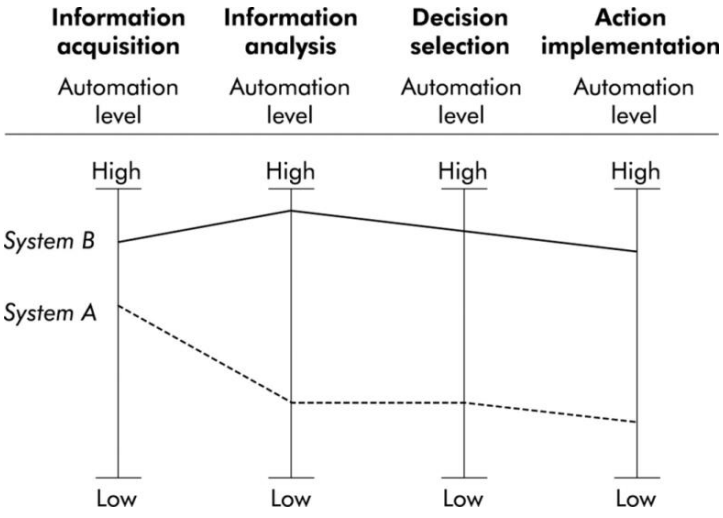

Fig. 6. Example of two systems with different levels of automation across functions.[106]

B. MAN + MACHINE DESIGN

The MABA-MABA dichotomous approach continued until work on "human-centered" design began to penetrate a number of fields in the 1980s.[107] An evaluation for automating

[105] *Id*.
[106] Parasuraman, Sheridan, & Wickens, *supra* note 7, at 289.
[107] Thomas B. Sheridan, *Human Centered Automation: Oxymoron or Common Sense?*, 1 SYSTEMS, MAN AND CYBERNETICS 823 (1995).

functions to a certain level by looking at impact on human operator performance, automation reliability,[108] and costs[109] was in place in 2000 and has been relied 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.

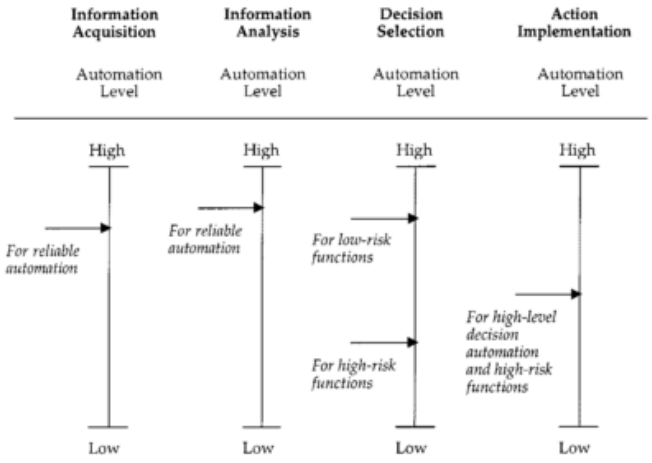

Fig. 7. Example of recommended levels for Air Traffic Control systems after evaluation of human performance consequences, automation reliability, and costs of actions.[110]

In relation to manual operations, if research shows that both human and system performance are enhanced by automation at levels of 5 but degrade above 7, then the reliability and social costs of automation should be considered within the bounds of a 5-7 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.[111]

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

[108] *Supra* Section III(A).
[109] *Infra* Section IV(C).
[110] Parasuraman, Sheridan, & Wickens, *supra* note 7, at 294.
[111] Mary L. Cummings, *Automation and Accountability in Decision Support System Interface Design*, 32:1 JOURNAL OF TECHNOLOGY STUDIES, 23, 24 (2006).

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 do to any number of the above factors. However, as these systems are integrated into social setting beyond factories, flight routes, and power plants, additional considerations must be considered. These social costs are where policy can play a more active role.

C. MAN + 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 paper includes only a small sample of the extraordinary work that has focused on making automation safe and effective, acknowledging the human, but little work has been done on integrating ethical considerations.[112] 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.[113] He toyed with many ethical possibilities but was unambiguous about his feelings on trusting machines to make critical decisions as substitute for a human:

> [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.[114]

[112] There is a large body of work on the robot ethics (*see* e.g. Patrick Lin, Keith Abney, and George Bekey (ed.), ROBOT ETHICS: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS (2011)), moral artificial intelligence (e.g. Wendell Wallach and Colin Allen, Moral Machines: Teaching Robots Right from Wrong (2010); Ronald Arkin, GOVERNING LETHAL BEHAVIOR IN AUTONOMOUS ROBOTS (2009)), and ethical impacts of automation (e.g., Cummings, *supra* note 85; Srinivasan Ramaswamy and Hermant Joshi, *Automation Ethics*, in Hand of Automation 809 (Shimon Y. Nof, 2009 (ed.)); Norbert Wiener, THE HUMAN USE OF HUMAN BEINGS: CYBERNETICS AND SOCIETY (1954)).

[113] Norbert Wiener, THE HUMAN USE OF HUMAN BEINGS: CYBERNETICS AND SOCIETY (1954).

[114] *Id*., at 184.

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! [115]

However, 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.[116]

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. All 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 the model below, designed by Parasuraman, Sheridan, and Wickens, impact of the resulting system has on the operator's performance is considered to establish initial types and levels of automation.[117] Then risk is assessed as part of the "secondary evaluation." These are separated into "Automation Reliability" and "Costs of Decision/Action Outcomes."

[115] *Id*., at 185.
[116] *Id*.
[117] Parasuraman, Sheridan, & Wickens, *supra* note 7.

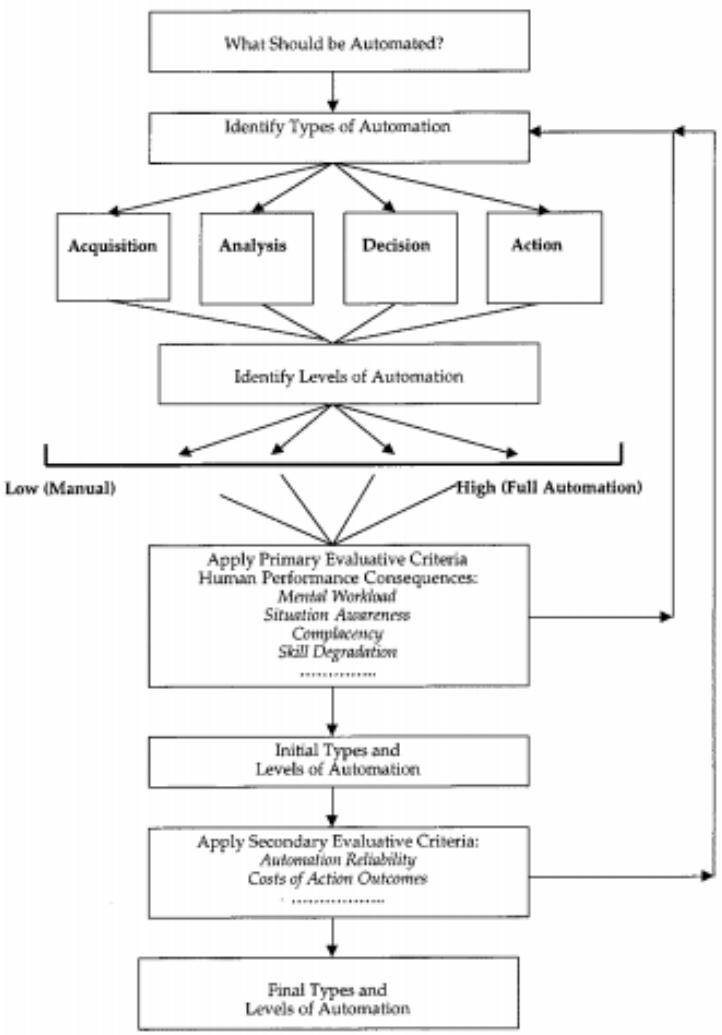

Fig. 8. Flow chart for determining appropriate levels of automation. A level (1-10 or manual to fully autonomous) for each type of automation is assigned, followed by an initial and secondary evaluation process applied to make further adjustments.

"Although it would be nice if constructed systems functioned well forever, they do not."[118] Reliability is defined as "the probability that an item will operate adequately for a specified period of time in its intended application."[119] While machine reliability and human reliability can be analyzed and combined to predict overall performance of the system,[120] automation reliability will dramatically influence the actual use of the system because of its tremendous impact on human trust.[121]

[118] Proctor & Van Zandt, *supra* note 12.
[119] Kyung S. Park, HUMAN RELIABILITY: ANALYSIS, PREDICTION, AND PREVENTION OF HUMAN ERRORS 149 (1987).
[120] Proctor and Van Zandt, *supra* note 12 at 59-79.
[121] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens, *Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs*, 2:2 JOURNAL OF

> Trust in automation is limited to the degree 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.[122]

Understanding trust is important to moving beyond rigid levels of autonomy designations to adaptive autonomy and supervisory control imitations to collaborative models,[123] 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.[124] 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 certain 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 involves 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.

COGNITIVE ENGINEERING AND DECISION MAKING 140 (2008); John D. Lee and Katrina A. See, *Trust in Automation: Designing for Appropriate Reliance*, 46 HUMAN FACTORS 50 (2004); John D. Lee and Neville Moray, *Trust, Self-Confidence, and Operator's Adaptation to Automation*, 40 INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES 153 (1994).

[122] Robert R. Hoffman, Matthew Johnson, Jeffrey M. Bradshaw, and Al Underbrink, *Trust in Automation*, 28 INTELLIGENT SYSTEMS 84, 85 (2013).

[123] *Id.*, at 84.

[124] Parasuraman, Sheridan, & Wickens, *supra* note 7, at 292.

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.[125]

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

This "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."[126] Reflective design,[127] critical technical practice,[128] participatory design,[129] value-sensitive design,[130] values in design,[131] engineering ethics[132] 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.

D. MAN + MACHINE LAW

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 be brought in line

---

[125] Parasuraman, Sheridan, & Wickens, *supra* note 7, at 292, footnote 3.
[126] Katie Shilton, *Value Levers: Building Ethics into Design*, 38:3 SCIENCE, TECHNOLOGY, & HUMAN VALUES 374 (2012).
[127] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph Kaye, *Reflective Design*, in PROCEEDINGS OF THE 4TH DECENNIAL CONFERENCE ON CRITICAL COMPUTING: BETWEEN SENSE AND SENSIBILITY (2005).
[128] Philip E. Agre, *Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI*, in SOCIAL SCIENCE, TECHNICAL SYSTEMS AND COOPERATIVE WORK: THE GREAT DIVIDE (Geoffrey C. Bowker, Les Gasser, Susan Leigh Star, and Bill Turner (eds.), 1997).
[129] Peter M. Asaro, *Transforming Society By Transforming Technology: The Science and Politics of Participatory Design*, 10 Accounting Management and Information Technology 257 (2000).
[130] Batya Friedman and Helen Nissenbaum, *Bias in Computer Systems*, in HUMAN VALUES AND THE DESIGN OF COMPUTER TECHNOLOGY (Batya Friedman (ed.), 1997)
[131] Knobel and Bowker (2011).
[132] Peter-Paul Verbeek, Materializing Morality: Design Ethics and Technological Mediation, 31:3 SCIENCE, TECHNOLOGY, & HUMAN VALUES 361 (2006).

with existing system design principles for safe and actual use of automation so as to be more meaningful and effective. 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. Locating an available position in existing automation design modeling for the law to exert input with potential methods of importing ethical conversations does not reach any of those goals. And so, I will expand on why and how utilizing existing automation models may be useful in creating effective socio-technical legal approaches, leaving unanalyzed other possibilities for recognizing man-machine interdependence in law.

1. LAW AS REGULATOR

Utilizing the models in place to uncover safe and actual automation use would naturally bring the law in line with these important factors, but the law could (and has) tried to shape automated systems more proactively using traditional prescriptive approaches. The problem with the examples from the past is that they focused on the capabilities of the automation and humans as completely separate entities with no influence or impact on the other and so legal treatment is binary and without nuance. Regulating specific levels of autonomy for different functions may provide more effective governance in some circumstances simply because it is more tailored. For instance, concerns about privacy may focus on the circumstances and duration of information acquisition (human or machine), or concerns about accuracy and accountability at the action phase may focus on reliability and intervention of the system.

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

> Specific rules often cannot reflect the large number of variable 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 broader regulatory goal in any particular circumstance. The problem is compounded when regulated entities are heterogeneous, and contexts are varied.[133]

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

Automation use is context specific and situational, and regulation can, as the above examples suggest, frustrate rather than further, broader regulatory goals. Thus, it may be a good candidate for a "more 'incomplete' regulatory instrument."[134] 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 punishing to preventing.[135] Relatively recently, this assessing and preventing risk in complex spaces has been delegated. 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 defining and monitoring it.[136]

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.[137] 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.[138] Coekelbergh suggests that this ethical responsibility may not be desired and argues that external constraints limit the moral imagination of engineers,[139] but Shilton suggests that ethical constraints, such as privacy, can be welcome additions to the design process.[140] 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. Methods for promoting ethical discourse and reflection need substantive supplementation, as well as structures for accountability and enforcement: constraints examining values related to human justice, well-being, welfare, and

[134] *Id*., at 389.
[135] *Id.*
[136] *Id*., at 390.
[137] Mark Coeckelbergh, *Regulation or Responsibility? Autonomy, Moral Imagination, and Engineering*, 31 SCIENCE, TECHNOLOGY, & HUMAN VALUES 237 (2006).
[138] *Id*.
[139] *Id*.
[140] Shilton, *supra* note 126, at 383-84.

rights in order to guide the integration of intelligent automation. These socio-technical principles can lay the foundation for socio-technical regulations and implementation.

The principles would serve a similar role as Fair Information Practices Principles – guidance for practices that specific regulations can be built upon. Fair Information Practices Principles, in some form or another, have been incorporated into nearly every information and privacy law at every level of legislation and regulation.[141] Fair Automation Practices Principles could similarly guide policy-makers and be incorporated into the design and implementation process similarly to a privacy by design approach,[142] 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."[143] 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.[144]

Because the interdependence between man and machine is complex and dynamic, the accountability approaches so en vogue in information policy may also be put in place to ensure risks and ethical considerations are being assessed internally. Depending on the potential harms and risks 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 risk to human values like safety, privacy, dignity, transparency, etc. 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 and reactive, goal-setting governance will remain industry specific, and delegation will be too lenient and overbroad.

Some problems with the existing model's secondary evaluation are that the costs of automation outcomes are at the end of the analysis, the limited scope of risk assessment, and the

[141] Colin J. Bennett and Charles D. Raab, THE GOVERNANCE OF PRIVACY: POLICY INSTRUMENTS IN GLOBAL PERSPECTIVE (2003).
[142] Ann Cavoukian, "Privacy by Design: The 7 Foundational Principles," (2011), available at http://www.ipc.on.ca/images/resources/7foundationalprinciples.pdf; Ira S. Rubinstein, *Regulating Privacy by Design*, 26 Berkeley Tech. L.J. 1409 (2011).
[143] Rubinstein, *supra* note 142, at 1411-12.
[144] 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. Douglas R. Richmond, *Human Factors Experts in Personal Injury Litigation*, 46 ARK. L. REV. 333 (1993); 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 (1984); Frank D. Fowler, *Railroad Litigation and the Human Factors Expert: Why the Plaintiff Missed the Train*, 4 AM. J. TRIAL ADVOC. 621 (1981).

lack of guidance on ethical constraints. Even still, based on this type of research, 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 human-centered aircraft automation are offered here by way of example.

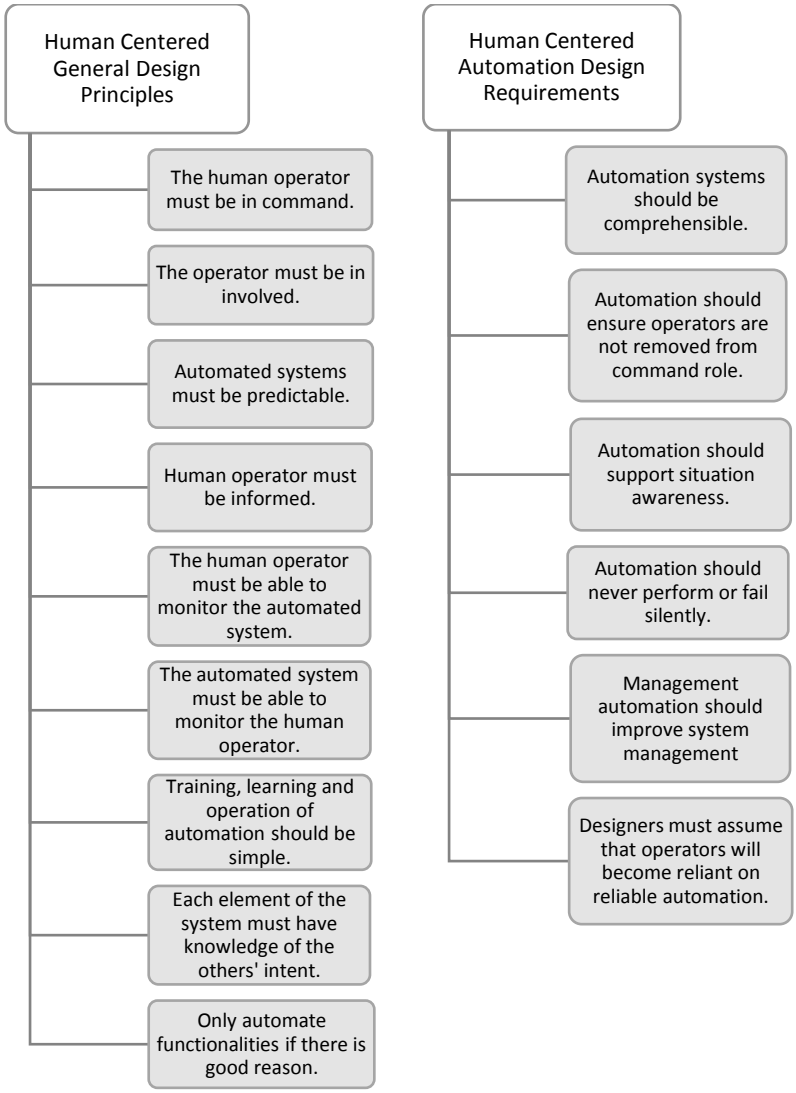

Fig. 9. Charles E. Billings' principles for aircraft automation, 1991.[145]

These principles were outlined for aircraft automation, but similar principles could be drafted for different fields or for more general purposes. The principles, focused on safe, functional, and

[145] Charles E. Billings, *Human-centered Aircraft Automation: A Concept and Guidelines*, NASA Technical Memorandum (1991).

optimized man-machine performance, will influence design, but as Dave Woods stated, "Automation that is strong, silent, and hard to direct is *not* a team player."[146] 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 process instead of fighting it. For instance, designers and implementers should consider the dignity and privacy of all humans in the loop, not just the operator. Human 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 should not be overlooked and will be more relevant with increasing automation. A discussion of what are critical decisions to be made by humans and how to limit automation bias and moral buffers in those instances would also help guide automation design. An exhaustive list of these principles is beyond the scope of this article, but research from big data and algorithmic living could certainly inform these principles. This will require designers and implementers of automation to expand a narrow view of evaluation[147] and those in legal roles to resist the urge to solve complex issues with a human or a machine or draw convenient lines between the two.

2. LAW AS IMPLEMENTER

Various legal entities have also served as implementers of automation and others have had to assess governmental implementation of automation. The examples above point to the automation of traffic violation enforcement and interpretation of automation in Fourth Amendment search decisions. In order to protect the values under their care, legal approaches should seek to establish socio-technical frameworks that do not rely on distinctions between man and machine. Often government use of emerging technologies occurs earlier than commercial or personal use, e.g., drones. When implementing an automated system to serve legal functions, evaluation of the humans in the loop should be undertaken, including individuals that will interact with the system, resolve errors, and something, as well as additional ethical constraints discussed in the previous section. Forcing a human into or out of the loop is not a substitute for evaluating the implementation of automation – public and private implementers should have guidance and structure for assessing their use of automation to protect and promote values.

[146] 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), available at http://www.ifp.illinois.edu/nsfhcs/talks/billings.html.
[147] Michael Davis, *Explaining Wrongdoing*, 20 JOURNAL OF SOCIAL PHILOSOPHY 74 (1989).

Danielle Citron highlights a number of administrative examples of automation implementation by the government that went awry causing Medicaid benefits to be terminated without notice, food stamps to be denied, and applicants for public services to be asked if they were "beggars."[148] These unfortunate events occurred due to miscoding of policies and incomplete audit trails.[149] Citron points to three central problems: (1) programmers translate policy into code in a distorted and often incorrect fashion; (2) misidentification occurs because crude algorithms are often embodied in the system; and (3) problems of notice occur because automated systems often lack audit trails that track the course an agency followed to take particular action.[150] Implementation of such a system should be scrutinized and issues of embedded bias, errors, and reliability should be resolved before automation is utilized to provide government services

Assessment of implementation should also strive for socio-technical recognition. Drawing lines between what a machine and humans can and cannot do has not been a lasting approach in design, because it has not resulted in desired performance, and in law it has become equally problematic.

The judicial case study above is housed in Fourth Amendment jurisprudence, which has a number of man-machine distinctions. Whether a machine can invade your privacy is just one question in this arena that ignores the fact that the collection and use of data by law enforcement is in a socio-technical context. Another man vs. machine debate occurs in Fourth Amendment jurisprudence when questions arise regarding law enforcement use of technology sense enhancing or sense creating. Whether search lights and binoculars,[151] beepers,[152] or thermal imaging[153] enhance or create senses for law enforcement officers have determined whether a warrant was necessary. More socio-technical approaches have recently been articulated by the *Klayman* court[154] and the Mosaic theory presented in *U.S. v. Maynard*,[155] somewhat reinforced by the Supreme Court in *U.S. v. Jones*.[156] The mosaic theory analyzes whether a search requiring

[148] Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2007).
[149] *Id*.
[150] *Id*.
[151] U.S. v. Lee, 274 U.S. 559 (1927).
[152] U.S. v. Knotts, 460 U.S. 276 (1983); U.S. v. Karo, 468 U.S. 705 (1984).
[153] Kyllo v. U.S., 533 U.S. 27 (2001).
[154] *Klayman*, Civil Action No. 13-0881.
[155] U.S. v. Maynard, 615 F. 3d 544 (D.C. Cir. 2010).
[156] U.S. v. Jones, 132 U.S. 945 (2012).

a warrant has occurred by looking at collective sequences instead of individual steps asking whether a full picture of a person's life has been created.[157]

Assessment of automation use to implement legal goals is not limited to the Fourth Amendment. Utilizing automation to remove copyrighted materials forces judges to ask whether fair use should have (and thus *could* have) been considered by a takedown system.[158] In the near future, drone use by agencies from border control[159] to search and rescue teams[160] will need to be assessed, as well as government use of intelligent systems utilizing predictive analytics[161] and e-government initiatives.[162]

Modeling levels of autonomy and humans in the loop presents a panoramic view of the larger socio-technical ecosystem that can help courts and agencies determine appropriate and responsible use of the loop. Additionally, the absence of this kind of assessment by the designer or implementer can be scrutinized by adjudicators to reveal internal risk assessment and accountability.

V. CONCLUSION

Human-automation systems researchers continue to be pressured to design within a dangerous 'automate everything' mindset, and while not necessarily driven by Wiener's social philosophy, are critical of automation and its implementation. These researchers have investigated the many ways in which placing a human (or her necessary involvement) in certain system loops impacts performance, speaking directly to concerns about safety. Through the various phases developed within the short lifespan of this research field, models have been produced to help guide system designers. The research is limited to the loosely defined objectives of the system, but should nonetheless be a starting point for informed legal approach

[157] Orin S. Kerr, *The Mosaic Theory of the Fourth Amendment*, 111 MICH. L. REV. 311 (2012).
[158] *See* Tuteur v. Crosley Corcoran, Civil Action 13-10159-RgS, 2013 WL 4832601 (D. Ma. 2013); Corynne McSherry, "Puzzling Ruling in Massachusetts Fair Use Case," Electronic Frontier Foundation (Sept. 12, 2013), https://www.eff.org/deeplinks/2013/09/puzzling-ruling-massachusetts-fair-use-case.
[159] Craig Whitlock and Craig Timberg, "Border-Patrol Drones Being Borrowed by Other Agencies More Often Than Previously Known," Washington Post (Jan. 14, 2014), available at http://www.washingtonpost.com/world/national-security/border-patrol-drones-being-borrowed-by-other-agencies-more-often-than-previously-known/2014/01/14/5f987af0-7d49-11e3-9556-4a4bf7bcbd84_story.html.
[160] Sonia Waharte and Niki Trigoni, *Supporting Search and Rescue Operations with UAVs*, in PROCEEDINGS OF THE 2010 INTERNATIONAL CONFERENCE ON EMERGENCY SECURITY TECHNOLOGIES 142 (2010).
[161] Lisa Shay, et al., *Confronting Automated Law Enforcement*, WEROBOT at 29 (2012).
[162] Hans J. Scholl, et al., *E-Government Field Force Automation: Promises, Challenges, and Stakeholders*, 4656 ELECTRONIC GOVERNMENT 127 (2007).

to the human in the loop. The law, to this point, has fallen prey to the man vs. machine way of thinking that focuses on the capabilities of the technology to draw lines and protect or promote values. The law does not have the luxury of taking decades to learn the lessons that have developed fields from system engineering to human-computer interaction, it must catch up with research and practice as it stands today and deal with the introduction of sophisticated automation as the complex interdependent socio-technical process that it is. Establishing a framework that compliments safe and actual use of automation has already become a pressing issue as the government seeks to regulate drones, big data, smart homes, and driverless cars. Initial work must be done to articulate principles to guide ethical design and lay a foundation to build policy upon. Relying on human-automation models to add layers of ethical or legal considerations can serve to guide responsible and accountable design and implementation of automation.