Artificial Intelligence Issues Affecting the Legal Profession

Christina Montgomery
Chief Privacy Officer, IBM
Artificial Intelligence and Cybersecurity Issues Affecting the Legal Profession

Table of Contents

An Overview of the Impact of Artificial Intelligence within the Legal Profession ................................................................. 1

A Primer on Using Artificial Intelligence in the Legal Profession
by Lauri Donahue (reprinted with permission from Lauri Donahue and JOLT Digest, Harvard Law Journal of Law & Technology, Copyright © 2019 Harvard Journal of Law and Technology) ............................................ 10

State v. Loomis, 881 N.W. 2d. 749, 762 (Wis. Sup. Ct. 2016) ............... 16

Remedies for Robots by Mark A. Lemley & Brian Casey (reprinted with permission from Mark A. Lemley, Stanford Law School, John M. Olin Program in Law and Economics, Copyright © 2018 Mark A. Lemley) ............... 52
Artificial Intelligence (“AI”), once only a figment of the imagination of the science fiction writer, is now a reality that permeates our lives. It has continued to expand, grow, and change at a rapid pace over a relatively short period of time, and its impact is widespread.¹ AI is the term used to describe how computers perform tasks normally viewed as requiring human intelligence, such as recognizing speech (think “Alexa” and “Siri”), identifying objects, making decisions based on data, and translating languages.² “Machine learning” is an application of AI in which computers use algorithms (rules) embodied in software to learn from data and adapt with experience. Some AI programs train themselves, while others need to be trained by humans feeding them data.³ AI has already impacted the legal profession and likely will continue to change the way in which law is practiced, how legal services are provided, and how clients are represented.⁴

It is Not So Elementary My Dear Watson

Watson, named after IBM’s first CEO, and not the dear doctor and trusty assistant of

---

³ Id. at 2.
Sherlock Holmes, was initially developed to compete on the game show “Jeopardy”.⁵ One night in February 2011, many viewers while watching the game show got the chance to see Watson in action. Three episodes made TV and science history as IBM’s Watson took on the two best players and won.⁶ Watson has many offspring now, in many sectors, with some bearing his name and some not. A scan of IBM’s Watson website will reveal that Watson has moved into social services, the health industry, data analytics, wearable technology, the banking sector, and even fantasy football.⁷

Significant to the legal profession is Watson’s lawyer son. The “son of Watson” called “Ross” was born out of a Watson University Competition, at which a group of University of Toronto students built a legal application on top of the Watson platform. The students placed second in the IBM contest.⁸ Ross is branded as an artificially intelligent attorney that helps human lawyers research faster. When asked a question in plain English, Ross returns an answer with readings from statutes, cases, and other sources.⁹

Ross’s handlers claim that Ross can think like a lawyer.¹⁰ Furthermore, by taking advantage of the natural language and cognitive computing platform that Watson offers, Ross can predict the outcome of court cases with a confidence rating, assess legal precedents and suggest readings to prepare for cases.¹¹ Interestingly, as of August 2015, Ross was learning everything there is to know

---


⁷ See id.

⁸ See, e.g., id. at 37.

⁹ See, e.g., Edward J. Walters, Jr., Ipse Dixit: Big Data, 64 LA. B.J. 248 (2016).

¹⁰ See id. at 248.

about U.S. bankruptcy law. Ross’s powerful software is currently changing the way bankruptcy and restructuring lawyers research the law. Its natural language search capabilities enable sophisticated queries that go beyond ordinary key word searches. Not surprisingly, Ross is being utilized in several elite law firms.

Besides Ross, there are numerous technologies in the subfield of AI and the law to consider in a discussion on the effect of AI and the legal profession. A few to mention here (although there are more, and the numbers are growing) are IBM’s Watson Debater, ModusP, Lex Machina, Modria, Premonition, BEAGLE, and COMPAS. Watson Debater is a new feature of IBM’s Watson computer. When asked to discuss any topic, it can autonomously scan its knowledge database for relevant content, comprehend the data, select what it believes are the strongest arguments, and then construct sentences in natural language to illustrate the points it had selected, in favor and against the topic.

The Israeli startup company, ModusP, has created an advanced search engine using sophisticated algorithms based on AI. The search function helps jurists reduce legal research hours by finding legal knowledge and insights more efficiently. Another AI tool, Lex Machina, acquired by LexisNexis in 2015, transforms data from federal court dockets into live charts. In Lexis Advance, one can access this data by clicking on the hyperlinked judge’s name in the text of a case that shows a summary with the judge’s biographical information, open cases by practice

---

12 Nelson & Simek, Watson’s Children, supra note 6, at 38.
14 Nelson & Simek, Watson’s Children, supra note 6, at 38.
16 Id. at 55.
17 Id. at 55.
area, comparisons to other judges in the district, cases filed by year and case timelines.\textsuperscript{19} The latter charts can give one a sense of how long the dispute may take to resolve, and the odds of winning at trial.\textsuperscript{20} Lex Machina’s Motion Chains enables one to analyze the odds of success for a specific type of motion based on historical data. Similarly, an attorney can see how opposing counsel has performed on similar cases or before a particular judge.\textsuperscript{21} Further, one can also research parties, case damages, venues, practice areas (and so on) for business development purposes.\textsuperscript{22}

Another company, Modria, seeks to increase efficiency still further, through automating the dispute resolution process through the use of algorithms, effectively removing humans from the justice delivery system.\textsuperscript{23} There is also Premonition, which is a legal analytics software that can rank arbitrators based on past decisions, help select expert witnesses based on their persuasiveness and past case results, and can help analyze the court, judge and opposing counsel based on their win rate and results. It can also help with tailoring client pitch data, finding local counsel and recruiting new litigators.\textsuperscript{24}

BEAGLE utilizes AI to quickly highlight the most important clauses in a contract.\textsuperscript{25} It also provides a real-time collaboration platform that enables lawyers to easily negotiate a contract or pass it around an organization for quick feedback. BEAGLE’s learning process allows the program to adapt to focus on what users care about most.\textsuperscript{26} COIN, which is short for Contract Intelligence, is another contract review tool, used since June 2017, by JPMorgan, is an AI powered program

\begin{footnotes}
\footnotesize
\begin{itemize}
\item[19] See id. at 24.
\item[20] See id. at 24.
\item[21] See id. at 24.
\item[22] See id. at 24.
\item[25] See, e.g., Ben-Ari et. al., supra note 15, at 55.
\item[26] Id. at 55.
\end{itemize}
\end{footnotes}
used to interpret commercial loan agreements. The bank plans to use the technology to interpret other legal documents as well.27

Finally, another AI tool developed by a private company called Equivant (formerly Northpointe), Its name COMPAS—or the Correctional Offender Management Profiling for Alternative Sanctions—purports to predict a defendant’s risk of committing another crime.28 It works through a proprietary algorithm that considers some of the answers to a 137-item questionnaire.29 The American Bar Association has urged states to adopt risk assessment tools in an effort to reduce recidivism and increase public safety. Some states that use COMPAS have conducted validation studies of COMPAS concluding that it is a sufficiently accurate risk assessment tool.30 Although not directly used by law firms, it can significantly impact the way counsel represents a criminal defendant.31

**Changing How Legal Services are Provided**

Although AI can’t yet replace advocacy, negotiation, or structuring of complex deals, it can do things like review documents during litigation and due diligence, analyze contracts to determine whether they meet pre-determined criteria, perform legal research, and predict case outcomes.32 The development of the field of AI and law starts with programs that analyze cases and continues with technologies that make lawyers’ tasks efficient, solve disputes, and replace human intervention.33 This is not to say that the legal profession is going away. Many argue that these AI technological advancements will allow and provide more opportunities for those willing to work

---

27 Donahue, *supra* note 2.
29 Id.
32 Donahue, *supra* note 2.
33 Ben-Ari et. al., *supra* note 15, at 56.
with the AI. To take advantage of these opportunities, it is vital that attorneys know what the emerging technologies are and how to work with them.

There are numerous examples to illustrate how AI can change the provision of legal services, but only a few are discussed here. One example, and probably AI’s most used aspect, is legal research. Any lawyer who has performed legal research using Lexis or Westlaw has used legal automation, but AI takes research to the next level. Ross, for an example, uses the power of Watson to find similar cases, even responding to queries in plain English. Andrew Arruda, CEO and co-founder of Ross, said this is just one of many success stories. Ross’s breakthroughs in legal research software using “natural language processing” means that “legal research now goes from a task that takes hours down to a task that might take half that time to find the same results.” Arruda added that lawyers are also “finding cases they literally never would have found anywhere else.” Because Ross is a cognitive computing platform, it learns from past interactions, that is, Ross’s responses increase in accuracy as lawyers continue to use it. This feature can help lawyers reduce the time spent on research. This will affect how much time a lawyer has in preparing a case which will better serve the client.

Another example is contract review. BEAGLE purports to review legal documents and reduce manual reviewing error rates, reducing the average time for legal review to less than 20 minutes, and increasing legal review accuracy by twenty percent (20%) (BEAGLE’s website

---

35 See, e.g., Id. at 6.
36 See, e.g., Donahue, supra note 2.
37 Id.
39 Id.
40 See, e.g. Ben-Ari et. al., supra note 15, at 55.
The contract review applications such as BEAGLE and COIN are not only being used by law firms, they are being used by companies to save money by not having to pay attorneys for contract review. In-house departments are increasingly pressured to become more efficient and measure performance. These pressures ensure that every in-house leader will need to consider AI at some point. Thomas Trujillo, who up until approximately 2017, served as a chief operating officer in Bank of America’s legal department, explains the typical plight: “There is more and more pressure on companies and law departments to evolve and reduce costs. We must balance competing demands and a limited budget.” Fortunately, a wealth of AI technology is already here, and is rapidly changing and automating the way law is practiced, from research and contract review to other areas of services. AI will put pressure on lawyer billing, as clients will expect less time to be spent on research and contract review. That is not to say that these items will not be a part of a lawyer’s billable practice, but clients will be seeking attorneys who use AI in a cost-effective manner.

Adapting Client Representation and the Practice of Law When AI is Involved

AI is affecting client representation both in- and outside of the firm and the courtroom. For an example, Watson Debater can assist lawyers by suggesting the most persuasive arguments and precedents when dealing with a legal matter, enabling them to represent clients more effectively. Other AI technologies such as Lex Machina that give lawyers more information on specific judges, a client’s history, and information on what they can do to have a better chance of winning could

---

43 Id.
44 Id.
45 See, e.g., Connell, supra note 41, at 7.
46 See, e.g., Ben-Ari et. al., supra note 15, at 55.
possibly impact not only representation in the courtroom, but also in whether to enter into settlement negotiations. The same can be said with Premonition and its suggestions on which lawyers win the most cases before which judges. This might impact leaning towards attempting settlement instead of litigation or litigating instead of accepting a settlement.47

Perhaps, a telling example of how AI can affect client representation is the use of COMPAS by the court. Initially, risk assessment tools such as COMPAS were only used by probation and parole departments to help determine the best supervision and treatment strategies for offenders.48 However, in State v. Loomis, the court held that if used properly, observing the limitations and cautions set forth in its opinion, a court’s consideration of a COMPAS risk assessment at sentencing does not violate a defendant’s right to due process.49 In Loomis, the defendant denied involvement in the crime but waived his right to trial by agreeing to a plea deal. The plea deal left the actual sentence to the discretion of the Wisconsin circuit court judge.50

The judge accepted the guilty plea from the defendant and ordered a risk assessment. The assessment predicted the defendant had a high risk of recidivism.51 “Instead of one year in county jail with probation, which the prosecution and defense had agreed upon, the circuit court sentenced the defendant to ‘seven years with four years initial confinement’ for operating a motor vehicle without the owner’s consent.”52 In addition, the court sentenced the defendant to four years with two years suspended for attempting to flee an officer. Although the judge in Loomis also relied on other factors in addition to COMPAS, the COMPAS results affected the way the judge sentenced...
the defendant.\textsuperscript{53} The defendant in \textit{Loomis} argued that it was impossible to challenge a risk assessment without sufficient information about how COMPAS functions, such as how risk is determined and how factors are weighed to calculate the assessment, but because the defendant could correct responses to questions, the court determined that he had the ability to determine the accuracy of his risk assessment.\textsuperscript{54} The takeaway is that attorneys need to be aware of and learn how to deal with these novel tools in court in order to reinforce the rule of law and protect their client’s rights.\textsuperscript{55}

\textbf{Conclusion}

Artificial Intelligence and its growth need not be feared. AI is not beating humans when it comes to many legal skills and tasks.\textsuperscript{56} Whether for data analytics or for streamlined legal research, the use of AI can keep client costs down. Some even theorize that the failure to use AI could be considered malpractice one day.\textsuperscript{57} AI, however, does not have judgment, creativity and most importantly empathy.\textsuperscript{58} It does not equate with emotional intelligence.\textsuperscript{59} AI is a tool and those in the legal profession, including lawyers and judges need to know how to utilize it.\textsuperscript{60} Lawyers and judges are only as good as the information they receive, and AI has the potential to significantly improve the quality of that information.\textsuperscript{61}

\textsuperscript{53} \textit{Id.} at 139.
\textsuperscript{54} \textit{Id.} at 139.
\textsuperscript{56} \textit{See, e.g.}, Connell, \textit{supra} note 41 at 43–44.
\textsuperscript{57} Browning and Downs, \textit{supra} note 38, at 509.
\textsuperscript{58} \textit{See, e.g.}, Connell, \textit{supra} note 41, at 44
\textsuperscript{59} \textit{See, e.g.}, \textit{id.} at 44 (quoting Microsoft Assistant General Counsel Dennis Garcia).
\textsuperscript{60} \textit{See, e.g.}, Connell, \textit{supra} note 41, at 44.
\textsuperscript{61} Browning & Downs, \textit{supra} note 38, at 509.
A Primer on Using Artificial Intelligence in the Legal Profession

By Lauri Donahue
January 03, 2018

Lauri Donahue is a 1986 graduate of Harvard Law School and was one of the co-founders of the Harvard Journal of Law & Technology. She is now the Director of Legal Content for LawGeek, a Tel Aviv legaltech startup.

What's artificial intelligence ("AI") and why should lawyers care about it? On a practical level, lawyers should be aware that software powered by AI already carries out legal tasks. Within a few years, AI will be taking over (or at least affecting) a significant amount of work now done by lawyers. Thirty-nine percent of in-house counsel expect that AI will be commonplace in legal work within ten years.

On a more philosophical level, lawyers should understand that the "decisions" made by AI-powered software will raise significant legal questions, including those of tort liability and of criminal guilt. For example, if AI is controlling a driverless car and someone's killed in an accident, who's at fault?

While the philosophical questions are important to resolve, this Comment will focus on the practical issues. To provide an overview of what AI is and how it will be used in the legal profession, this Comment addresses several questions:

- What is AI?
- How does AI work?
- What can AI do?
- How are lawyers using AI?
- How will AI affect the legal profession?

What is AI?

Let's start with a few definitions:

"Artificial Intelligence" is the term used to describe how computers can perform tasks normally viewed as requiring human intelligence, such as recognizing speech and objects, making decisions based on data, and translating languages. AI mimics certain operations of the human mind.

https://jolt.harvard.edu/digest/a-primer-on-using-artificial-intelligence-in-the-legal-profession
"Machine learning" is an application of AI in which computers use algorithms (rules) embodied in software to learn from data and adapt with experience.

A "neural network" is a computer that classifies information -- putting things into "buckets" based on their characteristics. The hot-dog identifying app from HBO's Silicon Valley is an example of one application of this technology.

**How Does AI Work?**

Some AI programs train themselves, through trial and error. For example, using a technique called neuroevolution, researchers at Elon Musk's OpenAI research center set up an algorithm with policies for getting high scores on Atari videogames. Several hundred copies of these rules were created on different computers, with random variations. The computers then "played" the games to learn which policies were most effective and fed those results back into the system. AI can also be used to build better AI. Google is building algorithms that analyze other algorithms, to learn which methods are more successful.

Other AI programs need to be trained by humans feeding them data. The AI then derives patterns and rules from that data. AI programs trained through machine learning are well-suited to solve classification problems. This basically means calculating the probability that certain information is either of type A or type B. For example, determining whether a given bear is a panda or a koala is a classification problem.
The training starts with showing the computer lots of samples of pandas and koalas. These initial samples are called the training set, and clearly identify which type of bear is being presented to the AI.

The AI builds a model—a set of rules—to distinguish between pandas and koalas. That model might be based on things like size, coloring, the shape of the ears, and what the animal eats (bamboo or eucalyptus).

After training, the AI can be tested with new pandas and koalas to see whether it classifies them correctly. If it doesn't do very well, the algorithm may need to be tweaked or the training set may need to be expanded to give the AI more data to crunch.

**What Can AI Do?**

At this point in its development, AI is good at finding items that meet human-defined criteria and detecting patterns in data. In other words, AI can figure out what makes a panda a panda and what...
distinguishes it from a koala—which lets it find the pandas in a collection of random bears. These are sometimes called "search-and-find type" tasks.

Once it's identified something, the AI can then apply human-defined rules and take actions. In the case of legal work, an AI can carry out tasks like:

- IF this document is a non-disclosure agreement, THEN send it to the legal department for review
- IF this NDA meets the following criteria, THEN approve it for signature
- FIND all my contracts with automatic renewal clauses and NOTIFY ME four weeks before they renew
- TELL ME which patents in this portfolio will expire in the next six months

According to Stefanie Yuen Thio, joint managing partner and head of corporate at TSMP Law Corp. in Singapore, legal work that's repetitive, requiring minimal professional intervention, or based on a template will become the sole province of software. In addition, she says,

> any legal work that depends on collating and analyzing historical data such as past judicial decisions, including legal opinions or evaluating likely litigation outcomes, will become the dominion of AI. No human lawyer stands a chance against the formidable processing power of a mainframe when it comes to sifting through voluminous data.

AI can help consumers by providing a form of "legal service" to clients who might otherwise not be able to afford a lawyer. The free service DoNotPay, created by a 19-year-old, is an AI-powered chatbot that lets users contest parking tickets in London and New York. In its first 21 months, it took on 250,000 cases and won 160,000 of them, saving users more than $4 million worth of fines. The same program is helping consumers file databreach-related suits against Equifax for up to $25,000—though it can’t help them litigate their cases.

**What AI Can't Do**

According to Yuen Thio, AI can't yet replicate advocacy, negotiation, or structuring of complex deals. The New York Times suggested that tasks like advising clients, writing briefs, negotiating deals, and appearing in court were beyond the reach of computerization, at least for a while. AI also isn't yet very good at the type of creative writing in a Supreme Court brief. Or a movie script.

**How Are Lawyers Using AI?**

Lawyers are already using AI to do things like reviewing documents during litigation and due diligence, analyzing contracts to determine whether they meet pre-determined criteria, performing legal research, and predicting case outcomes.

**Document Review**

Document review for litigation involves the task of looking for relevant documents—for example, documents containing specific keywords, or emails from Ms. X to Mr. Y concerning topic Z during March, 2016. Setting search parameters for document review doesn't require AI, but using AI improves the speed, accuracy, and efficiency of document analysis.

For example, when lawyers using AI-powered software for document review flag certain documents as relevant, the AI learns what type of documents it's supposed to be looking for. Hence, it can more accurately identify other relevant documents. This is called "predictive coding." Predictive coding offers many advantages over old-school manual document review. Among other things, it:

• levers small samples to find similar documents
• reduces the volume of irrelevant documents attorneys must wade through
• produces results that can be validated statistically
• is at least modestly more accurate than human review
• is much faster than human review

Predictive coding has been widely accepted as a document review method by US courts since the 2012 decision in Da Silva Moore v. Publicus Groupe.

Analyzing Contracts

Clients need to analyze contracts both in bulk and on an individual basis.

For example, analysis of all contracts a company has signed can identify risks, anomalies, future financial obligations, renewal and expiration dates, etc. For companies with hundreds or thousands of contracts, this can be a slow, expensive, labor-intensive, and error-prone process (assuming the contracts aren’t already entered into a robust contract management system). It’s also boring for the lawyers (or others) tasked with doing it.

On a day-to-day basis, lawyers review contracts, make comments and redlines, and advise clients on whether to sign contracts as-is or try to negotiate better terms. These contracts can range from simple (e.g., NDAs) to complex. A backlog of contracts to review can create a bottleneck that delays deals (and the associated revenues). Lawyers (especially inexperienced ones) can miss important issues that can come back to bite their clients later.

AI can help with both bulk and individual contract review.

At JPMorgan, an AI-powered program called COIN has been used since June 2017 to interpret commercial loan agreements. Work that previously took 360,000 lawyer-hours can now be done in seconds. The bank is planning to use the technology for other types of legal documents as well.

Some AI platforms, such as the one provided by Kira Systems, allow lawyers to identify, extract, and analyze business information contained in large volumes of contract data. This is used to create contract summary charts for M&A due diligence.

The company I work for, LawGeex, uses AI to analyze contracts one at a time, as part of a lawyer’s daily workflow. To start with, lawyers set up their LawGeex playbooks by selecting from a list of clauses and variations to require, accept, or reject. For example, a California governing law clause might be OK, but Genovian law isn’t. Then, when someone uploads a contract, the AI scans it and determines what clauses and variations are present and missing. The relevant language is highlighted and marked with a green thumbs-up or a red thumbs-down based on the client’s preset criteria.

In-house lawyers use LawGeex to triage standard agreements like NDAs. Contracts meeting predefined criteria can be pre-approved for signature; those that don’t are kicked to the legal department for further review and revision.

Legal Research

Any lawyer who’s ever done research using Lexis or Westlaw has used legal automation. Finding relevant cases in previous eras involving the laborious process of looking up headnote numbers and Shepardizing in paper volumes. But AI takes research to the next level. For example, Ross Intelligence uses the power of IBM’s Watson supercomputer to find similar cases. It can even respond to queries in plain English. The power of AI-enabled research is striking: using common research methods, a
bankruptcy lawyer found a case nearly identical to the one he was working on in 10 hours. Ross's AI found it almost instantly.

Predicting Results

Lawyers are often called upon to predict the future: If I bring this case, how likely is it that I'll win – and how much will it cost me? Should I settle this case (or take a plea), or take my chances at trial? More experienced lawyers are often better at making accurate predictions, because they have more years of data to work with.

However, no lawyer has complete knowledge of all the relevant data.

Because AI can access more of the relevant data, it can be better than lawyers at predicting the outcomes of legal disputes and proceedings, and thus helping clients make decisions. For example, a London law firm used data on the outcomes of 600 cases over 12 months to create a model for the viability of personal injury cases. Indeed, trained on 200 years of Supreme Court records, an AI is already better than many human experts at predicting SCOTUS decisions.

How Will AI Affect the Legal Profession?

A consensus has emerged that AI will significantly disrupt the legal market. AI will impact the availability of legal sector jobs, the business models of many law firms, and how in-house counsel leverage technology.

According to Deloitte, about 100,000 legal sector jobs are likely to be automated in the next twenty years. Deloitte claims 39% of legal jobs can be automated; McKinsey Global Institute estimates that 23% of a lawyer’s job could be automated. Some estimates suggest that adopting all legal technology (including AI) already available now would reduce lawyers’ hours by 13%.

How Law Firms are Responding to AI

Law firms are notoriously slow to adapt to new technologies. Enhancing efficiency is often seen as contrary to the economic goal of maximizing billable hours. Lawyers are also seen as being technophobic.

However, many law firms are trying to understand and use new legal technologies, including AI. According to the London Times, "the vast majority of the UK’s top 100 law firms are either using artificial intelligence or assessing the technology." Firms adopting AI systems include Latham & Watkins, Baker & McKenzie, Slaughter & May, and Singapore’s Dentons Rodyk & Davidson.

Ron Dolin, a senior research fellow at Harvard Law School’s Center on the Legal Profession, says that traditional law firm business models based on armies of first year associates racking up billable hours doing M&A contract review are doomed by the advent of AI. This isn’t necessarily bad news for junior associates—or at least for the ones who still have jobs—as many hated doing contract review in the first place.

Firms that fail to take advantage of AI-powered efficiencies may lag in competing with those who do—at least to the extent clients insist on fixed-rate billing. Thus, lawyers who understand technology, and educate themselves about the latest legaltech developments, may be of increasing value to their firms.

How In-House Counsel Are Using AI
Corporate counsel have obvious reasons to adopt AI. Unlike attorneys in law firms, corporate counsel have no incentive to maximize their hours. Indeed, many lawyers go in-house to improve their work-life balance, which includes getting home at a reasonable hour. They’re also often subject to strict budget and headcount constraints, so they have to figure out how to get more done with limited resources. AI helps in-house lawyers get home earlier without increasing their departmental budgets.

**AI and the Future of the Legal Profession**

The [ABA Model Rules of Professional Conduct](https://www.abanet.org/aboutaba/governance/modelrules.html) (“Model Rules”) require that lawyers be competent—and that they keep up with new technology. As Comment 8 states:

> To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology...

At least 27 states have adopted some form of this Model Rule. In January of 2017, [Florida](https://www.flsenate.gov/Session/Legislation/2017/Chapters/2017Ch8045CS) became the first state to require technology training as part of its continuing legal education requirement. Other states seem likely to follow suit. Indeed, failing to use commonly available technology, like email and e-discovery software, can be grounds for a [malpractice claim or suspension by the bar](https://www.flsenate.gov/Session/Legislation/2017/Chapters/2017Ch8045CS).

Of course, AI-powered legal automation is not yet common. But it soon will be. Spending on AI is expected to grow rapidly—from $8 billion in 2016 to $47 billion in 2020—as AI is seen as reducing costs and increasing efficiency. [Top MBA programs](https://www.flsenate.gov/Session/Legislation/2017/Chapters/2017Ch8045CS) already have courses on how managers can use AI applications.

As they come to rely on AI, C-level executives may expect that their inside and outside lawyers are also up-to-speed.
The lead opinion concludes that Lands’ End’s right to the 12 percent interest rate was “contingent on a subsequent determination by a court.” Lead op., ¶ 77. But the lead opinion forces Lands’ End to bear the burden of the right court making the wrong determination at a critical time. Had the same court decided the case six weeks later, the result would have been different.

¶ 229 For the foregoing reasons, I respectfully dissent.

¶ 230 I am authorized to state that Chief Justice PATIENCE DRAKE ROGGEN-SACK joins this dissent.

2016 WI 68
STATE of Wisconsin, Plaintiff–Respondent,
v.
Eric L. LOOMIS, Defendant–Appellant.
No. 2015AP157–CR.
Supreme Court of Wisconsin.
Argued April 5, 2016.
Decided July 13, 2016.
Background: Defendant was convicted pursuant to a guilty plea of attempting to flee a traffic officer and operating a motor vehicle without the owner’s consent and was sentenced to four years, with initial confinement of two years and extended supervision of two years on the attempting to flee charge and seven years, with four years of initial confinement and three years of extended supervision, to be served consecutively with the prior sentence, on the operating without consent charge. Defendant petition for post-conviction relief. The Circuit Court, La Crosse County, Scott L. Horne, J., denied the petition. Defendant appealed. The Court of Appeals, 2015 WL 5446731, certified the appeal to the Supreme Court.

Holdings: The Supreme Court, Ann Walsh Bradley, J., held that:

(1) use of risk assessment tool at sentencing did not violate defendant’s due process right to be sentenced based on accurate information;
(2) use of risk assessment tool at sentencing did not violate defendant’s due process right to an individualized sentence;
(3) risk assessment tool’s consideration of a defendant’s gender did not violate defendant’s due process rights; and
(4) sentencing court correctly considered read-in charges.

City’s issue preclusion argument is that the City miscasts the “issue” to which issue preclusion applies. The “issue” is not the proper 2008 assessed value of Lands’ End’s property. Rather, we determine here that issue preclusion applied only to the “issue” of the correct 2006 assessment. The resolution of that issue through the application of issue preclusion does not, by itself, establish the proper 2008 assessed value. Rather, it is the combination of issue preclusion and a new undisputed fact in the present case that persuades us that Lands’ End is entitled to summary judgment. The new undisputed fact is that the value of the subject property did not materially change between 2006 and 2008.

Giving preclusive effect to Judge Leineweber’s finding that the 2006 value of the property was $25,000,000, and combining that finding with the undisputed fact in this case that the value of the property essentially stayed the same, leads us to conclude that the value of the property in 2008 must be $25,000,000. Because there is no genuine dispute that the 2008 value of the property is $25,000,000, we conclude that Lands’ End is entitled to judgment as a matter of law.

1. **Criminal Law ⇨1134.29, 1179**

   Whether a defendant’s right to due process has been violated presents a question of law, which the Supreme Court reviews independently of the determinations of a circuit court or a court of appeals. U.S.C.A. Const.Amend. 14.

2. **Criminal Law ⇨1156.2**

   The Supreme Court reviews sentencing decisions under the erroneous exercise of discretion standard.

3. **Criminal Law ⇨1156.2**

   An erroneous exercise of discretion occurs when a circuit court imposes a sentence without the underpinnings of an explained judicial reasoning process.

4. **Criminal Law ⇨1156.2**

   A sentencing court erroneously exercises its discretion when its sentencing decision is not based on the facts in the record or it misapplies the applicable law.

5. **Sentencing and Punishment ⇨59**

   A sentencing court misapplies the law when it relies on clearly irrelevant or improper factors.

6. **Criminal Law ⇨1141(2)**

   The defendant bears the burden of proving that a sentencing court relied on clearly irrelevant or improper factors by clear and convincing evidence.

7. **Criminal Law ⇨1134.75**

   A discretionary sentencing decision will be sustained if it is based upon the facts in the record and relies on the appropriate and applicable law.

8. **Constitutional Law ⇨3875**

   The process that is due under the federal constitution differs with the types of decisions and proceedings involved. U.S.C.A. Const.Amend. 14.

9. **Constitutional Law ⇨3875**

   Due process is flexible and calls for such procedural protections as the particular situation demands. U.S.C.A. Const. Amend. 14.

10. **Constitutional Law ⇨4705**

    **Sentencing and Punishment ⇨283**

    Use of “Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)” risk assessment tool at sentencing did not violate defendant’s due process right to be sentenced based on accurate information; although defendant could not review and challenge how the proprietary algorithm calculated risk, he could at least review and challenge the resulting risk scores set forth in the report attached to the presentence investigation report (PSI), risk assessment was based upon defendant’s answers to questions and publicly available data about his criminal history, court and defendant had access to same copy of risk assessment, and studies of tool have determined that it was reasonably accurate. U.S.C.A. Const.Amend. 14.

11. **Constitutional Law ⇨4705**


12. **Constitutional Law ⇨4705**

    The due process right to be sentenced based upon accurate information includes the right to review and verify information contained in the presentence investigation report (PSI) upon which the circuit court bases its sentencing decision. U.S.C.A. Const.Amend. 14.
13. Constitutional Law $\supseteq$4705

**Sentencing and Punishment $\supseteq$283**

Use of “Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)” risk assessment tool at sentencing did not violate defendant’s due process right to an individualized sentence; although assessment was based partially on group data, consideration of a risk assessment at sentencing along with other supporting factors was helpful in providing the sentencing court with as much information as possible in order to arrive at an individualized sentence. U.S.C.A. Const. Amend. 14.

14. Constitutional Law $\supseteq$4705

**Sentencing and Punishment $\supseteq$283**

The “Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)” risk assessment tool’s consideration of a defendant’s gender did not violate defendant’s due process rights at sentencing, where there was a factual basis underlying the tool’s use of gender in calculating risk scores, and defendant failed to establish that sentencing court actually considered gender in imposing his sentence. U.S.C.A. Const. Amend. 14.

15. Sentencing and Punishment $\supseteq$1900

Although it cannot be determinative, a sentencing court may use a “Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)” risk assessment as a relevant factor for such matters as: (1) diverting low-risk prison-bound offenders to a non-prison alternative; (2) assessing whether an offender can be supervised safely and effectively in the community; and (3) imposing terms and conditions of probation, supervision, and responses to violations.

16. Sentencing and Punishment $\supseteq$283

A “Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)” risk assessment may be used to enhance a judge’s evaluation, weighing, and application of the other sentencing evidence in the formulation of an individualized sentencing program appropriate for each defendant.

17. Sentencing and Punishment $\supseteq$1900

“Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)” risk scores may not be used at sentencing as the determinative factor in deciding whether an offender can be supervised safely and effectively in the community.

18. Sentencing and Punishment $\supseteq$283

“Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)” risk scores may not be used to determine whether an offender is incarcerated or to determine the severity of the sentence.

19. Sentencing and Punishment $\supseteq$117, 373

A circuit court must explain the factors in addition to a “Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)” risk assessment that independently support the sentence imposed, and a COMPAS risk assessment is only one of many factors that may be considered and weighed at sentencing.

20. Sentencing and Punishment $\supseteq$283

Any presentence investigation report (PSI) containing a “Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)” risk assessment filed with the court must contain a written advisement listing the limitations.

21. Sentencing and Punishment $\supseteq$98

Sentencing court was permitted to consider read-in charges in sentencing defendant following guilty plea premised on plea agreement; although court initially
misstated that there was no distinction between read-in charges and dismissed charges, it ultimately corrected itself and proceeded under the correct framework.

22. Sentencing and Punishment ☞98

Read-in charges are expected to be considered at sentencing with the understanding that the read-in charges could increase the sentence up to the maximum that the defendant could receive for the conviction in exchange for the promise not to prosecute those additional offenses.

For the defendant-appellant, there were briefs by Michael D. Rosenberg and Community Justice, Inc., Madison, and oral argument by Michael D. Rosenberg.

For the plaintiff-respondent, the cause was argued by Christine A. Remington, assistant attorney general, with whom on the brief was Brad D. Schimel, attorney general.

ON CERTIFICATION FROM THE COURT OF APPEALS

ANN WALSH BRADLEY, J.

¶ 1 In 2007, the Conference of Chief Justices adopted a resolution entitled “In Support of Sentencing Practices that Promote Public Safety and Reduce Recidivism.” It emphasized that the judiciary “has a vital role to play in ensuring that criminal justice systems work effectively and efficiently to protect the public by reducing recidivism and holding offenders accountable.” The conference committed to “support state efforts to adopt sentencing and corrections policies and programs based on the best research evidence of practices shown to be effective in reducing recidivism.”

¶ 2 Likewise, the American Bar Association has urged states to adopt risk assessment tools in an effort to reduce recidivism and increase public safety. It emphasized concerns relating to the incarceration of low-risk individuals, cautioning that the placement of low-risk offenders with medium and high-risk offenders may increase rather than decrease the risk of recidivism. Such exposure can lead to negative influences from higher risk offenders and actually be detrimental to the individual’s efforts at rehabilitation.

¶ 3 Initially risk assessment tools were used only by probation and parole departments to help determine the best supervision and treatment strategies for offenders. With nationwide focus on the need to reduce recidivism and the importance of evidence-based practices, the use of such


2. Id.

3. Id.


5. Id. at 19.

6. Id.

tools has now expanded to sentencing.8 Yet, the use of these tools at sentencing is more complex because the sentencing decision has multiple purposes, only some of which are related to recidivism reduction.9

¶ 4 When analyzing the use of evidence-based risk assessment tools at sentencing, it is important to consider that tools such as COMPAS continue to change and evolve.10 The concerns we address today may very well be alleviated in the future. It is incumbent upon the criminal justice system to recognize that in the coming months and years, additional research data will become available. Different and better tools may be developed. As data changes, our use of evidence-based tools will have to change as well. The justice system must keep up with the research and continuously assess the use of these tools.

¶ 5 Use of a particular evidence-based risk assessment tool at sentencing is the heart of the issue we address today. This case is before the court on certification from the court of appeals.11 Petitioner, Eric L. Loomis, appeals the circuit court’s denial of his post-conviction motion requesting a resentencing hearing.

¶ 6 The court of appeals certified the specific question of whether the use of a COMPAS risk assessment at sentencing violates a defendant’s right to due process. Additionally, he contends that the circuit court erroneously exercised its discretion by assuming that the factual bases for the read-in charges were true.

¶ 8 Ultimately, we conclude that if used properly, observing the limitations and cautions set forth herein, a circuit court’s consideration of a COMPAS risk assessment at sentencing does not violate a defendant’s right to due process.

¶ 9 We determine that because the circuit court explained that its consideration of the COMPAS risk scores was supported by other independent factors, its use was not determinative in deciding whether Loomis could be supervised safely and effectively in the community. Therefore, the circuit court did not erroneously exercise its discretion. We further conclude that the circuit court’s consideration of the read-in charges was not an erroneous ex-


12. We are also asked to review whether the court of appeals’ decision in State v. Samsa, 2015 WI App 6, 359 Wis.2d 580, 859 N.W.2d 149, must be modified or overruled if this court determines that the right to due process prohibits consideration of COMPAS risk assessments at sentencing. For the reasons set forth below, we have not so determined and thus need not address this issue.
exercise of discretion because it employed recognized legal standards.

¶ 10 Accordingly, we affirm the order of the circuit court denying Loomis's motion for post-conviction relief requesting a resentencing hearing.

I

¶ 11 The facts of this case are not in dispute. The State contends that Loomis was the driver in a drive-by shooting. It charged him with five counts, all as a repeater: (1) First-degree recklessly endangering safety (PTAC); (2) Attempting to flee or elude a traffic officer (PTAC); (3) Operating a motor vehicle without the owner's consent; (4) Possession of a firearm by a felon (PTAC); (5) Possession of a short-barreled shotgun or rifle (PTAC).

¶ 12 Loomis denies involvement in the drive-by shooting. He waived his right to trial and entered a guilty plea to only two of the less severe charges, attempting to flee a traffic officer and operating a motor vehicle without the owner's consent. The plea agreement stated that the other counts would be dismissed but read in:

The other counts will be dismissed and read in for sentencing, although the defendant denies he had any role in the shooting, and only drove the car after the shooting occurred. The State believes he was the driver of the car when the shooting happened.

The State will leave any appropriate sentence to the Court's discretion, but will argue aggravating and mitigating factors.

After accepting Loomis's plea, the circuit court ordered a presentence investigation. The Presentence Investigation Report ("PSI") included an attached COMPAS risk assessment.

¶ 13 COMPAS is a risk-need assessment tool designed by Northpointe, Inc. to provide decisional support for the Department of Corrections when making placement decisions, managing offenders, and planning treatment. The COMPAS risk assessment is based upon information gathered from the defendant's criminal file and an interview with the defendant.

¶ 14 A COMPAS report consists of a risk assessment designed to predict recidivism and a separate needs assessment for identifying program needs in areas such as employment, housing and substance abuse. The risk assessment portion of COMPAS generates risk scores displayed in the form of a bar chart, with three bars that represent pretrial recidivism risk, general recidivism risk, and violent recidivism risk. Each bar indicates a defendant’s level of risk on a scale of one to ten.

¶ 15 As the PSI explains, risk scores are intended to predict the general likelihood that those with a similar history of offending are either less likely or more likely to commit another crime following release from custody. However, the COMPAS risk assessment does not predict the specific likelihood that an individual offender will reoffend. Instead, it provides a prediction based on a comparison of information about the individual to a similar data group.

13. "PTAC" refers to party to a crime.


15. Id. at 12, 16.

16. Id. at 3, 26.

17. Id. at 8.
¶ 16 Loomis’s COMPAS risk scores indicated that he presented a high risk of recidivism on all three bar charts. His PSI included a description of how the COMPAS risk assessment should be used and cautioned against its misuse, instructing that it is to be used to identify offenders who could benefit from interventions and to target risk factors that should be addressed during supervision.

¶ 17 The PSI also cautions that a COMPAS risk assessment should not be used to determine the severity of a sentence or whether an offender is incarcerated:

For purposes of Evidence Based Sentencing, actuarial assessment tools are especially relevant to: 1. Identify offenders who should be targeted for interventions. 2. Identify dynamic risk factors to target with conditions of supervision. 3. It is very important to remember that risk scores are not intended to determine the severity of the sentence or whether an offender is incarcerated.

(Emphasis added.)

¶ 18 At sentencing, the State argued that the circuit court should use the COMPAS report when determining an appropriate sentence:

In addition, the COMPAS report that was completed in this case does show the high risk and the high needs of the defendant. There’s a high risk of violence, high risk of recidivism, high pretrial risk; and so all of these are factors in determining the appropriate sentence.

¶ 19 Ultimately, the circuit court referenced the COMPAS risk score along with other sentencing factors in ruling out probation:

You’ve identified, through the COMPAS assessment, as an individual who is at high risk to the community.

In terms of weighing the various factors, I’m ruling out probation because of the seriousness of the crime and because your history, your history on supervision, and the risk assessment tools that have been utilized, suggest that you’re extremely high risk to re-offend.

¶ 20 In addition to the COMPAS assessment, the circuit court considered the read-in charges at sentencing. For sentencing purposes, it assumed that the factual bases for the read-in charges were true and that Loomis was at least involved in conduct underlying the read-in charges. The circuit court explained further that Loomis “needs to understand that if these shooting related charges are being read in that I’m going to view that as a serious, aggravating factor at sentencing.” Defense counsel protested the circuit court’s assumption that the read-in charges were true and explained that Loomis did not concede that he was involved in the drive-by shooting.

¶ 21 Although a review of the transcript of the plea hearing reveals miscommunications and uncertainty about the consequences of a dismissed but read-in offense, the circuit court ultimately quoted directly from a then-recent decision of this court explaining the nature of such a read-in offense. It explained to Loomis that a circuit court can consider the read-in offense at sentencing and that such consideration could increase a defendant’s sentence:

The Court: Mr. Loomis, I just—there is a recent Supreme Court decision in State v. Frey that describes what a read-in offense is. And I just want to quote from that decision so that you fully understand it.

“[T]he defendant exposes himself to the likelihood of a higher sentence within the sentencing range and the additional
possibility of restitution for the offenses
that are 'read in.' "
So you're limited in this agreement to a
sentencing range within—up to the max-
imums for the charges that you're plead-
guilty. You're agreeing, as the Su-
preme Court decision indicates, that the
charges can be read in and considered,
and that has the effect of increasing the
likelihood, the likelihood of a higher sen-
tence within the sentencing range. You
understand that?
Loomis: Yes.
¶ 22 The plea questionnaire/waiver of
rights form stated that the maximum pen-
alty Loomis faced for both charges was
seventeen years and six months imprison-
ment. The court sentenced him within the
maximum on the two charges for which he
entered a plea.18
¶ 23 Loomis filed a motion for post-
conviction relief requesting a new sentenc-
ing hearing. He argued that the circuit
court's consideration of the COMPAS risk
assessment at sentencing violated his due
process rights. Loomis further asserted
that the circuit court erroneously exercised
its discretion by improperly assuming that
the factual bases for the read-in charges
were true.
¶ 24 The circuit court held two hearings
on the post-conviction motion. At the first
hearing, the circuit court addressed Loom-
is's claim that it had erroneously exercised
its discretion in how it considered the
read-in charges. Considering the relevant
case law and legal standards, the circuit
court concluded that it had applied the
proper standard and denied Loomis's mo-
tion on that issue.

18. On the attempting to flee an officer charge,
the circuit court sentenced Loomis to four
years, with initial confinement of two years
and extended supervision of two years. For
operating a vehicle without the owner's con-

¶ 25 During the first post-conviction mo-
tion hearing, the circuit court reviewed the
plea hearing transcript and the sentencing
transcript and explained that it did not
believe Loomis's explanation:
I felt Mr. Loomis's explanation was in-
consistent with the facts. The State's
version was more consistent with the
facts and gave greater weight to the
State's version at sentencing.
¶ 26 At the second hearing, the circuit
court addressed the due process issues.
The defendant offered the testimony of an
expert witness, Dr. David Thompson, re-
grading the use at sentencing of a COM-
PAS risk assessment. Dr. Thompson
opined that a COMPAS risk assessment
should not be used for decisions regarding
incarceration because a COMPAS risk as-
essment was not designed for such use.
According to Dr. Thompson, a circuit
court's consideration at sentencing of the
risk assessment portions of COMPAS runs
a "tremendous risk of over estimating an
individual's risk and . . . mistakenly sen-
tencing them or basing their sentence on
factors that may not apply. . . ."
¶ 27 Dr. Thompson further testified that
sentencing courts have very little informa-
tion about how a COMPAS assessment
analyzes the risk:
The Court does not know how the COM-
PAS compares that individual's history
with the population that it's comparing
them with. The Court doesn't even
know whether that population is a Wis-
sconsin population, a New York popula-
tion, a California population. . . . There's
all kinds of information that the court
doesn't have, and what we're doing is
we're mis-informing the court when we
sent, the circuit court sentenced Loomis to
seven years, with four years of initial confine-
ment and three years of extended supervision,
to be served consecutively with the prior sen-
tence.
put these graphs in front of them and let them use it for sentence.

¶ 28 In denying the post-conviction motion, the circuit court explained that it used the COMPAS risk assessment to corroborate its findings and that it would have imposed the same sentence regardless of whether it considered the COMPAS risk scores. Loomis appealed and the court of appeals certified the appeal to this court.

II  
[1] ¶ 29 Whether the circuit court’s consideration of a COMPAS risk assessment violated Loomis’s constitutional right to due process presents a question of law, which this court reviews independently of the determinations of a circuit court or a court of appeals. See Jackson v. Buchler, 2010 WI 135, ¶ 39, 330 Wis.2d 279, 793 N.W.2d 826.

[2, 3] ¶ 30 “This court reviews sentencing decisions under the erroneous exercise of discretion standard.” State v. Frey, 2012 WI 99, ¶ 37, 343 Wis.2d 358, 817 N.W.2d 436. An erroneous exercise of discretion occurs when a circuit court imposes a sentence “without the underpinnings of an explained judicial reasoning process.” McCleary v. State, 49 Wis.2d 263, 278, 182 N.W.2d 512 (1971); see also State v. Gallion, 2004 WI 42, ¶ 3, 270 Wis.2d 535, 678 N.W.2d 197.

[4–6] ¶ 31 Additionally, a sentencing court erroneously exercises its discretion when its sentencing decision is not based on the facts in the record or it misapplies the applicable law. State v. Travis, 2013 WI 38, ¶ 16, 347 Wis.2d 142, 882 N.W.2d 491. It misapplies the law when it relies on clearly irrelevant or improper factors. McCleary, 49 Wis.2d at 278, 182 N.W.2d 512. The defendant bears the burden of proving such reliance by clear and convincing evidence. State v. Harris, 2010 WI 79, ¶ 3, 326 Wis.2d 685, 786 N.W.2d 409.

[7] ¶ 32 In a similar manner we review the issue of whether a circuit court erroneously exercises its discretion when it relies on the factual basis of the read-in charge when fashioning the defendant’s sentence. Frey, 343 Wis.2d 358, ¶¶ 37–39, 817 N.W.2d 436. “A discretionary sentencing decision will be sustained if it is based upon the facts in the record and relies on the appropriate and applicable law.” Travis, 347 Wis.2d 142, ¶ 16, 832 N.W.2d 491.

III  
¶ 33 At the outset we observe that the defendant is not challenging the use of a COMPAS risk assessment for decisions other than sentencing, and he is not challenging the use of the needs portion of the COMPAS report at sentencing. Instead, Loomis challenges only the use of the risk assessment portion of the COMPAS report at sentencing.

¶ 34 Specifically, Loomis asserts that the circuit court’s use of a COMPAS risk assessment at sentencing violates a defendant’s right to due process for three reasons: (1) it violates a defendant’s right to be sentenced based upon accurate information, in part because the proprietary nature of COMPAS prevents him from assessing its accuracy; (2) it violates a defendant’s right to an individualized sentence; and (3) it improperly uses gendered assessments in sentencing.

¶ 35 Although we ultimately conclude that a COMPAS risk assessment can be used at sentencing, we do so by circumscribing its use. Importantly, we address how it can be used and what limitations and cautions a circuit court must observe in order to avoid potential due process violations.
¶ 36 It is helpful to consider Loomis’s
due process arguments in the broader con-
text of the evolution of evidence-based
sentencing. Wisconsin has been at the
forefront of advancing evidence-based
practices. In 2004, this court’s Planning
and Policy Advisory Committee (PPAC)
created a subcommittee “to explore and
assess the effectiveness of policies and
programs . . . designed to improve public
safety and reduce incarceration.”

¶ 37 From that initial charge, Wiscon-
sin’s commitment to evidence-based prac-
tices and its leadership role have been well
documented. Initially, a variety of risk
and needs assessment tools were used by
various jurisdictions within the state. In
2012, however, the Wisconsin Department
of Corrections selected COMPAS as the
statewide assessment tool for its correc-
tional officers, providing assessment of
risk probability for pretrial release miscon-
duct and general recidivism.

¶ 38 The question of whether COMPAS
can be used at sentencing has previously
been addressed. In State v. Samsa, 2015
WI App 6, 359 Wis.2d 580, 859 N.W.2d 149,
the court of appeals approved of a
circuit court’s consideration of a COMPAS
assessment at sentencing. However, it was
not presented with the due process impli-
cations that we face here. Citing to a
principle for sentencing courts set forth in
State v. Gallion, 2002 WI App 265, ¶ 26,
258 Wis.2d 473, 654 N.W.2d 446, aff’d, 2004
WI 42, 270 Wis.2d 535, 678 N.W.2d 197,
the Samsa court emphasized that “COM-
PAS is merely one tool available to a court
at the time of sentencing.” 359 Wis.2d 580, ¶ 13, 859 N.W.2d 149.

¶ 39 In Gallion we warned of ad hoc
decision making at sentencing: “Experi-
ence has taught us to be cautious when
reaching high consequence conclusions
about human nature that seem to be intu-
tively correct at the moment. Better in-
stead is a conclusion that is based on more
complete and accurate information . . . .”
270 Wis.2d 535, ¶ 36, 678 N.W.2d 197. We
encouraged circuit courts to seek “more
complete information upfront, at the time
of sentencing. Judges would be assisted
in knowing about a defendant’s propensity
for causing harm [and] the circumstances
likely to precipitate the harm . . . .” Id., ¶ 34.

¶ 40 Concern about ad hoc decision mak-
ing is justified. A myriad of determina-
tions are made throughout the criminal
justice system without consideration of
tested facts of any kind. Questions such
as whether to require treatment, if so what
kind, and how long supervision should last
often have been left to a judge’s intuition
or a correctional officer’s standard prac-
tice.

19. Supreme Court of Wisconsin, Planning
and Policy Advisory Committee (PPAC), Effec-
tive Justice Strategies Subcommittee, Phase I:
June 2004—June 2007 Insights and Recom-
dendations, at 3–4, https://www.wicourts.gov/
courts/programs/docs/phase1final report.pdf.

20. See generally Suzanne Tallarico et. al., Na-
tional Center for State Courts (NCSC), Court
Services Division, Effective Justice Strategies
in Wisconsin: A Report of Findings and Recom-
gov/courts/programs/docs/ejsreport.pdf; Su-
preme Court of Wisconsin, Planning and Pol-
icy Advisory Committee (PPAC), Effective
Justice Strategies Subcommittee, Phase II:
Progress and Accomplishments (Nov. 13,
2013), https://www.wicourts.gov/courts/pro-
grams/docs/finalreport.pdf; Casey, Using Of-
fender Risk and Needs Assessment Information
at Sentencing, supra note 7, at 41.

21. Pamela M. Casey et al., National Center
for State Courts (NCSC), Center for Sentenc-
ing Initiatives, Research Division, Use of Risk
and Needs Assessment Information at Sentenc-
ing: La Crosse County, Wisconsin at 3 (Jan.
2014), http://www.ncsc.org/media/Microsites/
Files/CSI/RNA%20Brief%20-%20La%20Cr
osse%20County%20W1%20csi.ashx.
¶ 41 The need to have additional sound information is apparent for those working in corrections, but that need is even more pronounced for sentencing courts. Sentencing decisions are guided by due process protections that may not apply to many run of the mill correctional decisions. This distinction is of import given that the risk and needs assessment tools were designed for use by those within the Department of Corrections and that design is being transitioned to a sentencing venue governed by different guiding principles.

¶ 42 In response to a call to reduce recidivism by employing evidence-based practices, several states have passed legislation requiring that judges be provided with risk assessments and recidivism data at sentencing. Other states permit, but do not mandate, the use of risk assessment tools at sentencing.

¶ 43 But other voices are challenging the efficacy of evidence-based sentencing and raise concern about overselling the results. They urge that judges be made aware of the limitations of risk assessment tools, lest they be misused:

In the main, [supporters] have been reticent to acknowledge the paucity of reliable evidence that now exists, and the limits of the interventions about which we do possess evidence. Unless criminal justice system actors are made fully aware of the limits of the tools they are

22. The process that is due under the Constitution differs with the types of decisions and proceedings involved. "Due process is flexible and calls for such procedural protections as the particular situation demands." 
Schweiker v. McClure, 456 U.S. 188, 200, 102 S.Ct. 1665, 72 L.Ed.2d 1 (1982); see also Londoner v. City and Cty. of Denver, 210 U.S. 373, 386, 28 S.Ct. 708, 52 L.Ed. 1103 (1908) ("Many requirements essential in strictly judicial proceedings may be dispensed with [in the administrative forum]").

being asked to implement, they are likely to misuse them.


¶ 44 We heed this admonition. The DOC already recognizes limitations on the PSI, instructing that “[i]t is very important to remember that risk scores are not intended to determine the severity of the sentence or whether an offender is incarcerated.” We are in accord with these limitations. Further, we set forth the corollary limitation that risk scores may not be considered as the determinative factor in deciding whether the offender can be supervised safely and effectively in the community.25

¶ 45 In addressing Loomis’s due process arguments below, we additionally raise cautions that a sentencing court must observe in order to avoid potential due process violations.

IV

[10] ¶ 46 We turn to address Loomis’s first argument that a circuit court’s consideration of a COMPAS risk assessment violates a defendant’s due process right to be sentenced based on accurate information. Loomis advances initially that the proprietary nature of COMPAS prevents a defendant from challenging the scientific validity of the risk assessment. Accordingly, Loomis contends that because a COMPAS risk assessment is attached to the PSI, a defendant is denied full access to information in the PSI and therefore cannot ensure that he is being sentenced based on accurate information.


¶ 50 Skaff reasoned that given the discretion accorded the circuit court in sentencing decisions, any significant inaccuracies would likely affect the defendant’s sentence. Id. Therefore, denial of access to the PSI denied the defendant “an essential factor of due process, i.e., a procedure conducive to sentencing based on correct information.” Id. at 57, 447 N.W.2d 84 (citing Mathews v. Eldridge, 424 U.S. 319, 335, 96 S.Ct. 893, 47 L.Ed.2d 18 (1976)).

¶ 51 Loomis analogizes the COMPAS risk assessment to the PSI in Gardner and Skaff. Northpointe, Inc., the developer of COMPAS, considers COMPAS a proprietary instrument and a trade secret. Accordingly, it does not disclose how the risk scores are determined or how the factors are weighed. Loomis asserts that because COMPAS does not disclose this information, he has been denied information which the circuit court considered at sentencing.

¶ 52 He argues that he is in the best position to refute or explain the COMPAS risk assessment, but cannot do so based solely on a review of the scores as reflected in the bar charts. Additionally, Loomis contends that unless he can review how the factors are weighed and how risk scores are determined, the accuracy of the COMPAS assessment cannot be verified.

¶ 53 Loomis’s analogy to Gardner and Skaff is imperfect. Although Loomis cannot review and challenge how the COMPAS algorithm calculates risk, he can at least review and challenge the resulting risk scores set forth in the report attached to the PSI. At the heart of Gardner and Skaff is the fact that the court relied on information the defendant did not have any opportunity to refute, supplement or explain. Gardner, 430 U.S. at 362, 97 S.Ct. 1197. That is not the case here.

¶ 54 Loomis is correct that the risk scores do not explain how the COMPAS program uses information to calculate the risk scores. However, Northpointe’s 2015 Practitioner’s Guide to COMPAS explains that the risk scores are based largely on static information (criminal history), with limited use of some dynamic variables (i.e. criminal associates, substance abuse).

¶ 55 The COMPAS report attached to Loomis’s PSI contains a list of 21 questions and answers regarding these static factors such as:

1. How many times has this person been returned to custody while on parole? 5 +
2. How many times has this person had a new charge/arrest while on probation? 4
3. How many times has this person been arrested before as an adult or juvenile (criminal arrest only)? 12

Thus, to the extent that Loomis’s risk assessment is based upon his answers to questions and publicly available data about his criminal history, Loomis had the opportunity to verify that the questions and answers listed on the COMPAS report were accurate.

¶ 56 Additionally, this is not a situation in which portions of a PSI are considered by the circuit court, but not released to the defendant. The circuit court and Loomis had access to the same copy of the risk assessment. Loomis had an opportunity to challenge his risk scores by arguing that evidence do not apply at sentencing, we need not address that argument here. State v. Straszkowski, 2008 WI 65, ¶ 52, 310 Wis.2d 259, 750 N.W.2d 835.

27. In a similar vein, Loomis asserts that the COMPAS assessment would not pass the Daubert test and therefore would be inadmissible at trial. See Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 113 S.Ct. 2786, 125 L.Ed.2d 469 (1993). Given that the rules of evidence do not apply at sentencing, we need not address that argument here. State v. Straszkowski, 2008 WI 65, ¶ 52, 310 Wis.2d 259, 750 N.W.2d 835.

other factors or information demonstrate their inaccuracy.

¶ 57 Yet, regardless of whether Gardner and Skaff are analogous to this case, Loomis correctly asserts that defendants have the right to be sentenced based on accurate information. *Travis*, 347 Wis.2d 142, ¶ 17, 832 N.W.2d 491.

¶ 58 Some states that use COMPAS have conducted validation studies of COMPAS concluding that it is a sufficiently accurate risk assessment tool. New York State’s Division of Criminal Justice Services conducted a study examining a COMPAS assessment’s recidivism scale’s effectiveness and predictive accuracy and concluded that “the Recidivism Scale worked effectively and achieved satisfactory predictive accuracy.” Unlike New York and other states, Wisconsin has not yet completed a statistical validation study of COMPAS for a Wisconsin population.

¶ 59 However, Loomis relies on other studies of risk assessment tools that have raised questions about their accuracy. For example, he cites to a 2007 California Department of Corrections and Rehabilitation (“CDCR”) study which concludes that although COMPAS appears to be assessing criminogenic needs and recidivism risk, “there is little evidence that this is what [ ] COMPAS actually assesses.”

¶ 60 The California Study reached the further conclusion that there “is no sound evidence that the COMPAS can be rated consistently by different evaluators, that it assesses the criminogenic needs it purports to assess, and (most importantly) that it predicts inmates’ recidivism for...”

29. This analysis references current research studies in order to caution sentencing courts regarding concerns that have been raised about evidence-based risk assessment tools such as COMPAS. However, we are not in a position to evaluate or opine on the scientific reliability of this data. Accordingly, this opinion should not be read as an endorsement of any particular research study or article, regardless of whether its conclusion is critical or supportive of the COMPAS risk assessment tool.


32. The State acknowledges that no method of risk assessment is without error. For example, the State cites to a primer on COMPAS which summarizes multiple studies that have assessed COMPAS’s predictive validity as moderate, with scores ranging from .50–.73. Pamela M. Casey, et al., National Center for State Courts (NCSC), *Offender Risk & Needs Assessment Instruments: A Primer for Courts* at A–23 (2014), http://www.ncsc.org//media/Microsites/Files/CSI/BJA%20RNA%20Final%20Report_Combined%20Files%208–22–14.ashx. However, it also summarizes other studies that assign COMPAS higher scores of .70 for internal consistency and .70–1.0 for re-test reliability. *Id.*

Ultimately, the authors of the study could not recommend that CDCR use COMPAS for individuals.35

¶ 61 Subsequently, however, the CDCR published its 2010 final report on California’s COMPAS validation study.36 The 2010 study concluded that although not perfect, “COMPAS is a reliable instrument . . . .” 37 Specifically, it explained that the general recidivism risk scale achieved the value of .70, which is the conventional standard, though the violence risk scale did not.38

¶ 62 In addition to these problems, there is concern that risk assessment tools may disproportionately classify minority offenders as higher risk, often due to factors that may be outside their control, such as familial background and education.39 Other state studies indicate that COMPAS is more predictive of recidivism among white offenders than black offenders.40

¶ 63 A recent analysis of COMPAS’s recidivism scores based upon data from 10,000 criminal defendants in Broward County, Florida, concluded that black defendants “were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism.” 41 Likewise, white defendants were more likely than black defendants to be incorrectly flagged as low risk.42 Although Northpointe disputes this analysis, this study and others raise concerns regarding how a COMPAS assessment’s risk factors correlate with race.43

¶ 64 Additional concerns are raised about the need to closely monitor risk assessment tools for accuracy. At least one commentator has explained that in order to remain accurate, risk assessment tools “must be constantly re-normed for changing populations and subpopulations.” Klingele, The Promises and Perils of Evidence-Based Corrections, 91 Notre Dame L. Rev. at 576. Accordingly, jurisdictions that utilize risk assessment tools must ensure they have the capacity for maintaining those tools and monitoring their continued accuracy. Id. at 577.

¶ 65 Focusing exclusively on its use at sentencing and considering the expressed due process arguments regarding accuracy, we determine that use of a COMPAS risk assessment must be subject to certain cautions in addition to the limitations set forth herein.

¶ 66 Specifically, any PSI containing a COMPAS risk assessment must inform the sentencing court about the following cautions regarding a COMPAS risk assessment’s accuracy: (1) the proprietary

34. Id.
35. Id.
37. Id. at 29.
38. Id.
42. Id.
nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are to be determined; (2) risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed; (3) some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism; and (4) risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations.

Providing information to sentencing courts on the limitations and cautions attendant with the use of COMPAS risk assessments will enable courts to better assess the accuracy of the assessment and the appropriate weight to be given to the risk score.

V

¶ 67 We address next Loomis's argument that a circuit court's consideration of a COMPAS risk assessment amounts to sentencing based on group data, rather than an individualized sentence based on the charges and the unique character of the defendant. As this court explained in Gallion, individualized sentencing "has long been a cornerstone to Wisconsin's criminal justice jurisprudence." 270 Wis.2d 535, ¶ 48, 678 N.W.2d 197.

¶ 68 If a COMPAS risk assessment were the determinative factor considered at sentencing this would raise due process challenges regarding whether a defendant received an individualized sentence. As the defense expert testified at the post-conviction motion hearing, COMPAS is designed to assess group data. He explained that COMPAS can be analogized to insurance actuarial risk assessments, which identify risk among groups of drivers and allocate resources accordingly.

¶ 69 Similarly, the 2015 Practitioner's Guide to COMPAS explains that "[r]isk assessment is about predicting group behavior . . . it is not about prediction at the individual level." Risk scales are able to identify groups of high-risk offenders—not a particular high-risk individual. A pointed example of potential misunderstanding arising from the use of group data is that an individual who has never committed a violent offense may nevertheless be labeled as a high risk for recidivism on the violent risk scale. As the DOC explains: "[a]n offender who is young, unemployed, has an early age-at-first-arrest and a history of supervision failure, will score medium or high on the Violence Risk Scale even though the offender never had a violent offense." 46

¶ 70 To ameliorate this problem, the DOC explains that "staff are predicted to disagree with an actuarial risk assessment (e.g. COMPAS) in about 10% of the cases due to mitigating or aggravating circumstances to which the assessment is not sensitive." Thus, "staff should be encouraged to use their professional judgment and override the computed risk as appropriate . . . ." 48

¶ 71 Just as corrections staff should disregard risk scores that are inconsistent with other factors, we expect that circuit courts will exercise discretion when assess-

45. Id.
47. Id.
ing a COMPAS risk score with respect to each individual defendant.

¶ 72 Ultimately, we disagree with Loomis because consideration of a COMPAS risk assessment at sentencing along with other supporting factors is helpful in providing the sentencing court with as much information as possible in order to arrive at an individualized sentence. In Gallion, this court explained that circuit courts “have an enhanced need for more complete information upfront, at the time of sentencing.” 270 Wis.2d 535, ¶ 34, 678 N.W.2d 197.

¶ 73 COMPAS has the potential to provide sentencing courts with more complete information to address this enhanced need. The Indiana Supreme Court examined similar risk assessment tools and explained that these tools assist courts in weighing all the sentencing factors:

Such assessment instruments enable a sentencing judge to more effectively evaluate and weigh several express statutory sentencing considerations such as criminal history, the likelihood of affirmative response to probation or short term imprisonment, and the character and attitudes indicating that a defendant is unlikely to commit another crime. Malenchik v. State, 928 N.E.2d 564, 574 (Ind.2010) (internal quotations omitted).

¶ 74 However the due process implications compel us to caution circuit courts because COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders—not a particular high-risk individual. Accordingly, a circuit court is expected to consider this caution as it weighs all of the factors that are relevant to sentencing an individual defendant.

VI

¶ 75 We turn now to address Loomis’s argument that a COMPAS risk assessment’s use of gender violates a defendant’s due process rights. Relying on Harris, 326 Wis.2d 685, ¶ 33, 786 N.W.2d 409, Loomis asserts that because COMPAS risk scores take gender into account, a circuit court’s consideration of a COMPAS risk assessment violates a defendant’s due process right not to be sentenced on the basis of gender.

¶ 76 Due to the proprietary nature of COMPAS, the parties dispute the specific method by which COMPAS considers gender. Loomis asserts that gender is used as a criminogenic factor or solely for statistical norming, Loomis objects to any use of gender in calculating COMPAS’s risk scores. In response, the State contends that considering gender in a COMPAS risk assessment is necessary to achieve statistical accuracy. The State argues that if the instrument includes gender, then comparing men and women to each offender to a “norming” group of his or her own gender.

¶ 77 Regardless of whether gender is used as a criminogenic factor or solely for statistical norming, Loomis objects to any use of gender in calculating COMPAS’s risk scores. In response, the State contends that considering gender in a COMPAS risk assessment is necessary to achieve statistical accuracy. The State argues that because men and women have different rates of recidivism and different rehabilitation potential, a gender neutral risk assessment would provide inaccurate results for both men and women.

¶ 78 Both parties appear to agree that there is statistical evidence that men, on average, have higher recidivism and violent crime rates compared to women. Commentators also have noted and discussed statistical evidence differentiating men and women in this context. See, e.g., Sonja B. Starr, Evidence-Based Sentencing and the Scientific Rationalization of Discrimination, 66 Stan. L. Rev. 803, 813 (2014) (“if the instrument includes gender,
men will always receive higher risk scores than otherwise-identical women (because, averaged across all cases, men have higher recidivism rates...’); John Monahan, A Jurisprudence of Risk Assessment: Forecasting Harm Among Prisoners, Predators, and Patients, 92 Va. L. Rev. 391, 416 (2006) (‘That women commit violent acts at a much lower rate than men is a staple in criminology and has been known for as long as official records have been kept.’).

¶ 79 However, Loomis asserts that even if statistical generalizations based on gender are accurate, they are not necessarily constitutional. He cites to Craig v. Boren, 429 U.S. 190, 208–210, 97 S.Ct. 451, 50 L.Ed.2d 397 (1976), a case where the United States Supreme Court concluded that an Oklahoma law that prohibited the sale of 3.2% beer to men under the age 21 and to women under the age of 18 violated the equal protection clause of the Fourteenth Amendment. The Court explained that although state officials offered sociological or empirical justifications for the gender-based difference in the law, “the principles embodied in the Equal Protection Clause are not to be rendered inapplicable by statistically measured but loose-fitting generalities concerning the drinking tendencies of aggregate groups.” Id. at 208–09, 97 S.Ct. 451.

¶ 80 Notably, however, Loomis does not bring an equal protection challenge in this case. Thus, we address whether Loomis’s constitutional due process right not to be sentenced on the basis of gender is violated if a circuit court considers a COMPAS risk assessment at sentencing. See Harris, 326 Wis.2d 685, ¶ 33, 786 N.W.2d 409.

¶ 81 Loomis misinterprets Harris in arguing that a sentencing court cannot consider a COMPAS risk assessment because it takes gender into account in calculating risk scores. In Harris, the defendant asserted that the circuit court imposed its sentence on the basis of gender when it criticized him for being unemployed while his child’s mother worked. Id., ¶ 61. Harris argued that he should not be penalized for being a stay-at-home father and that the circuit court used this fact as an aggravating factor when fashioning his sentence. Id.

¶ 82 In determining whether the circuit court improperly considered gender in sentencing Harris, this court concluded that there was a factual basis underlying the circuit court’s statements that was not related to Harris’s gender. Id., ¶¶ 62–63. The record revealed that Harris was not his daughter’s stay-at-home primary caregiver and that other factors demonstrated that he was not a responsible father. Id., ¶ 63.

¶ 83 Likewise, there is a factual basis underlying COMPAS’s use of gender in calculating risk scores. Instead, it appears that any risk assessment tool which fails to differentiate between men and woman will misclassify both genders. As one commenter noted, “the failure to take gender into consideration, at least when predicting recidivism risk, itself is unjust.” Melissa Hamilton, Risk–Needs Assessment: Constitutional and Ethical Challenges, 52 Am. Crim. L. Rev. 231, 255 (Spring 2015). Thus, if the inclusion of gender promotes accuracy, it serves the interests of institutions and defendants, rather than a discriminatory purpose. Id.

¶ 84 Additionally, Harris concluded that the defendant had not met his burden of proving by clear and convincing evidence that the circuit court actually relied on gender as a factor in imposing its sentence. 326 Wis.2d 685, ¶ 64, 786 N.W.2d 409. It explained that the “circuit court considered the proper factors—it evaluated the gravity of the offense, Harris’s character, and the public’s need for protection.” Id., ¶¶ 65, 67. “The circuit court
thoroughly explained its reasons for the sentence it imposed, and all of the potentially offensive comments flagged by both Harris and the court of appeals bear a reasonable nexus to proper sentencing factors."

Here, as in Harris, Loomis has not met his burden of showing that the circuit court actually relied on gender as a factor in imposing its sentence. The circuit court explained that it considered multiple factors that supported the sentence it imposed:

In terms of weighing the various factors, I’m ruling out probation because of the seriousness of the crime and because your history, your history on supervision, and the risk assessment tools that have been utilized, suggest that you’re extremely high risk to re-offend.

In addition to the COMPAS risk assessment, the seriousness of the crime and Loomis’s criminal history both bear a nexus to the sentence imposed. See Gallion, 270 Wis.2d 555, ¶ 46, 678 N.W.2d 197 (“we require that the court, by reference to the relevant facts and factors, explain how the sentence’s component parts promote the sentencing objectives.”). See also Harris v. State, 75 Wis.2d 513, 519, 250 N.W.2d 7 (1977) (relevant sentencing factors include past record of criminal offenses, history of undesirable behavior patterns, and results of presentence investigation).

We determine that COMPAS’s use of gender promotes accuracy that ultimately inures to the benefit of the justice system including defendants. Additionally, we determine that the defendant failed to meet his burden of showing that the sentencing court actually relied on gender as a factor in sentencing. Thus, we conclude that the use of the COMPAS risk assessment at sentencing did not violate Loomis’s right to due process.

VII

Next, we address the permissible uses for a COMPAS risk assessment at sentencing. Then we set forth the limitations and cautions that a sentencing court must observe when using COMPAS.

Although it cannot be determinative, a sentencing court may use a COMPAS risk assessment as a relevant factor for such matters as: (1) diverting low-risk prison-bound offenders to a non-prison alternative; (2) assessing whether an offender can be supervised safely and effectively in the community; and (3) imposing terms and conditions of probation, supervision, and responses to violations.

First, COMPAS may be useful in identifying prison-bound offenders who are at low risk to reoffend for purposes of diverting them to non-prison alternatives and aids in the decision of whether to suspend all or part of an offender’s sentence.

Second, risk assessment tools such as COMPAS can be helpful to assess an offender’s risk to public safety and can inform the decision of whether the risk of re-offense presented by the offender can be safely managed and effectively reduced through community supervision and services.

Third, COMPAS may be used to inform decisions about the terms and con-
ditions of probation and supervision. Risk assessments can be useful in identifying low-risk offenders who do not require intensive supervision and treatment programs. Together with a need assessment, a risk assessment may inform public safety considerations related to offender risk management, and therefore may be used to provide guidance about the level of supervision and control needed for an offender placed on probation or released to extended supervision. Specifically, it may inform decisions such as reporting requirements, drug testing, electronic monitoring, community service, and the most appropriate treatment strategies.

¶ 92 Thus, a COMPAS risk assessment may be used to "enhance a judge's evaluation, weighing, and application of the other sentencing evidence in the formulation of an individualized sentencing program appropriate for each defendant." See Malenchik, 928 N.E.2d at 573. As the court of appeals explained in Samsa, "COMPAS is merely one tool available to a court at the time of sentencing and a court is free to rely on portions of the assessment while rejecting other portions." 359 Wis.2d 580, ¶ 13, 859 N.W.2d 149.

¶ 93 However, the use of a COMPAS risk assessment at sentencing must be subject to certain limitations. As noted above, the DOC already recognizes these limitations on the PSI, instructing that "[i]t is very important to remember that risk scores are not intended to determine the severity of the sentence or whether an offender is incarcerated." This is also the first "Guiding Principle" of the National Center for State Courts ("NCSC") report on using offender risk and needs assessment at sentencing, which instructs that:

Risk and need assessment information should be used in the sentencing decision to inform public safety considerations related to offender risk reduction and management. It should not be used as an aggravating or mitigating factor in determining the severity of an offender's sanction.

¶ 94 Additionally, we set forth the corollary limitation that risk scores may not be used as the determinative factor in deciding whether the offender can be supervised safely and effectively in the community. This is consistent with the second "Guiding Principle" of the National Center for State Courts.

¶ 95 The decision of the Indiana Supreme Court in Malenchik, 928 N.E.2d at 575, provides additional guidance here. It limited the use of risk assessments, explaining that they "are not intended to serve as aggravating or mitigating circumstances nor to determine the gross length of sentence, but a trial court may employ such results in formulating the manner in which a sentence is to be served."

¶ 96 Additionally, a COMPAS risk assessment was not designed to address all of the goals of a sentence. Its aim is addressing the treatment needs of an individual and identifying the risk of recidivism. Sentencing, on the other hand, is meant to address additional purposes. See State v. Dowdy, 2012 WI 12, ¶ 97, 338 Wis.2d 565, 808 N.W.2d 691 (Abrahamson, 52. Id. at 16.

53. Id.

54. Id. at 11; Casey, Offender Risk & Needs Assessment Instruments: A Primer for Courts, supra note 32, at 2.

55. Casey, Using Offender Risk and Needs Assessment Information at Sentencing, supra note 7, at 14, 16.

56. Id. at 11.

57. Id. at 14.
C.J., dissenting) (“It is commonly understood that there are four main purposes of sentencing: (1) deterrence; (2) rehabilitation; (3) retribution; and (4) segregation.”).

¶ 97 Because of these disparate goals, using a risk assessment tool to determine the length and severity of a sentence is a poor fit. As scholars have observed, “[a]ssessing the risk of future crime plays no role in sentencing decisions based solely on backward-looking perceptions of blameworthiness, … is not relevant to deterrence, … and should not be used to sentence offenders to more time than they morally deserve.” 58

[18] ¶ 98 Thus, a sentencing court may consider a COMPAS risk assessment at sentencing subject to the following limitations. As recognized by the Department of Corrections, the PSI instructs that risk scores may not be used: (1) to determine whether an offender is incarcerated; or (2) to determine the severity of the sentence. Additionally, risk scores may not be used as the determinative factor in deciding whether an offender can be supervised safely and effectively in the community.59

[19] ¶ 99 Importantly, a circuit court must explain the factors in addition to a COMPAS risk assessment that independently support the sentence imposed. A COMPAS risk assessment is only one of many factors that may be considered and weighed at sentencing.


[20] ¶ 100 Any Presentence Investigation Report (“PSI”) containing a COMPAS risk assessment filed with the court must contain a written advisement listing the limitations.60 Additionally, this written advisement should inform sentencing courts of the following cautions as discussed throughout this opinion:

• The proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined.

• Because COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders—not a particular high-risk individual.

• Some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism.

• A COMPAS risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed. Risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations.

• COMPAS was not developed for use at sentencing, but was intended for use by the Department of Corrections in mak-
ing determinations regarding treatment, supervision, and parole.

¶ 101 It is important to note that these are the cautions that have been identified in the present moment. For example, if a cross-validation study for a Wisconsin population is conducted, then flexibility is needed to remove this caution or explain the results of the cross-validation study. Similarly, this advisement should be regularly updated as other cautions become more or less relevant as additional data becomes available.

VIII

¶ 102 We apply next the relevant permissible uses, limitations and cautions to an examination of the record in this case. Loomis argues that he is entitled to a resentencing hearing because the circuit court considered the COMPAS risk assessment in imposing his sentence. According to Loomis, this is a violation of his due process rights because he argues that a COMPAS risk assessment should never be considered at sentencing.

¶ 103 Notably, Loomis does not argue that the other factors the court considered at sentencing were insufficient to support the sentence he received. In fact, at oral argument Loomis’s counsel acknowledged that he would not be challenging the sentence imposed if it were devoid of any reference to the COMPAS risk assessment. He argues instead that even if there are other bases for the circuit court’s sentence, this does not overcome the error of considering the COMPAS risk assessment.

¶ 104 As discussed above, if used properly with an awareness of the limitations and cautions, a circuit court’s consideration of a COMPAS risk assessment at sentencing does not violate a defendant’s right to due process. The circuit court here was aware of the limitations. Two limitations were set forth by the DOC in the PSI containing the COMPAS report. Thus, when Loomis was sentenced, the circuit court was aware that “risk scores are not intended to determine the severity of the sentence or whether an offender is incarcerated.” The third limitation, that a COMPAS risk assessment may not be determinative in deciding whether a defendant may be supervised safely and effectively in the community is a corollary limitation to those already set forth in the PSI.

¶ 105 With respect to the cautions this opinion requires, we intend that they be used to inform courts of due process implications. These cautions will enable sentencing courts to better assess the weight to be given to the COMPAS risk scores, circumventing potential due process violations. Here, however, the record reflects that although the circuit court referenced the risk assessment at sentencing, the court essentially gave it little or no weight.

¶ 106 At the post-conviction motion hearing, the circuit court explained that it used the COMPAS risk assessment to corroborate its findings and that it would have imposed the same sentence regardless of whether it considered the COMPAS risk scores:

I think it’s accurate and safe for this court to say that had there been absolutely no mention of the risk assessment tool in the Presentence Report, had the COMPAS not been attached to the presentence report, that the sentence would have been exactly the same because of the court’s evaluation of the sentencing factors that are required under the [ ] law.

¶ 107 The circuit court explained that it considered the COMPAS risk assessment as “an observation” to reinforce its assessment of the other factors it considered:
In other words, the factors that were cited by the court suggested low probability of success on supervision and serious crime. And that was reinforced by the fact that the risk assessment tool shows high risk in all areas.

The court identified factors that were apparent to Mr. Loomis’s history and the nature of the offenses. And then went to the COMPAS as an instrument, basically, that supported that evaluation.

¶ 108 This is consistent with the sentencing transcript, in which the circuit court explained that it was also considering the seriousness of the crime and Loomis’s criminal history and history on supervision in ruling out probation. A review of the sentencing transcript reveals that the circuit court also addressed and discussed the gravity of the offense, the character and rehabilitative needs of the defendant, and the need to protect the public. See Gallion, 270 Wis.2d 535, ¶ 13, 678 N.W.2d 197.

¶ 109 Thus, the record reflects that the sentencing court considered the appropriate factors and was aware of the limitations associated with the use of the COMPAS risk assessment. Ultimately, although the circuit court mentioned the COMPAS risk assessment, it was not determinative in deciding whether Loomis should be incarcerated, the severity of the sentence or whether he could be supervised safely and effectively in the community.

¶ 110 Additionally, although the circuit court was unaware of the cautions set forth above, those cautions are required in part to ensure that undue weight is not given to the COMPAS risk scores. As the circuit court explained at the post conviction hearing, it would have imposed the exact same sentence without it. Accordingly, we determine that the circuit court’s consideration of COMPAS in this case did not violate Loomis’s due process rights.

IX

¶ 111 As a final matter, Loomis argues that the circuit court improperly gave undue weight to read-in charges at sentencing. He asserts that not only did the circuit court appear to misunderstand the difference between dismissed and read-in charges, but it improperly assumed that the factual basis for the read-in charges was true.

¶ 112 In Frey, 343 Wis.2d 358, ¶ 61, 817 N.W.2d 436, this court clarified how the read-in procedure and dismissed charges fit into the plea bargaining process. The circuit court can consider uncharged or unproven offenses regardless of whether the defendant consented to having the charges read in or dismissed outright. Id., ¶ 47.

¶ 113 Frey explained that it is preferable for the circuit court to “acknowledge and discuss dismissed charges, if they are considered by the court, giving them appropriate weight and describing their relationship to a defendant’s character and behavioral pattern or to the incident that serves as the basis for a plea.” Id., ¶ 54. Open discussion of dismissed charges is consistent with the sentencing methodology set forth in Gallion and allows the defendant the opportunity to explain or dispute the charges. Id.

¶ 114 Additionally, read-in charges are expected to be considered at sentencing “with the understanding that the read-in charges could increase the sentence up to the maximum that the defendant could receive for the conviction in exchange for the promise not to prosecute those additional offenses.” Id., ¶ 68.

¶ 115 Loomis asserts that the circuit court appeared to misunderstand the difference between dismissed charges and those that are dismissed but read in. At
sentencing, the circuit court initially erred in its statement regarding dismissed and read-in charges when it stated that it could not consider the dismissed charges at all, but would consider the read-in charges as true. However, the circuit court took a break from the hearing to review Frey and continued the hearing under the proper framework.

¶ 116 Although the circuit court may have initially misstated that under Frey there is no distinction between read-in charges and dismissed charges, the circuit court’s consideration of the read-in charges was not an erroneous exercise of discretion. It subsequently clarified the proper use.

¶ 117 During the plea hearing, quoting directly from Frey, the circuit court advised Loomis of the proper legal standard regarding how it would consider the read-in offenses at sentencing. It allowed both sides to debate the merits of the charges and ultimately believed the State’s version of events was more credible.

¶ 118 At the post-conviction motion hearing, the circuit court reviewed the plea hearing transcript and the sentencing transcript and explained how it weighed the facts in addressing the read-in charges:

The Court had to give weight, greater weight or lesser weight to the facts that’s relating [sic] to the shooting. I felt Mr. Loomis’s explanation was inconsistent with the facts. The State’s version was more consistent with the facts and gave greater weight to the State’s version at sentencing.

¶ 119 Thus, the circuit court weighed the facts, assessed the credibility and the recognized legal standards for read-in offenses. Accordingly, we conclude that the circuit court’s consideration of the read-in charges was not an erroneous exercise of discretion.

X

¶ 120 Ultimately, we conclude that if used properly as set forth herein, a circuit court’s consideration of a COMPAS risk assessment at sentencing does not violate a defendant’s right to due process and that the circuit court did not erroneously exercise its discretion here.

¶ 121 We further conclude that the circuit court’s consideration of the read-in charges was not an erroneous exercise of discretion.

¶ 122 Accordingly, we affirm the order of the circuit court.

The order of the circuit court is affirmed.

PATIENCE DRAKE ROGGENSACK, Chief Justice. (concurring).

¶ 123 I agree with much of the majority opinion’s discussion and I concur in its result; however, I write to clarify that while our holding today permits a sentencing court to consider COMPAS, we do not conclude that a sentencing court may rely on COMPAS for the sentence it imposes. Because at times the majority opinion interchangeably employs consider and rely when discussing a sentencing court’s obligations and the COMPAS risk assessment tool, our decision could mistakenly be read as permitting reliance on COMPAS.1 Therefore, I write to clarify for the reader.2

1. See, e.g., majority op., ¶¶ 8, 31, 48, 82, 85, 98–99.

2. Contrary to the manner in which the majority opinion sometimes employs "consider" and "rely," they are not interchangeable. "Rely" is defined as "to be dependent" or "to place full confidence." Webster’s New Collegiate Dictionary 977 (1974). Therefore, to permit circuit courts to rely on COMPAS is to
¶ 124 At sentencing, the circuit court is to consider three primary factors: gravity of the offense, character of the offender and the need to protect the public. State v. Alexander, 2015 WI 6, ¶ 22, 360 Wis.2d 292, 858 N.W.2d 662. A circuit court’s proper exercise of sentencing discretion includes an individualized sentence based on the facts of the case and may include explaining how the sentence imposed furthers the circuit court’s objectives. Id. (citing State v. Harris (Landray M.), 2010 WI 79, ¶ 29, 326 Wis.2d 685, 786 N.W.2d 409).  

¶ 125 A sentencing court must articulate the factors that it considered at sentencing and how they affected the sentence imposed. State v. Harris (Denia), 119 Wis.2d 612, 623, 350 N.W.2d 633 (1984). It is through this articulation that we determine whether the circuit court properly exercised its sentencing discretion. Id. Defendants have a due process right not to be sentenced in reliance on improper factors such as on race or gender. Harris (Landray M.), 326 Wis.2d 685, ¶ 33, 786 N.W.2d 409.  

¶ 126 The circuit court’s consideration of various sentencing factors is afforded a “strong presumption of reasonableness because the circuit court is best suited to consider the relevant factors and demeanor of the convicted defendant.” State v. Gallion, 2004 WI 42, ¶ 18, 270 Wis.2d 535, 678 N.W.2d 197 (internal quotation marks omitted). Therefore, a circuit court’s sentencing decision is upheld unless it exhibits an erroneous exercise of discretion by sentencing based on irrelevant or improper factors. Id., ¶ 17; Harris (Landray M.), 326 Wis.2d 685, ¶ 30, 786 N.W.2d 409. In addition, any reference to a potentially improper sentencing factor is reviewed in the context of the circuit court’s sentencing record as a whole. Harris (Landray M.), 326 Wis.2d 685, ¶ 45, 786 N.W.2d 409.  

¶ 127 As the majority opinion aptly explains, the circuit court here appropriately considered numerous sentencing factors when imposing sentence and merely mentioned the defendant’s COMPAS risk assessment in passing. The circuit court detailed the three primary sentencing factors and explained how the facts of the case warranted the sentence imposed. Therefore, I agree with the majority opinion that circuit courts may consider a COMPAS risk assessment along with a multitude of other relevant factors at sentencing, as was done in this case.  

¶ 128 However, one of my concerns is that the certified question frames the issue presented as “whether the right to due process prohibits circuit courts from relying on COMPAS assessments when imposing sentence.” The majority opinion concludes that “if used properly with an awareness of the limitations and cautions, a circuit court’s consideration of a COMPAS risk assessment at sentencing does permit circuit courts to depend on COMPAS in imposing sentence. On the other hand, ‘consider’ is defined as ‘to observe’ or to ‘contemplate’ or to ‘weigh.’ Id. at 241–42. Therefore, to permit circuit courts to consider COMPAS is to permit circuit courts to observe a COMPAS risk assessment and weigh it along with other relevant factors in imposing sentence.  

3. See State v. Gallion, 2004 WI 42, ¶ 43 n. 11, 270 Wis.2d 535, 678 N.W.2d 197 which identifies numerous, supplemental sentencing factors that circuit courts may consider under the appropriate circumstances of each case.


5. Id., ¶¶ 85, 104–10.


not violate a defendant’s right to due process.’’\(^8\) I agree that “consideration” of COMPAS does not contravene defendant’s right to due process.

\(\S\) 129 However, the question presented on certification is whether due process prohibits circuit courts from relying on COMPAS, and then the majority opinion’s answering that question in the negative, even though it employs the word “consideration,” may cause the majority opinion to be read as permitting circuit court reliance on COMPAS. Stated otherwise, rather than merely considering COMPAS as one of many factors relevant to sentencing, the majority opinion, due to its interchangeable use of “rely” and “consider,” together with the certified question, may be read to permit a circuit court to rely on COMPAS to determine the appropriate sentence. Reliance would violate due process protections. Accordingly, I write to clarify our holding in the majority opinion: consideration of COMPAS is permissible; reliance on COMPAS for the sentence imposed is not permissible.

SHIRLEY S. ABRAHAMSON, J. (concurring).

\(\S\) 130 I join the majority opinion. It describes the salient issues raised by considering COMPAS at sentencing, and informs the bench and the bar of the limitations and cautions that should be observed in considering COMPAS in sentencing. It underscores that we are addressing the use of a research-based tool and that it is incumbent upon actors in the criminal justice system to recognize that additional research data may become available in the future and different, better tools may be developed.

\(\S\) 131 I write separately to make two points:

\(\S\) 132 First, I conclude that in considering COMPAS (or other risk assessment tools) in sentencing, a circuit court must set forth on the record a meaningful process of reasoning addressing the relevance, strengths, and weaknesses of the risk assessment tool.

\(\S\) 133 Second, this court’s lack of understanding of COMPAS was a significant problem in the instant case. At oral argument, the court repeatedly questioned both the State’s and defendant’s counsel about how COMPAS works. Few answers were available.

\(\S\) 134 Northpointe, the company that created COMPAS, sought to file an amicus brief in the instant case to discuss the history, accuracy, and efficacy of COMPAS, as well as the use of risk assessment tools like COMPAS throughout the criminal justice system.

\(\S\) 135 The court denied (over my dissent and without comment) Northpointe’s motion to file an amicus brief. The denial was a mistake. The court needed all the help it could get. The majority opinion considers publications by Northpointe. Why could it not consider an amicus brief by Northpointe?

\(\S\) 136 For these reasons, I write separately.

I

\(\S\) 137 I would hold that a circuit court, in considering COMPAS (or another risk assessment tool) in sentencing, must evaluate on the record the strengths, weaknesses, and relevance to the individualized sentence being rendered of the evidence-based tool (or, more precisely, the research-based or data-based tool).

\(\S\) 138 Such an explanation is needed, I think, because the use of risk assessment tools like COMPAS has garnered mixed

---

\(^8\) Majority op., \(\S\) 104 (emphasis added).
reviews in the scholarly literature and in popular commentary and analysis.

¶ 139 For example, although then-Attorney General Eric Holder endorsed the use of risk assessment tools in preparing and planning for the reentry of offenders into society, he cautioned against using risk assessment tools in sentencing. Attorney General Holder warned that using “static factors and immutable characteristics, like the defendant’s education level, socioeconomic background or neighborhood” in sentencing could have unintended consequences, including undermining our goal of “individualized justice, with charges, convictions, and sentences befitting the conduct of each defendant and the particular crime he or she commits.”


Wisconsin law also recognizes the need for individualized sentences. See State v. Gallion, 2004 WI 42, ¶ 48, 270 Wis.2d 535, 678 N.W.2d 197 (recognizing that individualized sentencing “has long been a cornerstone to Wisconsin’s criminal justice jurisprudence.”).

University of Wisconsin Law Professor Cecelia Klingele summarized the challenges inherent in using these tools in The Promises and Perils of Evidence-Based Corrections, 91 Notre Dame L. Rev. 537, 576–78 (2015).

2. See Jeff Larson et al., How We Analyzed the COMPAS Recidivism Algorithm, Pro Publica (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm; see also Julia Angwin et al., Machine Bias, Pro Publica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (reviewing the findings of Pro Publica’s study and discussing numerous anecdotal examples of individuals whose risks of recidivism were incorrectly assessed).

3. See, e.g., Sheldon X. Zhang et al., An Analysis of Prisoner Reentry and Parole Risk Using COMPAS and Traditional Criminal History Measures, 60 Crime & Delinquency 167, 187 (2014) (finding that a model assessing just four static variables—gender, age, age of first arrest, and number of prior arrests—performed just as well as COMPAS in predicting prior arrests); Jennifer L. Skeem & Jennifer Eno Louden, Assessment of Evidence on the Quality of Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) 28 (2007), http://www.cdcr.ca.gov/adult_research_branch/Research_Documents/COMPAS_Skeem_EnoLouden_Dec_2007.pdf (last visited July 1, 2016) (stating that “there is little evidence that the COMPAS predicts recidivism,” and “there is no evidence that the COMPAS assesses risk state, or change over time in criminogenic needs.”); but see Sharon Lansing, New York State COMPAS-Probation Risk and Need Assessment Study: Examining the Recidivism Scale’s Effectiveness and Predictive Accuracy, N.Y. Div. Justic. Servs., Office of Justice Research & Performance, at i (Sept. 2012) (concluding that COMPAS’s “Recidivism Scale worked effectively and achieved satisfactory predictive accuracy,” namely 71% accuracy); Tim Brennan et al., Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System, 36 Crim. Just. & Behavior 21, 30 (2009) (determining, in a study by three individuals for Northpointe, the company that markets COMPAS, that COMPAS’s risk models are “satisfactorily predictive . . . ”).
of new developments in evidence-based decision making and cognizant of the qualities of the tools utilized. Such a process also provides appellate courts with a meaningful record to review and provides the State, the defendant, and the public with a transparent and comprehensible explanation for the sentencing court’s decision.

II

¶ 143 With evidence-based decision making on the rise, amicus briefs evaluating the research and data will, in all likelihood, become more important. As Judge Richard Posner has written, “[m]ost judges are generalists, and increasingly we are confronted by complexities that most of us have difficulty understanding.” One way of addressing these complexities is taking a more expansive view toward accepting amicus briefs.

¶ 144 The court denied Northpointe’s motion to file an amicus brief over my dissent. See Attachment A. COMPAS is proprietary, and Northpointe considers COMPAS’s algorithms trade secrets. As a result, Northpointe does not disclose how COMPAS determines individual risk scores or how it weighs various factors in arriving at a risk score.

¶ 145 Northpointe has an obvious financial and proprietary interest in the continued use of COMPAS. The court could have taken Northpointe’s interests into account in weighing Northpointe’s amicus brief.

¶ 146 This court’s orders accepting and rejecting amicus briefs have generally not explained the court’s decision, and the orders have not been consistent. Perhaps Northpointe’s brief was rejected because Northpointe had an interest in the use of its tool.

¶ 147 In contrast, in another order denying a motion to file an amicus brief (Attachment B), the amicus had no legally cognizable interest in the case.

¶ 148 Yet in Thompson v. Craney, 199 Wis.2d 674, 546 N.W.2d 123 (1996), the court accepted an amicus brief filed by then-Assembly Speaker David T. Prosser arguing that the legislative enactment at issue in that case was constitutional.

¶ 149 In a recent case addressing similar issues to those raised in Thompson, the court permitted an amicus to file a brief and raise issues that the parties did not address. See Coyne v. Walker, No. 2013AP416, unpublished orders dated Sept. 22, 2015 and October 1, 2015.

¶ 150 Without providing an explanation for the court’s acceptance or denial of amicus briefs, we provide no guidance to lawyers and other interested persons wishing to file amicus briefs in future cases. The court should, in my opinion, take a more expansive view toward granting motions to file amicus briefs.

¶ 151 For the reasons set forth, I concur and write separately.

---

March 11, 2016

You are hereby notified that the Court has entered the following order:


The court having considered the motion of Northpointe Inc. for leave to file a non-party brief amicus curiae;

IT IS ORDERED that the motion is denied.


¶2 The court issues an order denying the motion. I would grant the motion.
¶3 Once again, the court offers no explanation of its order to assist this movant or future movants.

¶4 Northpointe, Inc. states that it wishes to file an amicus brief because it developed COMPAS, the sentencing tool at issue in the instant case. Northpointe argues that it is uniquely positioned to discuss the history, accuracy, and efficacy of COMPAS, as well as the use of actuarial risk needs assessments across the criminal justice continuum, and may bring to the court's attention certain issues and points of law that may assist the court in reaching its decision on the merits.

¶5 I surmise that the court may be denying this motion because, as the developer of COMPAS, Northpointe, Inc. would have an interest in promoting its product and might benefit or be harmed financially by the court's writing on COMPAS in State v. Loomis. Northpointe does not, however, appear to have a specific interest in the circuit court's decision about the sentence to be imposed on this particular defendant, Eric Loomis.

¶6 In addition to Northpointe's financial interest, it claims the software is proprietary information.

¶7 I would not necessarily deny a motion to file an amicus brief because the movant has a financial or other interest in the subject matter of the case or because of the claim of the software's proprietary nature. Amici generally have a strong interest in a case or an important stake in an outcome. Why else would a person or entity take the time and trouble to seek amicus status?

¶8 The propriety of using COMPAS and risk assessment instruments to reach evidence-based decisions is an emerging area of the law. COMPAS risk assessment instruments present issues that are novel, technical, and complex. The court's decision in the instant case may affect far more people than the parties.

¶9 The briefs of both parties cite to reference materials relating to COMPAS, risk-needs assessment, and evidence-based sentencing. Why cannot an amicus brief by Northpointe play the same role as the citations in the briefs to reference books or articles by Northpointe or about Northpointe's COMPAS? The court knows of Northpointe's financial and proprietary interests and can take these into account in evaluating Northpointe's amicus brief.

¶10 Judges across the country disagree regarding the merits of amicus briefs or the approach a court should take to requests to file amicus briefs. Compare, e.g., Judge Richard Posner's views expressed in Voices for Choice v. Ill. Bell. Tel. Co., 339 F.3d 542, 545 (7th Cir. 2003) (generally viewed as hostile to amicus briefs), with then-Judge Samuel Alito's views expressed in Neonatology Assoc., P.A. v. C.I.R., 293 F.3d 128, 133 (3d Cir. 2002) (generally viewed as more friendly toward amicus briefs).

¶12 My analysis of this court's orders accepting and rejecting amicus briefs over the years is that they generally do not explain the court's decision; they do not guide lawyers and other interested persons in filing amicus briefs in future cases; and they do not provide the benefit of reasoned decisions so that the court can be thoughtful and consistent in its approach to amicus briefs.

¶13 The court should, in my opinion, take a more expansive view to granting motions to file an amicus brief. A court can use all the help it can get. An amicus may have particular expertise and may present a perspective that the parties cannot or will not advocate. I agree with Judge Samuel Aalto (now Supreme Court Justice Samuel Aalto) in writing that requiring amici to be limited to persons or entities that are impartial or disinterested "is contrary to the fundamental assumption of our adversary system that strong (but fair) advocacy on behalf of opposing views promotes sound decision making." Neonatology Assocs., 293 F.3d at 131.

¶14 The court knows of Northpointe's financial and proprietary interests. If the amicus brief proves useless to the court, the court can easily set it aside.

¶15 In closing, I write to object once again to unilateral directives imposed by one member of the court on a justice's separate writings. For my prior writings objecting to this practice, see my dissent to the December 4, 2015 unpublished order of four justices setting a deadline for motions to intervene in the John Doe trilogy; my concurrence/dissent to the January 12, 2016 unpublished order of four justices granting the motion of three district attorneys to intervene in the John Doe trilogy; my separate writings in the unpublished orders granting review in three cases, State v. Finley, No. 2014AP2488-CA; Wis. Cnty v. City of Madison, No. 2015AP146; and Regency West Apis. v. City of Racine, No. 2014AP2947; my concurrence/dissent to In re Disciplinary Proceedings Against Rutkow; 2015 WI 1, ¶35-54, Wis. 2d ___, ___ N.W.2d ___; and my concurrence/dissent to the March 8, 2016 order denying judicial notice in Clark v. Am. Cyanamid Co., No. 2014AP775.
¶16 As I have often written, I favor deadlines for writings of justices and staff. But deadlines should be imposed by the court and uniformly and consistently applied.

¶17 For the reasons set forth, I dissent and would grant the request to file an amicus brief.

Prosser, J., dissents and would grant the motion.
May 19, 2016

To:

Hon. Richard J. Sankovitz
Milwaukee County Circuit Court Judge
821 W. State St.
Milwaukee, WI 53233

John Barrett
Milwaukee County Clerk of Circuit Court
901 N. 9th St., Rm. G-8
Milwaukee, WI 53233

Timothy A. Bascom
Bascom, Budish & Ceman, S.C.
2600 North Mayfair Rd., #1140
Milwaukee, WI 53226

Donald H. Carlson
Crivello Carlson S.C.
710 N. Plankinton Ave., Ste. 500
Milwaukee, WI 53203

Stephen J. McManus
McManus & Associates LLC
12700 W. Bluemound Rd., Ste. 130
Elm Grove, WI 53122

*Additional Parties listed on Page Three

You are hereby notified that the Court has entered the following order:


The court having considered the motion of Charles P. Dykman for leave to file non-party brief amicus curiae;

IT IS ORDERED that the motion is denied.

DAVID T. PROSSER and REBECCA G. BRADLEY, J.J., did not participate.

SHIRLEY S. ABRAHAMSON, J. (dissenting). The court issues an order denying Charles Dykman's motion to file an amicus brief under Wis. Stat. § 809.19(7). I would grant the motion and would grant the Brenners' request to respond to the amicus brief.

Once again, the court offers no explanation of its order to assist this movant or future movants.

That Charles Dykman has no legally cognizable interest in the case does not bar him from filing an amicus brief. Indeed sometimes such an interest may be a barrier to filing an amicus brief. See *State v. Loomis*, No. 2015AP157-CR, unpublished order dated Mar. 11, 2016 (Abrahamson, J., dissenting).

I recognize that the amicus brief would raise issues that the parties did not address. We recently allowed such an amicus brief in *Coyne v. Walker*, No. 2013AP416. See unpublished order dated Sept. 22, 2015; see also unpublished order dated Oct. 1, 2015.

Judges across the country disagree regarding the merits of amicus briefs or the approach a court should take to requests to file amicus briefs. Compare, e.g., Judge Richard Posner’s views expressed in *Voices for Choice v. Ill. Bell. Tel. Co.*, 339 F.3d 542, 545 (7th Cir. 2003) (generally viewed as hostile to amicus briefs), with then-Judge Samuel Alito’s views expressed in *Neomatology Assoc. P.A. v. C.I.R.*, 293 F.3d 128, 133 (3d Cir. 2002) (generally viewed as more friendly toward amicus briefs).

My analysis of this court’s orders accepting and rejecting amicus briefs over the years is that the orders generally do not explain the court’s decision; they do not guide lawyers and other interested persons in filing amicus briefs in future cases; and they do not provide the benefit of reasoned decisions so that the court can be thoughtful and consistent in its approach to amicus briefs.

The court should, in my opinion, take a more expansive view to granting motions to file an amicus brief. A court should use all the help it can get. An amicus may have particular expertise and may present a perspective that the parties cannot or will not advocate.

For the reasons set forth, I dissent and would grant the request to file an amicus brief.

Diane M. Fremgen
Clerk of Supreme Court

*Additional Parties:
Bryan J. Paradise
Hennessey & Roach PC
414 E. Walnut St., Ste. 200
Green Bay, WI 54301-5020

Pamela M. Schmidt
Scopelitis Garvin Light Hanson & Feary
330 E. Kilbourn Ave., Ste. 827
Milwaukee, WI 53202

Charles P. Dykman
4611 Tonyawatha Trail
Monona, WI 53716

Timothy S. Trecek
Habush, Habush & Rotter, S.C.
777 E. Wisconsin Ave., #2300
Milwaukee, WI 53202-5302

Susan R. Tyndall
Habush Habush & Rotter, S.C.
N14 W23755 Stone Ridge Dr., Ste. 100
Waukesha, WI 53188-1147

370 Wis.2d 54
2016 WI App 42
KLISMET'S 3 SQUARES INCORPORATED, Plaintiff–Respondent,
v.
NAVISTAR, INC., Defendant–Appellant.
No. 2014AP1830.

Court of Appeals of Wisconsin.

Background: Purchaser of truck brought action against manufacturer, alleging violation of Lemon Law. Following a grant of partial summary judgment in favor of purchaser, the Circuit Court, Waupaca Coun-
ty, Raymond S. Huber, J., entered judgment in favor of purchaser after a trial. Manufacturer appealed.

Holdings: The Court of Appeals, Sherman, J., held that:
(1) accord and satisfaction did not bar Lemon Law claim;
(2) Lemon Law permitted use of 300,000 as denominator in formula for determining reasonable allowance for use;
(3) trial court’s calculation of allowance for use was reasonable; and
(4) purchaser did not intentionally prevent manufacturer from issuing refund within 30–day statutory period.

Affirmed.

1. Appeal and Error 842(9)

The Court of Appeals' review of a circuit court's determination regarding ac-
Remedies for Robots

Mark A. Lemley  
Stanford Law School  

Bryan Casey  
Stanford Law School

John M. Olin Program in Law and Economics  
Stanford Law School  
Stanford, CA 94305

Working Paper Series  
Paper No. 523

This paper can be downloaded without charge from the Social Science Research Network Electronic Paper Collection  
http://ssrn.com/abstract=3223621
Remedies for Robots

Mark A. Lemley & Bryan Casey

Engineers training an artificially-intelligent self-flying drone were perplexed. They were trying to get the drone to stay within a predefined circle and to head towards its center. Things were going well for a while. The drone received positive reinforcement for its successful flights, and it was improving its ability to navigate towards the middle quickly and accurately. Then, suddenly, things changed. When the drone neared the edge of the circle, it would inexplicably turn away from the center, leaving the circle.

What went wrong? After a long time spent puzzling over the problem, the designers realized that whenever the drone left the circle during tests, they had turned it off. Someone would then pick it up and carry it back into the circle to start again. From this pattern, the drone’s algorithm had learned—correctly—that when it was sufficiently far from the center, the optimal way to get back to the middle was to simply leave it altogether. As far as the drone was concerned, it had discovered a wormhole. Somehow, flying outside of the circle could be relied upon to magically teleport it closer to the center. And far from violating the rules instilled in it by

---

1 © 2018 Mark A. Lemley and Bryan Casey.
2 William H. Neukom Professor, Stanford Law School; partner, Durie Tangri LLP.
3 Lecturer in Law, Stanford Law School; Legal Fellow, Center for Automotive Research at Stanford (CARS).
4 This example comes from a presentation at the June 2014 Stanford Ecommerce Best Practices Conference. As far as we know it has not been previously described in print.
its engineers, the drone had actually followed them to a T. In doing so, however, it had discovered an unforeseen shortcut—one that subverted its designers’ true intent.

What happens when artificially intelligent robots misbehave, as the drone did here? The question is not just hypothetical. As robotics and artificial intelligence (AI) systems increasingly integrate into our society, they will do bad things. Sometimes they will cause harm because of a design or implementation defect: we should have programmed the self-driving car to recognize a graffiti-covered stop sign but failed to do so. Sometimes they will cause harm because it is an unavoidable byproduct of the intended operation of the machine. Cars, for example, kill thousands of people every year, sometimes unavoidably. Self-driving cars will too. Sometimes the accident will be caused by an internal logic all of its own—one that we can understand but that still doesn’t sit well with us. Sometimes they will do the things we ask them to (minimize recidivism, for instance) but in ways we don’t like (such as racial profiling). And sometimes, as with our drone, robots will do unexpected things for reasons that doubtless have their own logic but which we either can’t understand or predict.

These new technologies present a number of interesting substantive law questions, from predictability, to transparency, to liability for high stakes decision making in complex computational systems. A growing body of scholarship is beginning to address these types of questions.5 Our focus here is different. We seek to explore what remedies the law can and should provide once a robot has caused harm.

---

The law of remedies is trans-substantive. Where substantive law defines who wins legal disputes, remedies law asks, “What do I get when I win?” Remedies are sometimes designed to make plaintiffs whole by restoring them to the condition they would have been in “but for” the wrong. But they can also contain elements of moral judgment, punishment, and deterrence. For instance, the law will often act to deprive a defendant of its gains even if the result is a windfall to the plaintiff, because we think it is unfair to let defendants keep those gains. In other instances, the law may order defendants to do (or stop doing) something unlawful or harmful.

Each of these goals of remedies law, however, runs into difficulties when the bad actor in question is neither a person nor a corporation but a robot. We might order a robot—or, more realistically, the designer or owner of the robot—to pay for the damage it causes. (Though, as we will see, even that presents some surprisingly thorny problems.) But it turns out to be much harder for a judge to “order” a robot, rather than a human, to engage in or refrain from certain conduct. Robots can’t directly obey court orders not written in computer code. And bridging the translation gap between natural language and code is often harder than we might expect. This is particularly true of modern AI techniques that empower machines to learn and modify their decision making over time, as the drone in the opening example did. If we don’t know how the robot “thinks,” we won’t know how to tell it to behave in a way likely to cause it to do what we actually want it to do.

---


6 More on this below.
One way to avoid these problems may be to move responsibility up the chain of command from a robot to its human or corporate masters—either the designers of the system or the owners who deploy it. But that too is easier said than done. Robot decision making is increasingly likely to be based on algorithms of staggering complexity and obscurity. The developers—and certainly the users—of those algorithms won’t necessarily be able to deterministically control the outputs of their robots. To complicate matters further, some systems—including many self-driving cars—distribute responsibility for their robots between both designers and downstream operators. For systems of this kind, it has already proven extremely difficult to allocate responsibility when accidents inevitably occur.7

Moreover, if the ultimate goal of a legal remedy is to encourage good behavior or discourage bad behavior, punishing owners or designers for the behavior of their robots may not always make sense—if only for the simple reason that their owners didn’t act wrongfully in any meaningful way. The same problem affects injunctive relief. Courts are used to ordering people and companies to do (or stop doing) certain things, with a penalty of contempt of court for noncompliance. But ordering a robot to abstain from certain behavior won’t be trivial in many cases. And ordering it to take affirmative acts may prove even more problematic.

In this paper, we begin to think about how we might design a system of remedies for robots. It may, for example, make sense to focus less of our doctrinal attention on moral guilt and more of it on no-fault liability systems (or at least ones that define fault differently) to compensate plaintiffs. But addressing payments for injury solves only part of the problem. Often

7 See infra notes ___ - ___ and accompanying text.
we want to compel defendants to do (or not do) something in order to prevent injury. Injunctions, punitive damages, and even remedies like disgorgement are all aimed, directly or indirectly, at modifying or deterring behavior. But deterring robot misbehavior too is going to look very different than deterring humans. Our existing doctrines often take advantage of “irrational” human behavior like cognitive biases and risk aversion. Courts, for instance, can rely on the fact that most of us don’t want to go to jail, so we tend to avoid conduct that might lead to that result. But robots will be deterred only to the extent that their algorithms are modified to include sanctions as part of the risk-reward calculus. These limitations may even require us to institute a “robot death penalty” as a sort of specific deterrence against certain bad behaviors. Today, speculation of this sort may sound far-fetched. But the field already includes examples of misbehaving robots being taken offline permanently\(^8\)—a trend which only appears likely to increase in the years ahead.

Finally, remedies law also has an expressive component that will be complicated by robots. We sometimes grant punitive damages—or disgorge ill-gotten gains—to show our displeasure with you. If our goal is just to feel better about ourselves, perhaps we might also punish robots simply for the sake of punishing them. Christina Mulligan half-jokingly suggests that we should have the right to punch a robot.\(^9\) But if our goal is to send a slightly more nuanced signal than that through the threat of punishment, robots will require us to rethink many of our current doctrines.

\(^8\) See infra notes ___ - ___ and accompanying text.

In Part I, we discuss the development of robots and learning AIs, as well as the sorts of robot wrongdoing that will increasingly draw the attention of the legal system. In Part II, we outline the basic principles of remedies law and consider how those remedies will work—or not work—when applied to robots and AIs. Finally, in Part III, we consider how we might remake remedies law with robots in mind.

I. Bad Robots

A. Rise of the Machines

“Robots again.” When Judge Kozinski opened his dissent in *Wendt v. Host International* with this line, he could count on it fetching an ironic grin because it was, well, ironic.\(^{10}\) *Wendt* prominently featured an animatronic version of two television personas,\(^{11}\) much like another case the jurist had overseen some three years prior.\(^{12}\) And in the late 1990s, suits of this sci-fi-esque variety represented such a novelty that the judge’s reference was unmissable. Robots again? Sure. But only because two cases in three years involving robots felt, at the time, like a freak recurrence.

Fast forward just two decades to the present, and Judge Kozinski’s quip appears quaint by comparison. Nowadays, robots are ubiquitous. Industries as far flung as finance, transportation, defense, and healthcare regularly invest billions in the technology. Patent filings

\(^{10}\) *Wendt v. Host Int’l, Inc.*, 197 F.3d 1284 (9th Cir. 1999) (Kozinski, J., dissenting).

\(^{11}\) *Wendt v. Host Int’l, Inc.*, 125 F.3d 806, 809 (9th Cir. 1997).

\(^{12}\) The case referred to here is *White v. Samsung Elec. Am, Inc.*, 971 F.2d 1395 (9th Cir. 1992), cert. denied, 508 U.S. 951 (1993) which involved an animatronic version of Vanna White, a television game show persona.
for robotics and AI applications have skyrocketed. Even octogenarian Senators can be heard fumbling over phrases once confined exclusively to computer science departments, such as “botnet,” “machine learning algorithm,” and “deep neural network.” Robots again, indeed.

Comparing these two moments—separated by just twenty years—puts on full display the field’s breathtaking progress. Today, technological feats that read like pages torn from sci-fi novels have become regular fixtures of the news. Robots have driven millions of miles on U.S. roadways, humbled human professionals at the pinnacle of their fields, and even performed high-stakes surgical procedures on cardiac patients. And as innovators continue to compete against each other in increasingly diverse domains, “robots” themselves are taking on new and expansive forms. Gone are the days of robots confined to assembly lines or warehouse floors. With each passing week, robots infiltrate deeper into our public spaces, places of work, and even bedrooms.


16 See infra notes ___ - ___ and accompanying text.


20 See infra notes ___.
The disruptive forces unleashed by this ascendant technology are challenging long-held assumptions about the limits of machine capabilities—forcing the rest of society to adapt not only economically and politically, but also legally. In the last few years alone, autonomous robots have killed and maimed our fellow citizens, helped determine who goes to prison and who stays there, spouted racist and homophobic remarks on our social media platforms, and even shaped the course of our national elections. Far from anomalous, all signs suggest that these types of events are destined to become the new normal as robots continue to march into the social mainstream in the decades ahead.

In the view of many leading experts, the challenges posed by this impending “robot revolution” could precipitate a jurisprudential revolution of similar magnitude. And though numerous scholars have begun to explore the ramifications robots pose for our substantive legal

---


25 See Charles Duhigg, The Case Against Google, N.Y. TIMES (Feb. 20, 2018), https://www.nytimes.com/2018/02/20/magazine/the-case-against-google.html (noting that prominent lawmakers and critics have accused Google of “creating an automated advertising system so vast and subtle that hardly anyone noticed when Russian saboteurs co-opted it in the last election”).

26 See Andrew Berg et al., Should We Fear the Robot Revolution? (The Correct Answer is Yes), INT’L MONETARY FUND (May 21, 2018) (arguing that global society is on the cusp of a second industrial revolution thanks to advances in robotics and artificial intelligence).

27 See infra note ___ and accompanying notes.
rules, comparatively little attention has been paid to the rules governing remedies. Our goal is to change that. But in order to understand the impact that robots may have on this area of law, it is helpful to first review the technology’s defining characteristics, as well as the ways legal issues will most likely arise.

**A. Defining “Robot”**

Though “robot” has appeared in common parlance for nearly a century, the term is still notoriously resistant to definition. For many outside of computer science circles, it continues to evoke 1950s-era stock images of ironclad humanoids adorned with flashing lights, accompanied by the obligatory monotone voice. More recently, though, “robot” and its derivative “robotics” have come to take on more exacting definitions within broader expert communities.

Among legal scholars, efforts have been made to define robots by their so-called “essential qualities.” Such qualities refer to the fundamental, legally-pertinent “characteristics that distinguish [robots] from prior or constituent technology such as computers or phones.”

---

28 See infra note ___ and accompanying notes.


31 See id. at 514.
One leading scholar, Ryan Calo, argues that robots exhibit at least three “essential qualities”: namely, “embodiment,”32 “emergence,”33 and “valence.”34 In Calo’s telling:

Robotics combines, arguably for the first time, the promiscuity of information with the [embodied] capacity to do physical harm. Robots display increasingly emergent behavior, permitting the technology to accomplish both useful and unfortunate tasks in unexpected ways. And robots, more so than any technology in history, feel to us like social actors—a tendency so strong that soldiers sometimes jeopardize themselves to preserve the "lives" of military robots in the field.35

In light of these qualities, Calo argues that “robots are best thought of as artificial objects or systems that sense, process, and act upon the world to at least some degree.”36 Thus, “a robot in the strongest, fullest sense of the term exists in the world as a corporeal object with the capacity to exert itself physically.”37

As innovation in robotics continues to advance apace, however, the sharp dividing lines of even these recently established “essential qualities” are rapidly blurring. Nowadays, disembodied systems that exist purely as bits and bytes regularly go by the monikers of “bot,” “chatbot,” “crawler bot,” “spam bot,” “social bot,” and so forth. When systems of these types

32 Calo describes “embodiment” as the “capacity to act physically upon the world [and], in turn, to the potential to physically harm people or property.” Calo, Robotics and the Lessons of Cyberlaw supra note __ at 534.
33 Calo describes “emergence” as the ability to “do more than merely repeat instructions but adapt to circumstance.” Calo, Robotics and the Lessons of Cyberlaw supra note __ at 538.
34 Calo describes “social valence” as the heightened emotion response triggered in humans due to our tendency to anthropomorphize them. See Calo, Robotics and the Lessons of Cyberlaw supra note __ at 538.
35 Id. at 515.
36 Id. at 535.
37 Id.
operate in parallel, the collective is often referred to by the ominous title of “botnet.” And when gaming or strategy robots run metaphorical circles around human champions in the likes of Go or DOTA, they do so in entirely ethereal forms with the capacity to exert themselves only digitally.

Thus, unlike some technologies that have stabilized as their commercial and social presence has increased, robots appear to have done the opposite. As Jack Balkin recently observed, a similar phenomenon occurred in the cell phone industry. According to the scholar, “Thirty years ago people might have argued that an essential characteristic of a cell phone was its ability to make a phone call outside of one’s home. . . . But this feature of cell phones is by no means the primary way that people use them today.” So, too, it seems is true of the “essential qualities” of yesteryears’ robots. Already, those that Calo enumerated less than five years read like relics of a bygone era—a testament to the field’s engine of innovation firing on all cylinders.

Today, the terms “robotics” and “artificial intelligence” are often used interchangeably, referring to both embodied and disembodied systems that affect the physical and digital worlds alike. And while there are important technical distinctions to be made between the two

---

38 Cade Metz, *In a Huge Breakthrough, Google’s AI Beats a Top Player at the Game of Go*, WIRED (Jan. 27, 2016), https://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-go/. “Go” is an ancient Eastern strategy game that is comparable to chess, though far more computationally complex. *Id.*

39 Tom Simonite, *Can Bots Outwit Humans In One Of The Biggest Esports Games?*, WIRED (Jun. 25, 2018), https://www.wired.com/story/can-bots-outwit-humans-in-one-of-the-biggest-esports-games/. DOTA is one of the internet’s most popular real time strategy games and is more difficult for AI systems than Go or chess.


41 See *id*.

42 See *supra* note ___ - ___ and accompanying text. See also, e.g., Balkin (stating he does “not think it is helpful to speak in terms of ‘essential qualities’ of a new technology that we can then apply to law”).
concepts, we adopt the convention of construing “robot” to encompass both robots in Calo’s “essentialist” sense and artificially intelligent systems embodied only in software. Our goal is to include any hardware or software system exhibiting intelligent behavior.

1. What Makes Robots Smart?

But what, then, does it mean for a robot to be “intelligent?” Experts operating at the cutting edge of the field describe “artificial intelligence”—in somewhat circular fashion—as the “science of making machines smart.” And though the definition may be wanting for precision, it is this singular feature—the ability to execute complex behaviors such as planning, language processing, or object recognition—that differentiates a robot from a barren hunk of metal, plastic, or bits.

Robots exhibit their “smart[s]” by executing “algorithms.” Although the term has a certain cerebral ring to it, it actually describes a simple concept. Algorithms are merely sequences of instructions for performing a given task. When translated into software, these instructions can be simplified further still. In fact, all commands given to a computational system are reducible

---

43 See Kamal Ahmed, Google’s Demis Hassabis—Misuse of Artificial Intelligence ‘Could do Harm’, BBC NEWS (Sept. 16, 2015), http://www.bbc.com/news/business-34266425. While some scholars have suggested that “there is a continuum between ‘robots’ and ‘artificial intelligence,’” Jack B. Balkin, The Path of Robotics Law, 6 CALIF. L. REV. CIRCUIT 45 (2015), the distinction is actually artificial (if you’ll pardon the expression). Without the ability to exhibit intelligent behavior, any so-called robot would be little more than an inanimate composite of metal, plastic, or bits. Accordingly, AI is better understood as a component feature of any robotics system, rather than an entity separate from it.


45 See id.
to one of three logical operators: AND, OR, and NOT.\textsuperscript{46} If chained together in the right way, these basic operators can produce behaviors of breathtaking complexity. Yet at bottom, even the most sophisticated algorithms are comprised of simple, logic-based building blocks.

For much of AI’s history as a scientific field, the prevailing paradigm of system design involved explicitly encoding the algorithms that governed robots.\textsuperscript{47} This approach—sometimes termed the “classic,” “symbolic,” or “GOFAI” approach (short for “Good Old-Fashioned AI”)—required that scientists or engineers hand-code robot behaviors through “explicit, logical representation of facts about the world.”\textsuperscript{48} The expression \textit{dogs have four legs}, for example, might be represented as:\textsuperscript{49}

\[
\forall x \ (\text{is}_a\_\text{dog}(x) \Rightarrow \text{number}_of\_\text{legs}(x) = 4)
\]

In plain English, this statement translates to: \textit{For every entity, if that entity is a dog, it has four legs.}

The precision and austerity of the GOFAI approach has obvious appeal. Among other features, explicitly encoded algorithms are inherently predictable and explainable. And robots programmed using this approach are still capable of exhibiting astonishingly complex behaviors,

\textsuperscript{46} See \textit{id}.

\textsuperscript{47} See \textit{id}.

\textsuperscript{48} See David Auerbach, \textit{The Programs That Become the Programmers}, SLATE (Sept. 25, 2015), http://www.slate.com/articles/technology/bitwise/2015/09/pedro_domingos_master_algorithm_how_machine_learning_is_reshaping_how_we.html.

\textsuperscript{49} This example derives from David Auerbach’s piece. See \textit{id}. 
ranging from mathematical calculations far surpassing human capabilities, to conquering world chess champions.50

But GOFAI also has its limits. How, for example, is an AI system embedded with a four-legged representation of dogs to categorize the small fraction that do not have four legs, either through accident or genetics? Without prospectively accounting for these types of outliers, hand-coded machines have no means of learning such distinctions on the fly.

In many instances, programmers can teach their robots how to handle these types of “edge cases”51 by prospectively encoding fail-safe measures that anticipate them. But even robust GOFAI approaches that account a wide array of edge cases are often no match for amorphous and ambiguous real-world environments.

Take, for example, the task of navigation. Classically encoded robots have long excelled at getting from point A to point B in warehouses or factories—whether traversing a floor on four wheels or a three-dimensional space with an articulated arm.52 This aptitude owed to the fact that warehouses and factories are, by and large, tightly controlled environments. As such, “programmers could anticipate the range of scenarios a [robot] may encounter, and c[ould] program if-then-else-type decision algorithms accordingly.”53

51 An “edge case” is a technical term that refers to scenarios which occur at the extremes of a given operating parameter—whether expected or unexpected.
53 See id.
On a smooth, clearly demarcated surface with little chance of encountering obstacles (much less inclement weather) the number of uncertainties and edge cases presented was reduced to manageable proportions. But translating a similar navigation task to a bustling city street was another matter entirely. Because the number of uncertainties a robot might encounter in most uncontrolled environments approaches infinity, navigating using a GOFAI approach requires a commensurate number of *a priori* if-then-else statements. Hand-coded algorithms, in other words, simply do not scale.

In the AI field’s earliest years, this inherent limitation of GOFAI—what Pedro Domingo terms the “knowledge acquisition bottleneck”\(^{54}\)—went largely unnoticed. At the time, microprocessing technology itself was in its infancy, meaning that roboticists typically curbed their enthusiasm.\(^{55}\) But as Moore’s law took hold, and computer scientists began expanding their ambitions, the shortcomings of GOFAI became increasingly apparent.

By the 1980s, it was clear a new approach would be necessary to move the field forward.\(^{56}\) But for decades, none appeared—leading to a painful period of stagnation that came to be known as the “AI Winter.”\(^{57}\) Thanks to recent breakthroughs in an innovative approach known as “machine learning,” however, the AI winter is emphatically over.\(^{58}\)

---

\(^{54}\) See DOMINGO, *supra* note __.

\(^{55}\) See Auerbach, *supra* note __ (noting a “lack of computational processing ability” limited AI’s potential in its early years).


\(^{57}\) See id.

\(^{58}\) See RAY KURZWEIL, THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY (Viking Press, 2006) (writing "the AI winter is long since over").
2. How Do Machines Learn?

Machine learning refers to a subfield of AI that turns the GOFAI approach to algorithmic design on its head. Rather than laying out a specific set of instructions for the robot follow, engineers instead specify a goal or set of goals for the robot to achieve when tackling a given problem, often referred to as an “optimizing function.” Having established the desired goal, the robot is then left to author its own algorithms for achieving it, which it does by practicing on illustrative examples of the problem at hand.

At the outset, the robot usually just flails around in the dark—trying things essentially at random without a good idea of what will or won’t work. But each time its experimental efforts move it closer to the goal specified by its designers, the robot receives positive feedback and uses statistical techniques to improve its algorithms accordingly.\(^\text{59}\) Thus, instead of repeatedly executing an unchanging set of instructions, machine learning approaches enable robots to iteratively write their own instructions as they go.\(^\text{60}\) And if given enough examples to train on, these systems can prove remarkably adept at solving staggeringly complex tasks that admit of no obvious GOFAI solutions.

Therein lies the promise of machine learning. In situations where the endless fine-tuning of algorithmic instructions would be impossible to do by hand, machines themselves are able to successfully navigate the “knowledge acquisition bottleneck.”\(^\text{61}\) The program, thus, becomes the

\(^{59}\) And when it performs poorly, vice versa.

\(^{60}\) See DOMINGO, supra note __.

\(^{61}\) See id.
programmer—obviating the need for engineers to anticipate a near-infinite number of edge cases.

When embedded in a broader software or hardware application, the possibilities created by this powerful approach are seemingly endless. Indeed, many leading experts now view machine learning as one among a rarified number of “general purpose technologies” (GPTs), the likes of which include the modern engine, the internet, and electricity.62 Such technologies are distinguished by their ability to “significantly enhance productivity or quality across a wide number of fields or sectors.”63 Paul David’s canonical study established three criteria of GPTs that machine learning appears to possess in abundance: “they have pervasive application across many sectors; they spawn further innovation in application sectors, and they themselves are rapidly improving.”64

Today, companies as diverse as Walmart, Facebook, and General Motors are adopting machine learning systems at “unprecedented rates . . . due to their ability to radically improve data-driven decision making at a cost and scale incomparable to that of humans.”65 It is this engineering approach that allows autonomous vehicles, self-flying drones, and warehouse


64 See id. (citing Paul A. David, The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox, 80 AM. ECON. REV. 355, 355–361 (1990)).

“fetching” robots⁶⁶ to function with seeming ease in unimaginably complex environments. And beyond these robots of the more “essentialist” variety, machine learning also powers a vast array of entities classified as “cyber-physical systems” (e.g. Internet of Things devices), as well as disembodied digital systems often classified as software “bots.”⁶⁷

B. When Robots Do Harm

Machine learning is not without its limitations, however. By breaking from the GOFAI paradigm, robots powered by this technique must also embrace a higher degree of uncertainty than their classically-encoded counterparts. Because machines share in the task of writing their algorithms, using machine learning requires sacrificing some degree of fine-grained control over a machine’s algorithms. Accordingly, designers seeking to implement this powerful approach also understand that it can produce robots which are difficult to predict, tricky to debug, and hard or even impossible to understand.⁶⁸

For many years, this engineering reality limited the most successful machine learning applications to domains with high degrees of fault tolerance. After all, it is one thing for a song recommendation engine to miss its mark 20% of the time. But it is quite another for an autonomous vehicle’s LIDAR system to miss oncoming vehicles at a similar clip.

---


⁶⁸ See DOMINGO, supra note __.
In the last decade, however, advances in the field have enabled engineers to dramatically improve the accuracy, predictability, and performance of numerous machine learning applications—thus, enabling them to entrust robots with positions of greater decision making authority than ever before. It is these advances that have allowed for the introduction of high-stakes robotics systems including self-driving cars, medical diagnostic robots, and even experimental autonomous passenger drones. Yet, even the most performant of these systems remains imperfect—much like the human decision makers they seek to emulate.

Accepting imperfection means also accepting the possibility that robotics systems will sometimes cause harm to others. Indeed, robots acting in harmful, occasionally catastrophic, ways are already a regular fixture of modern life. Robotic cars, aircraft, and manufacturing systems have killed and maimed third-parties; robots tasked with making online purchases have

---


been “arrested” for illicitly buying narcotics on the dark web; and robots powering our largest social media platforms have even influenced the course of national elections.

How is the legal system to remedy the harms caused by these, and countless other, robots? The easy cases will be those involving identifiable defendants who deliberately use robots against others. But all signs suggests that such cases will be the exception, not the rule. Far more often, robots comprised of complex amalgamations of software and hardware, designed by vast numbers of contributors, operating along diffuse causal chains, and executing algorithms that range from enigmatic to outright inscrutable will take actions that hurt others. In such instances, it may be hard to tell who owns the robot, who operates it, who trained it, whether it operated as intended, whether the harm could have been avoided, and perhaps even who the victims are. It is these types of scenarios that will pose the greatest challenges for remedies law in the years and decades ahead.

The following sections survey some of the harms complex robotics systems are likeliest to cause, providing contemporary examples of each.

1. Unavoidable Harms

Many robots operating free from software bugs, hardware errors, or failures of engineering precaution will nevertheless harm others. Some dangers, after all, are inherent to a

---


product or service. In such instances, calling for the total elimination of the danger is tantamount to calling for a prohibition on a product or service itself.

Harms of this variety are often referred to as “unavoidable harms” or, in some tort circles, as “comment k harms.” Conceptually, the notion of such harms tends to evoke products such as cigarettes, pharmaceuticals, alcohol, or knives. But as Robert Peterson notes, virtually no product or service is perfectly “safe,” whether it is a peanut butter jar or a tea cozy—much less a complex robotics application.

An illustrative example of the types of unavoidable harms that robots will cause can be found in the autonomous vehicle (AV) context. Ever since the transition from the horse drawn buggy to the modern automobile, vehicular transportation has entailed error-prone humans, strapped to hulking masses of steel, navigating highly complex environments at highly dangerous speeds. Accordingly, “[f]or more than a century, safety professionals have begun with the assumption that cars would crash, and focused their efforts on reducing the damage.” Experts too numerous to list have convincingly argued that this same assumption will also hold for cars

76 See Welge v. Planters Lifesavers Co., 17 F.3d 209 (7th Cir. 1994).
77 See Peterson, supra note __.
driven by robots as opposed to humans.\textsuperscript{79} For even superhumanly safe self-driving systems are subject to the laws of physics. And if AVs driven by such systems unexpectedly encounter an individual or object without sufficient time or distance to prevent a collision, harm of some variety may be unavoidable.\textsuperscript{80}

2. Deliberate Least-Cost Harms

A close relative of the “unavoidable harms” detailed above involves “deliberate least-cost harms.” These harms are similar to unavoidable ones insofar as they are foreseeable by designers and, in some sense, cannot be avoided. But unlike their entirely unavoidable counterparts, deliberate least-cost harms fall into a grey area where there is sufficient forewarning to meaningfully react to an impending harmful event, but no way to avoid the harm entirely. The question, thus, becomes one of triage: Which of the harmful outcomes is the least costly?\textsuperscript{81}

\textsuperscript{79} Today’s human-driven car accidents can cause unavoidable injuries to drivers, passengers, bystanders, and property. But there is an important difference between contemporary cars and the robocars of the future. Injury from a car crash today is typically the result either of the design of the car or, far more commonly, the behavior of the humans. The law distinguishes those two types of harm, holding manufacturers responsible for injuries caused by product design and human drivers responsible for the injuries they cause.\textsuperscript{79} But self-driving cars, as the name implies, drive themselves. The “design” of the product, in other words, is also responsible for its behavior on the road.

\textsuperscript{80} See, e.g., Noah Goodall, Ethical Decision Making During Automated Vehicle Crashes, J. OF THE TRANSP. RESEARCH BD. (Dec. 2014), doi:10.3141/2424-07 (noting that “[w]hile any engineering system can fail, it is important to distinguish that, for automated vehicles, even a perfectly-functioning system cannot avoid every collision”).

\textsuperscript{81} Not in strictly monetary terms.
This type of lesser-of-evils dilemma, where injury is both inevitable and variable, was canonized by the philosopher Judith Thomson in a thought experiment known as the “trolley problem.” In its most popular formulation, the trolley problem proceeds as follows:

[A]n observer is witness to a runaway trolley car barreling toward five unwitting workers on the tracks ahead. The observer, however, is standing at a switch. If pulled, it will divert the trolley onto another track where only one unlucky worker awaits. Tragedy, of some kind, is foreordained. But the observer holds the proverbial power to steer fate: Turn the trolley, killing the one or refrain from turning the trolley, killing the five?

Ever since the introduction of experimental AVs to U.S. roadways, scenarios involving killer robocars thrust into trolley problem-like dilemmas have captured the public and academic imagination. But situations of this kind will likely be the exception, not the rule, when it comes to deliberate least-cost harms. Far likelier, albeit subtler, scenarios involving least-cost harms

---


83 Thomson’s original experiment asked subjects to imagine themselves as the trolley driver rather than as an outside observer at a switch.


85 See, e.g., WENDELL WALLACH & COLIN ALLEN, MORAL MACHINES: TEACHING ROBOTS RIGHT FROM WRONG, 16 (Oxford U. Press 2009); Joel Achenbach, Driverless Cars Are Colliding With the Creepy Trolley Problem, WASH. POST (Dec. 29, 2015) (arguing “we’re suddenly in a world in which autonomous machines, including self-driving cars, have to be programmed to deal with Trolley Problem-like emergencies in which lives hang in the balance); John Markoff, Should Your Driverless Car Hit a Pedestrian to Save Your Life, N.Y. TIMES (June 23, 2016), https://www.nytimes.com/2016/06/24/technology/should-your-driverless-car-hit-a-pedestrian-to-save-your-life.html (discussing the “dilemma of robotic morality” and its implications for engineers designing robotic decision making systems); Matt Simon, To Make Us All Safer, Robocars Will Sometimes Have to Kill, WIRED (Mar. 13, 2017) (The “trolley problem . . . illustrates a strange truth: Not only will robocars fail to completely eliminate traffic deaths, but on very, very rare occasions, they’ll be choosing who to sacrifice—all to make the roads of tomorrow a far safer place.”).

86 One curious approach is to ignore the problem altogether. German law simply forbids consideration of the trolley problem in programming AVs, saying that an AV headed for an accident cannot alter its
will involve robots that make decisions with seemingly trivial implications at an individual level, but which result in non-trivial impacts at scale.\textsuperscript{87}

Self-driving cars, for example, will rarely face a stark choice between killing a child or killing two elderly people. But thousands of times a day, they will have to choose precisely where to change lanes, how closely to trail another vehicle, when to accelerate on a freeway on-ramp, and so forth. Each of these decisions will entail some probability of injuring someone. And making the “right” decision will require weighing the probability of causing harm, exploring what alternatives exist, and specifying how the car should value the different types of harms that will foreseeably impact different stakeholders.

Consider, for example, the seemingly trivial engineering choice of how much buffer to provide a cyclist. Suppose that a vehicle were programmed to give an extra inch or two of room to any cyclists it passed, out of an abundance of caution. From any single cyclist’s perspective the change would be scarcely perceptible. The vehicle would overtake them at a distance that appeared identical to any other self-driving cars programmed to not provide the extra distance.

But if that same design choice scaled to an entire fleet of vehicles that regularly encountered hundreds of thousands—or even millions—of cyclists, even a difference of such miniscule proportions could be expected to impact cyclist collision rates. Given this reality, providing the additional buffer room may seem like a no-brainer.

\textsuperscript{87} See Casey, \textit{Amoral Machines, supra} note ___ (discussing how minute differences in how individual vehicles operate could have profoundly consequential macroscopic effects).

\textsuperscript{87} See Dave Gershgorn, \textit{Germany's Self-Driving Car Ethicists: “All Lives Matter,” Quartz}, Aug. 24, 2017. That does leave open the question of what the AV should do in an unavoidable accident situation, though. “Nothing” may often be the worst response.
But not so fast. The same inch that benefits the cyclists might have the opposite effect for the vehicles’ passengers, putting them at a marginally higher risk of head-on collisions due to the vehicles’ position closer to the center of the roadway. Once again, the seemingly infinitesimal uptick in risk implicated by such a decision would be all but imperceptible during the course of any individual journey. But when viewed at scale, decisions of this kind will carry a profound ethical and legal weight—requiring designers to grapple with complex, highly fraught tradeoffs inherent to deliberate least-cost harms.

3. **Defect-Driven Harms**

One of the more obvious ways robots will cause harm is through traditional hardware or software “defects.” Harms of this variety occur when a software bug, hardware failure, or insufficient level of precaution by designers causes a robot to injure others. For much of the field’s history, these types of defect-driven harms have been relatively easy to define and identify. They typically occur when designers intend a robot to work in a certain way, but screw up, causing it to behave differently, as was recently alleged in a case involving a robot that “escaped” from its section of a trailer hitch assembly plant, “entered [a technicians] work area, surprise[ed] her, and crushed her head between hitch assemblies.”

As robots continue take on increasingly sophisticated forms, however, defining and identifying these types of “defects” will likely become more challenging. Is a self-driving car to be

---

88 This piece does not discuss substantive tort law distinctions found in modern tort doctrine.

deemed “defective” if it brakes more slowly than a human driver? What if it brakes faster than humans, but not as fast as other self-driving cars? Or as fast as other self-driving cars, but not as fast as it might possibly brake if reprogrammed?

Additional legal wrinkles involving defect-driven harms will also arise in systems involving “humans-in-the-loop,” where responsibility for controlling a robot is distributed between algorithmic and human decision makers. A boundary-pushing example of this phenomenon recently occurred in Tempe, Arizona, when a self-driving car deployed by Uber fatally struck a pedestrian. Although the vehicle was capable of autonomy under certain design parameters, it also relied on a backup driver to take control in the event of an emergency. Yet one night, when a pedestrian unexpectedly walked out in front of one such vehicle, neither the backup driver nor the self-driving system took steps to avoid the collision. As a result, the vehicle collided with the pedestrian at speeds in excess of 30mph without breaking or swerving. Should the backup driver be held responsible for failing to take over? Or was it unreasonable for Uber to put the operator in such a position to begin with? Does it matter how the car was programmed?

---


92 See id.


94 See id.
How the legal system will eventually resolve controversies involving these types of “moral crumple zones”\textsuperscript{95} remains an open question, even among experts.\textsuperscript{96} But none question the reality that robots exhibiting increasingly complex design defects will continue to harm individuals for the foreseeable future.

4. Misuse Harms

Sometimes, people will misuse robots in a manner that is neither negligent nor criminal but nevertheless threatens to harm others. Given the unpredictable nature of machine learning systems, and the nearly infinite variety of ways humans can interact with modern robotics applications, these types of harms are particularly difficult to guard against. Already, media reports are rife with examples of individuals attempting to manipulate robot behaviors, deceive or “trick” robot perception systems, probe robots for safety or security vulnerabilities, or deploy robots in ways that adversely impact others.\textsuperscript{97} Whether such forms of meddling are deemed to have been preventable by manufacturers, or to have fallen within the scope of the robot’s intended design, will have significant implications for the substantive legal doctrines that will govern the ultimate outcomes and for who bears the resulting liability.


\textsuperscript{96} Whether these types of questions will, ultimately, be resolved under the umbrella of negligence, breach of warranty, enterprise liability, or traditional product “defect” remains unclear.

A now infamous example of robot misuse comes from Microsoft’s Twitter chatbot, “Tay.” Unlike chatbots designed to maintain a static internal state upon deployment, Tay’s system updated itself in real time by learning from interactions with users. Within hours of going live, however, hundreds of Twitter users began intentionally tweeting “misogynistic, racist, and Donald Trumpist remarks” at the robot. Thanks to this barrage of unforeseen misuse, “Tay rapidly morphed from a fun-loving bot . . . into an AI monster.” Tay lasted a mere sixteen hours on the platform before Microsoft intervened. After initially declining to comment, the company eventually noted that a “coordinated effort by some [Twitter] users to abuse Tay’s commenting skills” led it to shut the robot down.

One notable feature of the Tay example is that Microsoft itself did not engage in misuse. Nor is there any reason to think that Tay’s design was defective. Rather, the robot’s rogue conduct resulted from the input of third-parties. But owners, too, will misuse robots, or at least use them in ways we may not expect. Drone owners, for example, might use them to spy on neighbors or invade their privacy. Similarly, self-driving car owners might modify their vehicles to protect occupants at all costs, even if doing so imposes greater risks on bystanders. And

---


99 But we repeat ourselves.

100 Misconduct from the perspective of Microsoft, at least.

predictive learning algorithms that might decide everything from the cost of your life insurance to where you end up in an emergency room queue to whether you are granted parole are all dependent on the training data they are fed. And that training is only as good as the (often imperfect) data users feed the robot.\textsuperscript{102}

\section{5. Unforeseen Harms}

Many harms attributable to robots will be neither defect-driven, unavoidable, nor the result of misuse, but will simply be unforeseen by those who designed them.\textsuperscript{103} Harms of this variety are by no means unique to the field of robotics. Indeed, unpredictability is part and parcel of any sufficiently complex system. It’s why your computer periodically crashes\textsuperscript{104} and perhaps why new typos seem to pop up in our writing even though we’ve read through a draft at least 30 times.\textsuperscript{105}

But if the last decade of progress in the field of robotics has taught us anything, it is that robotics systems using machine learning techniques can be extremely hard to predict, rendering them particularly susceptible to causing unforeseen harms. This phenomenon owes, in large part, to the fact that machine learning systems “enter[] into a social world already in motion, with an existing set of assumptions and expectations about what is likely and unlikely, possible

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{102} We discuss this problem in more detail \textit{infra} notes \textsuperscript{\_\_\_\_} and accompanying text.
\item \textsuperscript{103} Either because of resource constraints involving safety testing or because they were genuinely unforeseeable.
\item \textsuperscript{104} See Clay Shield, \textit{Why do computers crash?}, \textsc{Sci. Amer.} (Jan. 6, 2003), https://www.sciencemag.org/science/acsfulltext/81/3754/263.full.pdf.
\item \textsuperscript{105} OK, maybe not that last one.
\end{itemize}
\end{footnotesize}
and impossible.” Yet because such systems are, by definition, empowered to learn with limited direct human intervention, the behaviors that they develop can also be unconstrained by the norms, assumptions, and expectations that implicitly govern humans.

Sometimes, this lack of constraint can lead to astonishing, utterly unintuitive results. Robots deployed using machine learning techniques, for example, have devised wholly new tactics for conquering strategy games, have inadvertently set off wars of proliferation with bots on online platforms (leading to bizarre pricing decisions), and have even invented “codewords” to communicate with other AI systems that were indecipherable by their designers. Because of this unpredictability, many complex robots will carry an enormous range of unforeseeable risks—even where numerous precautions are taken in advance of deployment.

106 See Balkin, The Path of Robotics Law, supra note __.
107 See Domingo, supra note __.
108 See Balkin, The Path of Robotics Law, supra note __.
109 See David Silver et al., Mastering the Game of Go Without Human Knowledge, 555 NATURE 354 (Oct. 19, 2017); Nicola Twilley, Artificial Intelligence Goes to the Arcade, NEW YORKER (Feb. 25, 2015), https://www.newyorker.com/tech/elements/deepmind-artificial-intelligence-video-games (writing that “without any human coaching, [an AI system designed to play arcade games] not only bec[a]me better than any human player but [] also discovered a way to win that its creator never imagined”).
To be clear, the unpredictability inherent to machine learning is also one of its greatest strengths. An AI that just engages in rote calculation of equations we already know the answer to might get to the result faster than humans can, but it won’t be any better at understanding or predicting outcomes than humans. We want AIs to do unpredictable things, so long as those things lead to good results. If an AI can reliably conclude that butterfly population variance in Tibet affects the weather in Indonesia, it will be better than humans at predicting the weather. And if a self-driving car can conclude from subtle changes in the velocity of the cars surrounding it that a crash is imminent, it offers greater hope of avoiding such crashes than a human driver might.\footnote{112 See Rob Ludacer, \textit{Watch a Tesla Predict an Accident and React Before It Even Happens}, \textit{BUS. INSIDER} (Dec. 29, 2016), \url{http://www.businessinsider.com/tesla-avoids-accident-before-happens-2016-12} (showing a video of Tesla’s Autopilot doing just that).}

But the unpredictability of the path that robots will take to their goals means that they may do things that make perfect sense given what they were asked to maximize, but which turn out to reflect either poorly specified goals or flawed training data. The introduction’s example of a drone learning to intentionally sabotage its flight path provides just one of the now countless documented instances of unforeseen robot behaviors. Another comes from the healthcare domain.

In the 1990’s, a pioneering multi-institutional study sought to use machine learning techniques to predict health-related risks prior to hospitalization.\footnote{113 B. Buchanan et al., \textit{An Evaluation of Machine-learning Methods for Predicting Pneumonia Mortality}, \textit{9 ARTIFICIAL INTELLIGENCE IN MED.}, 107 (1997); R. Ambrosino et al, \textit{The Use of Misclassification Costs to Learn Rule-Based Decision Support Models for Cost-Effective Hospital Admission Strategies} in \textit{PROCEEDINGS OF THE ANNUAL SYMP. ON COMP. APPLICATION IN MEDICAL CARE} (1995).} After ingesting an enormous quantity of data covering patients with pneumonia, the system learned the rule:
has_asthma(x) ⇒ lower_risk(x)

The colloquial translation being “that patients with pneumonia who have a history of asthma have lower risk of dying from pneumonia than the general population.”

The machine-derived rule was curious, to say the least. Far from being protective, asthma can seriously complicate pulmonary illnesses, including pneumonia. Perplexed by this counterintuitive result, the researchers dug deeper. And what they found was troubling.

They discovered that “patients with a history of asthma who presented with pneumonia usually [had been] admitted not only to the hospital but directly to the Intensive Care Unit (ICU).” Once in the ICU, asthmatic pneumonia patients went on to receive more aggressive care, thereby raising their survival rates compared to the general population.

The rule, in other words, reflected a genuine pattern in data. But the machine had confused correlation with causation—"incorrectly learn[ing] that asthma lowers risk, when in fact asthmatics have much higher risk." Thankfully, the relative simplicity of the machine learning model deployed by the researchers in this instance allowed them to detect, reverse engineer, and remedy the situation

---


115 See Buchanan et al., *supra* note __.

116 *Id.*

117 *Id.*
before any harmful behavior resulted.\textsuperscript{118} Indeed, the algorithm taught humans something about the flaws in existing care techniques. But that is a luxury which will not be afforded to all robot designers.\textsuperscript{119} Indeed, as Marc Canellas et al. have convincingly argued, the likelihood of these types of unpredictable events actually tends to rise alongside the complexity of computational models, even though the overall likelihood of an abnormal event may remain constant.\textsuperscript{120} This phenomenon owes to the highly leptokurkic\textsuperscript{121} failure curves often observed in complex systems, where a “reduced likelihood of failure in a general sense” tends to be accompanied by “an increase[d] likelihood of more severe failures.”\textsuperscript{122}

\textsuperscript{118} See Caruana, supra note __.

\textsuperscript{119} IMB’s Watson, for example, was recently reported as displaying “multiple examples of unsafe and incorrect treatment recommendations.” Jennings Brown, IBM Watson Reportedly Recommended Cancer Treatments That Were ‘Unsafe and Incorrect’, Gizmodo (Jul. 25, 2018), https://gizmodo.com/ibm-watson-reportedly-recommended-cancer-treatments-tha-1827868882.


\textsuperscript{121} Leptokurkic distributions show higher peaks around mean values and higher densities of values at the tail ends of the probability curve.

\textsuperscript{122} Canellas et al., supra note __ at 41.
6. Systemic Harms

People have long assumed that robots are inherently “neutral” and “objective,” given that robots simply intake data and systematically output results. But they are actually neither. Robots are only as “neutral” as the data they are fed and only as “objective” as the design choices of those who create them. When either bias or subjectivity infiltrates a system’s inputs or design choices, it is inevitably reflected in the system’s outputs. Accordingly, those responsible for overseeing the deployment of robots must anticipate the possibility that algorithmically biased applications will cause harms of this systemic nature to third-parties.

Robots trained on poorly curated data sets, for example, run the risk of simply perpetuating existing biases by continuing to favor historical *haves* against *have-nots*. In such

---


instances, different outcome distributions in the data reflecting racial, ethnic, social, or economic disparities can become self-fulfilling prophecies—leaving already marginalized groups at the mercy of past injustices.

Similarly, the algorithmic goals and sub-goals that define robot behavior can also lead to biased results. After all, each decision in the process of developing an algorithm necessarily reflects the values of its designers. And when designers fail to consider particular stakeholders, or fail to specify goals that accurately map onto their desired outcomes, their robots may unfairly privilege certain individuals or groups over others. Hence, Cathy O’Neil’s provocative description of an algorithm as an “opinion embedded in mathematics.”125

Instances of bias or subjectivity infiltrating robotics systems are already well documented. A recent example comes from the car insurance industry. U.S. law obliges all car owners to purchase insurance for their vehicles. But not all premiums are created equal. A recent study by Consumer Reports found that contemporary premiums depended “less on driving habits and increasingly on socioeconomic factors,” including an individual’s credit scores.126 After analyzing “2 billion car insurance price quotes from more than 700 companies,” the study found that “[c]redit scores . . . factored into [insurance] algorithms so heavily that perfect drivers with low credit scores often paid substantially more than terrible drivers with high scores.” The study’s findings raised widespread concerns that AI systems used to generate these quotes could “create

125 See id.

126 A credit score “summarizes an individual’s credit history and financial activities in a way that informs the bank about their creditworthiness.” See Lydia T. Liu et al., supra note __.
negative feedback loops that are hard to break.” According to one expert, “Higher insurance prices for low-income people can translate to higher debt and plummeting credit scores, which can mean reduced job prospects, which allows debt to pile up, credit scores to sink lower, and insurance rates to increase in a vicious cycle.” Similar examples of robotics systems causing, or threatening to cause, systemic harms have been documented in the domains of predicting policing, criminal sentencing, targeting advertising, search optimization, and facial recognition, among many others.

To be sure, all advantages are comparative. AI may replicate bias in existing legal systems. But it also has the potential to reduce that bias by replacing human instinct with actual metrics. But it is important that those new objective measures don’t simply replicate the problems of their subjective predecessors.

---


128 See id. (quoting Cathy O’Neil, a data scientist and author of Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy).


7. Collateral Harms

As we continue to invite robots into our homes, personal lives, and places of work, the types of collateral risks they pose to our privacy, security, environment, and even livelihoods will also grow in kind. Some harms, after all, simply arise as a byproduct of pervasiveness. And the threat of these types of harms emerging will be especially true in modern robots, given that they often combine the uncertainties of machine learning, the “promiscuity of data,” the inherent security risks of computational systems, and the threat of physically affecting the real world.

Take, for example, the now commonplace phenomenon of inviting “Internet of Things” (IoT) devices, such as an Amazon Echo or Google Home, into our homes to monitor our every utterance. For many (ourselves included), the convenience of simply issuing a voice command to set a cookie timer, play a song, or order a cab can be too good to pass up. Yet, in exchange for the capabilities offered by these powerful voice recognition bots, we must also accept the reality of their 24-7 surveillance of our most intimate settings.

All signs suggest that the invasiveness of robotics applications like these will only increase in the years ahead. As the capabilities and price points of home security bots, robo vacuums, office assistants, and even “robomaids” improve with time, so too will the scope and the granularity of the data they capture. Data collection practices of this magnitude will not only present legal oversight challenges to those tasked with gathering it, but they will also present novel challenges for those seeking to secure robots against external threats. As James

---

131 This phrase comes from the scholar Ryan Calo and refers to the fact that digital information “faces few natural barriers to dissemination.” See Calo, Robotics and the Lessons of Cyberlaw, supra note ___ at 532.

132 Internet of Things refers to the embedding of networked devices in everyday objects, thereby allowing them to gather, send, and receive data.
Grimmelmann notes, “the more opportunities for innovation, the more possible targets for hacking.” Accordingly, the very same applications that now gather unprecedented amounts of data from users are also likely to pose unprecedented risks in the event that such data gets into the wrong hands.

Even if they aren’t hacked, the mere presence of these devices can change human behavior. People act differently when they think they are being watched or listened to, even if the thing doing the watching is only a picture of a pair of eyes taped to the computer. And if a robot is in your house, you’re not just imagining it: it probably is watching and listening to you.

Add to this brave new reality the awesome power of cloud computing and networking technologies, and the threat of collateral harms is only exacerbated. Armies of robots linked through networking technologies will enable single, centralized systems to impact our physical and digital environments in profound new ways. Seemingly microscopic design choices within systems controlling fleets of tens of thousands of autonomous vehicles, for example, could produce macroscopic effects including changes to traffic patterns, transportation pricing, congestion, and even energy grid usage. We may, for example, wake up one morning to discover that Google Maps has routed highway traffic through our quiet neighborhood streets. Such a decision harms people who never use Google Maps or self-driving cars. But so might its opposite.


134 Ryan Calo, People Can Be So Fake: A New Dimension to Privacy and Technology Scholarship, 114 Penn St. L. Rev. (2010); Margot E. Kaminski et al., Averting Robot Eyes, 76 Md. L. Rev. 983 (2017) (noting this problem and offering design principles to minimize it).

Suppose, instead, that the same routing algorithm avoided residential areas entirely, causing greater congestion on highways and interstates than was socially optimal.

Finally, some of the collateral harms presented by robots may not feel like traditional “harms” at all, but will be the unintended economic effects of certain behaviors, including net positive ones. Robots, for example, displace jobs. And though delivery drones, manufacturing robots, and driverless trucks may serve as the usual suspects in this regard, many other less obvious applications pose similar threats. A more accurate parole prediction algorithm, for example, could result in a smaller incarcerated population. As a consequence, cities and towns across rural America may need fewer prison guards and fewer construction workers to build their prisons. In the long term, algorithms like these will almost certainly prove to be a net benefit to society. But their short-term negative consequences may be especially pronounced for discrete segments of society—raising new questions surrounding the law’s role in remediying them.  

II. Remedies and Robots

The injuries we described in the last part will lead to lawsuits of various types. Indeed, they already have. We don’t intend to discuss all the ways courts might apply the substantive

---

136 For discussions of the net effects of robots on jobs, which are likely to be complex, see Jason Furman & Robert Seamans, AI and the Economy (working paper May 29, 2018); Mark A. Lemley, IP in a World Without Scarcity, 90 N.Y.U. L. Rev. 460, 510-15 (2013).

law to those legal harms. There is a growing literature doing just that.\(^\text{138}\) Rather, our focus is on the practical end game of these coming lawsuits: the law of remedies. Having identified a wrong, courts try to make it right by applying various remedies. But as we will see, when the defendant is a robot (or its owner) that can be easier said than done.

**A. The Law of Remedies**

A remedy, broadly defined, is anything that a judicial body can do for an individual who has been harmed or is threatened with harm. Remedies are the means by which substantive law is given its actual effect. Once a plaintiff is adjudged to have suffered harm under the laws governing primary rights and duties, the law must provide a remedy for those rights and duties to have meaning. Without a remedy, lawfulness and unlawfulness are rendered merely nominal distinctions—or, as it is often more pithily phrased, “No right without a remedy.”\(^\text{139}\)

There are two fundamental kinds of remedies: those that are “compensatory” and those that are “preventative.”\(^\text{140}\) Compensatory remedies aspire to address the wrongs suffered by an individual through monetary transfers between plaintiff and defendant, compensating the


\(^{140}\) DOUGLAS LAYCOCK, MODERN AMERICAN REMEDIES 3-7 (Aspen Publishers 4th ed. 2011).
plaintiff for the injury suffered. Preventative remedies, meanwhile, aspire to avoid this transfer entirely. They seek to discourage, avert, or literally undo harm, rather than retrospectively compensating victims once harm has occurred. Some preventative remedies accomplish this aim by threatening lawbreakers with damages, specific performance, or restitution in an effort to deter unlawful conduct. But sometimes courts seek to prevent harm more directly by enjoining individuals from acting or, less commonly, ordering them to take affirmative steps to avoid violating the law.¹⁴¹

One goal of remedies law is to make plaintiffs whole by restoring them to the condition they would have been in “but for” the wrong—what Doug Laycock calls restoring the “plaintiff’s rightful position.”¹⁴² Traditionally, this compensatory goal has focused on the plaintiff in the dispute—presumably a legal person.¹⁴³ Compensation is normally accomplished through the award of legal damages.

¹⁴¹ Id. at 11.
¹⁴² Id. at 11-15.
¹⁴³ It is possible to imagine robot plaintiffs. Robots can certainly be injured by humans. You might run a stop light and hit my self-driving car, for example. Or people might attack a robot. See, e.g., Isobel Hamilton, People Kicking These Food Delivery Robots is an Early Insight Into How Cruel Humans Could be to Robots, BUS. INSIDER (Jun. 9, 2018), https://www.businessinsider.com/people-are-kicking-starship-technologies-food-delivery-robots-2018-6?r=US&IR=T (the headline says it all); Russ Mitchell, Humans Slapped and Shouted at Robot Cars in Two of Six DMV Crash Reports This Year, L.A. TIMES (Mar. 5, 2018), http://www.latimes.com/business/autos/la-fi-hy-human-attacks-robot-cars-20180305-story.html; Silicon Valley Security Robot Attacked by Drunk Man—Police, BBC NEWS (Apr. 26, 2017), https://www.bbc.com/news/world-us-canada-39725535. The robot itself presumably won’t have a right to sue, at least for the foreseeable future. But the owner of the robot might sue for damages. That doesn’t seem to present significant remedies issues different from ordinary property damages cases, though. Valuing the loss of an individual robot or AI that has learned in ways that differ from factory settings may present difficulties akin to the valuation of any unique asset. But that’s likely to be rare, since people will presumably back up their unique AIs periodically.
But remedies law also focuses substantial attention on defendants. Equitable restitutionary remedies such as unjust enrichment, disgorgement, and constructive trust are designed not to compensate plaintiffs but to deprive defendants of the benefit of wrongful acts. These remedies are designed not to make the plaintiff whole, but to make the defendant “whole” (in the sense that he is no better off than he would have been but for the wrongdoing).

Injunctive relief can serve the purpose of putting either the plaintiff or the defendant in their rightful position. Injunctions order the defendant not to act (or, less commonly, to take some affirmative act). Generally, injunctions are designed to prevent a future harm or stop an ongoing one. But they can also aim to make affirmative changes in the world, by seeking to change existing structures that have led to past injuries.\footnote{Courts do this when they order structural reforms to prisons, hospitals, or schools, for instance. See, e.g., Missouri v. Jenkins, 515 U.S. 70 (1995); Hutto v. Finney, 437 U.S. 678 (1978).}

Remedies law also contains many elements of moral judgment, punishment, and deterrence. For instance, the law will often act to deprive the defendant of gains, even if the result is a windfall to the plaintiff, because we think it is unfair to let the defendant keep those gains. Courts may also enhance damages beyond what is necessary to compensate plaintiffs or deprive defendants of profits in order to punish behaviors we deem reprehensible.

Most of these non-compensatory remedies laws were explicitly designed to change the behavior of people. But the remedial mechanisms used to shape human behavior cannot be relied upon to do the same when machines, not people, engage in harmful conduct. The remainder of this Part considers some of the complications that robots bring to various remedies rules.
B. The Nature of Remedies

1. Normative Versus Economic Perspectives

The choice of remedy for a given legal violation often stems from fundamental assumptions regarding the nature of the substantive law itself. Two views predominate. A “normative” view of substantive law sees it as a prohibition against certain conduct, with the remedy being whatever is prescribed by the law itself. The defendant, on this view, has engaged in a wrongful act that we would stop if we could. But because it is not always possible to do so—commonly because the act has already occurred—remedies law seek to do the next best thing: compensate the plaintiff for the damage done. This view is consistent with laws enforced by property rules.145

An alternative view of substantive law, however, conceptualizes the role of remedies differently. Under this “economic” view, the substantive law alone forbids nothing. Rather, it merely specifies the foreseeable consequences of various choices, with the available remedies simply signaling the particular penalties associated with particular conduct. Damages, on this view, are simply a cost of doing business—one we want defendants to internalize but not

---

necessarily to avoid the conduct altogether.\textsuperscript{146} This approach is more commonly associated with liability rather than property rules.\textsuperscript{147}

To help illustrate the difference between these two views, consider an everyday encounter with a traffic light. Under the normative view, a red light stands as a prohibition against traveling through an intersection, with the remedy being a ticket or fine against those who are caught breaking the prohibition. We would stop you from running the red light if we could. But because policing every intersection in the country would be impossible, we instead punish those we do catch in hopes of deterring others.

Under the economic view, however, an absolute prohibition against running red lights was never the intention. Rather, the red light merely signals a consequence for those who do, in fact, choose to travel through the intersection. As in the first instance, the remedy available is a fine or a ticket. But under this view, the choice of whether or not to violate the law depends on the willingness of the lawbreaker to accept the penalty.

In one of his more arresting turns of phrase, Oliver Wendell Holmes’ famously described the economic view of substantive law as that of a “bad man.” According to Holmes:

\textit{If you want to know the law and nothing else, you must look at it as a bad man, who cares only for the material consequences which such knowledge enables him to predict, not as a good one, who finds his reasons for conduct, whether inside the law or outside of it, in the vaguer sanctions of conscience.}\textsuperscript{148}

\begin{flushleft}
\textsuperscript{146} See Ian Ayres & Eric Talley, \textit{Solomonic Bargaining: Dividing a Legal Entitlement to Facilitate Coasean Trade}, 104 YALE L.J. 1027 (1995); see also Louis Kaplow & Steven Shavell, \textit{Do Liability Rules Facilitate Bargaining? A Reply to Ayres and Talley}, 105 YALE L. J. 221 (1995) (responding the Ayres’s and Talley’s argument that, when bargaining is imperfect, ”liability rules possess an 'information-forcing' quality" that "may induce both more contracting and more efficient contracting than property rules).  \\
\textsuperscript{147} See Calabrese & Melamed, \textit{supra} note __.  \\
\textsuperscript{148} Oliver Wendell Holmes, Jr., \textit{The Path of the Law}, 10 HARV. L. REV. 457, 458-459 (1897).  \\
\end{flushleft}
The measure of the substantive law, in other words, is not to be mixed up with moral qualms, but is simply coextensive with its remedy—no more and no less.

While some law and economics scholars accept this precept as fundamental, in many behavioral contexts it does not tell the entire story. Although the actual consequences associated with lawbreaking play a substantial role in much of human decision making, many individuals nonetheless view law as having distinctly normative underpinnings. As Doug Laycock notes, “It is certainly true that some individuals will obey the law only if the consequences of violation are more painful than obedience. . . . [But the fact that] some individuals are unmoved does not eliminate the statement’s moral force for the rest of us.”

An illustrative example of this phenomenon in action comes from the Ohio case, French v. Dwiggins, involving a fatal motorcycle accident. At issue was a recently passed statute expanding the avenues of recovery available to plaintiffs who pursued wrongful death claims. The court wrote that, although the expansion of remedies coincided with the timing of the accident, the defendant “could not be reasonably expected to conduct his affairs any differently” than under the prior regime. The court reasoned that when it came to this life and death matter, the marginal differences in available remedies played no role in the defendant’s decision making leading up to the accident.

\[\text{149 Laycock, Modern American Remedies, supra note } \_\_\_\_ \text{ at 7. See also Yuval Feldman, The Law of Good People: Challenging States’ Availability to Regulate Human Behavior (Cambridge Univ. Press forthcoming 2018) (arguing that we should focus legal rules on the signals they send to good people rather than just constraining the behavior of bad people).}
\]

\[\text{150 French v. Dwiggins, 458 N.E.2d 827.}\]
Holmes, himself, could hardly have been said to disagree with the court’s reasoning.\textsuperscript{151} Despite his provocative use of the “bad man” metaphor to clarify the role of the legal rules for those acting out of pure self-interest, he understood the complex—oftentimes competing—roles that normative concerns play in everyday decision making.

2. **Bad Men and Good Robots**

People are rarely forced to grapple with the distinctions between the normative or economic view of substantive law.\textsuperscript{152} But robots, or at least their programmers, are afforded no such luxury. Sure, robots can be prohibited from engaging in certain types of conduct, assuming their designers understand and control the algorithm by which they make decisions. But implementing a legal remedy via computer code necessarily involves adopting either a normative or economic view of the substantive law.

That’s because a true “prohibition” can only be communicated to a computer system in one of two basic ways: . It can be encoded in the form of an “IF, THEN”\textsuperscript{153} statement that prevents a robot from engaging in particular types of conduct, or it can be coded as a negative weight for

\textsuperscript{151} See, e.g., Marco Jiminez, *Finding the Good in Holmes’s Bad Man*, 79 FORD. L. REV. 2069, 2069 (observing that “a careful reading of Holmes suggests that he was himself well aware of the intimate relationship between law and morality, and seems to have recognized, somewhat surprisingly, that only by engaging in an analytical separation of these two concepts can they then be normatively reunited in an intellectually consistent and satisfying manner”).

\textsuperscript{152} Corporations are more likely to do so. Because we can’t put a corporation itself in jail, corporate compliance—even with penalties designed to stop conduct rather than just internalize costs—might nonetheless be viewed as a cost of doing business for the corporation.

\textsuperscript{153} An IF, THEN statement—or “if-then-else statement”—refers to an expression that conditionally executes a statement or group of statements.
engaging in that same conduct. An IF, THEN statement operates like an injunction, while a weight in a decision-making algorithm operates like a liability rule.

Returning to the example of the red light, a programmer seeking to prohibit a robot from breaking the law could do so with an IF, THEN statement along the lines of: “If the robot encounters a red light, then it will not travel into the intersection.” Similarly, a programmer seeking to achieve that same prohibition in a probabilistic system could do so by assigning an infinitely high negative consequence to traveling into the intersection when the light is red.

An IF, THEN statement is an absolute rule. If a triggering event occurs, then a particular consequence must inexorably follow. As a practical matter, so is an infinitely negative weight. Both achieve the functionally equivalent result of prohibiting the unlawful conduct—the goal of a normative vision of substantive law. But in order to achieve this normative vision, the prohibition must be implemented without regard for the cost of a ticket.

Because the law is encoded as an absolute in its programming, the robot will always obey the law. That’s not true of people. If we want legal rules to be self-executing, the ability to impose perfect obedience may be a good thing.

By contrast, if the underlying theory of a remedy is economic, the machine’s decisionmaking calculus is fundamentally different. Once more, the example of the traffic light helps to clarify this distinction. To an economist, the substantive law and its remedy do not signal a “self-executing refusal never run a red light” but instead an understanding that “running a red light is associated with a small chance of a modest fine and a somewhat increased chance of a

154 No matter how improbable, any non-zero probability multiplied negative infinity returns negative infinity.
traffic accident which will damage the car and may require the payment of damages to another.”

Under this view, the remedy, and its risks, are both expressed in stochastic terms. They translate into probabilistic costs within the robot’s overall decisionmaking calculus. Those costs won’t be infinite, unless perhaps the penalty is death. They will instead reflect a “price” for running a red light that the algorithm might decide to pay depending on what benefits light-running offers.

Thus, under the economic view, the choice of whether or not to obey a law is, of necessity, the choice of a Holmesian “bad man.” Normative views of substantive law—which we know shape certain aspects of human behavior—cannot be expected to translate cleanly into the robotics context with their associated remedies intact. If we want robots to adopt normative views of the law, prohibitions against unlawful conduct will need to be embedded in robots without regard to their economic remedies, requiring outright prohibitions of the type that famously got Asimov’s robots into so much trouble. After all, it’s hard to operate a robot with too many absolute prohibitions. And this will be particularly true of machine learning systems that develop their own algorithms, making it difficult for engineers to reliably predict how encoded prohibitions will interact with other rules.

Encoding the rule “don’t run a red light” as an absolute prohibition, for example, might sometimes conflict with the more compelling goal of “not letting your driver die by being hit by

155 And probably not even then, unless the robot’s algorithm preferences its own survival over most other outcomes (which it probably won’t).

156 ISAAC ASIMOV, THE REST OF THE ROBOTS 43 (1964) (remarking “[t]here was just enough ambiguity in the Three Laws [of robotics found in his works] to provide the conflicts and uncertainties required for new stories, and, to my great relief, it seemed always to be possible to think up a new angle out of the sixty-one words of the Three Laws”).

157 “Don’t become Skynet” does seem like a good one to include, though. See Genetic Algorithms Tip: Always Include This in Your Fitness Function, XKCD https://xkcd.com/534/.
an oncoming truck.” Humans know that “don’t run a red light” doesn’t really mean “don’t ever run a red light.” Rather it translates, roughly, to “don’t run a red light unless you have a sufficiently good reason and it seems safe.” Likewise, even weightier normative prohibitions, such “though shalt not kill,” come with an implied “unless . . .” But designers can’t put that in an IF, THEN statement unless they understand and specify all the exceptions to the rule.

More plausibly, robots operating in the real world will have to adopt algorithmic approaches to almost all complex problems that weigh particular actions against various goals and risks. As a result, the role of remedies in discouraging socially detrimental conduct will need to be reimagined in terms of cost internalization, as opposed to normative sanction or punishment. Deterrence makes sense where we are trying to affect individual behavior. But the logical way to "deter" a machine is to put the actual costs into the calculus it uses to make the decision. In practice, that translates into quantifying, and then operationalizing, the price we want robots to have to pay if they take certain actions we want to deter. And under the broadest interpretation of the economic view, even doctrines seemingly designed to prevent or deter conduct—like injunctions or prison sentences—could simply be construed as costs, albeit very high ones.

That said, we think it makes more sense to distinguish between remedies designed to internalize costs and those designed to enjoin, deter, or punish behavior. While some defendants faced with the latter may treat punitive damages or even prison sentences as mere costs of doing

---

158 See Casey, Amoral Machines, supra note __.
business, the remedy’s ultimate intent is to deter unlawful conduct, not to simply internalize its social costs.

For the vast majority of applications, legal remedies will likely be incorporated into machines through their “economic” formulation—resulting in robots that, by design, adopt this view of substantive law exclusively. Unless specifically programmed otherwise, distinctions between normative and economic goals will be utterly lost on robots. Thus, while it may be true to say that it is the rare “individual [who] will obey the law only if the consequences of violation are more painful than obedience,”\textsuperscript{159} this will be definitionally true of robots. And for reasons made clear in virtually every sci-fi plot line featuring robots, it will only be on the rarest of occasions that it actually makes sense to completely bar robots from engaging in certain types of conduct.

It, thus, appears that Holmes’s archetypical “bad man” will finally be brought to corporeal form, ironically, not as a man at all. And if Holmes’s metaphorical subject is truly “morally impoverished and analytically deficient,” as some accuse, it will have significant ramifications for robots.\textsuperscript{160}

\textbf{C. Teaching Robots to Behave}

Each of the major types and purposes of remedies we identified in section A will face challenges as applied to robots and AI. In this section we consider each in turn.

\textsuperscript{159} Laycock, Modern American Remedies, supra note ___ at 7.

\textsuperscript{160} See CHRISTOPH BEZEMEK, BAD FOR GOOD - PERSPECTIVES ON LAW AND FORCE, THE FORCE OF LAW REAPPLIED 1 (Springer Bezemek/Ladavic eds. 2016).
1. **Who Pays?**

The first purpose of damages—to compensate plaintiffs for their losses and so return them to their rightful position—is perhaps the easiest to apply to robots. True, robots don’t have any money, so they can’t actually pay damage awards themselves. In fact, the European Union Parliament specifically cited this fact in its recommendation against giving robots personhood, noting that they are not fully functioning members of society that could afford to pay their debts.\(^{161}\)

But this problem is hardly insurmountable. The law will rise to challenge. Someone built the robots, after all. And someone owns them. So if a robot causes harm, it may make sense for the company behind it to pay, just as when a defective machine causes harm today.

But it’s not that easy. Robots are composed of many complex components, learning from their interactions with thousands, millions, or even billions of data points, and often designed, operated, leased, or owned by multiple different companies. Which party is to internalize these costs? The one that designed the robot or AI in the first place? The one that collected and curated the data set used to train its algorithm in unpredictable ways? The users who bought the robot and deployed it in the field? Sometimes all of these roles will be one in the same, falling upon individuals operating in a single company, as was arguably the case when a self-driving Uber car killed a pedestrian in Tempe, Arizona.\(^{162}\)

---

\(^{161}\) See Amanda Wurah, *We Hold These Truths to Be Self-Evident, That All Robots Are Created Equal*, J. OF FUTURE STUD., DOI:10.6531/JFS.2017.22(2).A61.

\(^{162}\) See Wakabayashi, *surpra* note __.
In such instances, assigning responsibility may be easy. But often the chain of legal responsibility will be more complicated. Is a self-flying passenger drone an inherently dangerous product? If so, one set of rules might apply depending on whether it is the passenger or, instead, a third-party who is injured. Is the injury caused by this hypothetical drone the result of a design defect? If so, it may be the designer who should bear the risk.\footnote{Mark Geistfeld, \textit{A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation}, 105 \textit{Cal. L. Rev.} 102, 125 (2017).} But suppose instead that it was the result of a software defect that a different designer introduced through an aftermarket modification. Here, the law commonly shifts responsibility away from the manufacturer, if the modification was one that it didn’t intend.\footnote{Daniel A. Crane et al., \textit{A Survey of Legal Issues Arising from the Deployment of Autonomous and Connected Vehicles}, 23 \textit{Mich. Telecommunications and Tech. L. Rev.} 190, 215.} Indeed, companies regularly void warranties when third-parties modify their products or use them in unexpected ways. Things will get even more complicated if, as seems likely, some or all of the robot code is open source, raising the question of who ultimately is responsible for the code that goes into the car.\footnote{See Lothar Determann & Bruce Perens, \textit{Open Cars}, 32 \textit{Berkeley Tech. L.J.} 915 (2017); Bryan Casey, \textit{Open Source Robots} (working paper 2018).}

Robot designers, owners, operators, and users will, of course, fight over who bears true legal responsibility for causing the robot to behave the way it did. And these complex distinctions don’t even account for the role of third-parties causing robots to behave in adverse ways, as recently happened when Microsoft’s chatbot, Tay, turned into a proverbial Nazi after interacting with trolls on Twitter.\footnote{See Vincent, \textit{supra} note __. In retrospect, this event probably should have been a wake-up call for 2016.}
These problems can, and will, eventually be resolved by the courts. But long before any consensus is reached, we should expect no shortage of finger-pointing, as different companies and individuals clamor to shift responsibility for harms to others in the causal chain—whether just to minimize their costs or because there are legitimate disputes about how the behavior of different actors in the chain interacted to cause the harm. And if the AI is self-learning, we may really never know who is to blame.

2. **Law as Action: Shaping the Behavior of Rabota Economicus**

The second prong of the remedies triad—damage awards and equitable remedies designed to internalize costs and deter socially unproductive behavior—will likely prove even more problematic. If we want to deter a robot, we need to make sure that it is programmed to account for the consequences of its actions. Embedding this type of decision making in robots often means quantifying the various consequences of actions and instructing the robot to maximize the expected net monetary benefits of its behavior.

This might sound like heaven to an economist. Finally, we will have a truly rational *homo economicus* (or, more accurately, a *rabota economicus*)\(^ {167} \) who will internalize the social costs of its actions (at least insofar as those costs are accurately calculated in the courts) and modify its behavior accordingly. And if machine learning systems estimate these costs correctly, robots will

---

be Learned indeed—presumably deciding to do harm only when it is socially optimal (i.e. when \( B < PL \)).

But not so fast. Things are more complicated. Robots won’t reflexively care about money. They will do whatever we program them to do. We can align robot incentives with social incentives by properly pricing, punishing, or deterring the companies that design, train, own, or operate robots. Those companies, in turn, should internalize the relevant costs of their robots’ actions. It might be reasonable to assume that corporations and people want to maximize their rational self-interest and will, thus, program their robots accordingly. But not all will, either intentionally or unintentionally. There are at least three potential problems.

First, the goal of cost internalization through legal liability can only be accomplished by proxy. And it isn’t clear who the proxy will be. All the problems we noted in the prior section about assigning responsibility to compensate victims will return in spades as we try to force robots to account for the costs of their conduct. Even truly rational, profit maximizing companies with perfect information about the costs of their actions won’t internalize those costs unless they expect the legal system to hold them liable. If they are wrong, either in fearing liability when none exists or in believing someone else will foot the bill, their pricing will not accurately reflect reality.

Second, we are unlikely to have anything resembling “perfect” information about the potential harms robots may cause. As noted in Part I, robots operating in complex environments can do a wide variety of harmful things. Some of those things we want to stop altogether. Some

\[ \text{United States v. Carroll Towing Co., 159 F.2d 169, 171-73 (2d Cir. 1947) (the case in which Judge Learned Hand first expressed his canonical negligence formula).} \]
we want to discourage except in unusual circumstances. Some we want to outright permit but still price appropriately to account for externalities imposed on others. And some we want to permit despite their costs to society because the alternatives are worse.

Getting robots to make socially beneficial, or morally “right,” decisions means we first need a good sense of all the things that could go wrong. Unfortunately, we’re already imperfect at that. Then we’d need to decide whether the conduct is something we want to ban, discourage, tax, or simply permit. Having done so, we would then need to decide who in the chain of robot design, training, ownership, and operation should be responsible for the harm, if anyone. Then, we would need to figure out how likely each adverse outcome is in any given situation. Finally, we would need to assign a price to those potential harms—even the amorphous ones, such as a reduction in consumer privacy. And we’d want to balance those harms against reasonable alternatives to make sure the decision the robot made was the right one, even if it did cause harm.

Our entire system of tort law has been trying to accomplish this feat for centuries. And it hasn’t worked very well. Indeed, most of tort is composed of standards, as opposed to hard and fast rules, for good reason. Standards give us the leeway to reserve judgement for later, when we might have a better idea of the actual facts leading up to an event.

Tort law, for example, requires us to value injury, and—if we are to deter conduct—to decide on a multiplier to that value that serves as an optimal deterrent. While there are some circumstances in which we calculate these values formulaically,\(^{169}\) the primary way we do so is

---

\(^{169}\) See, e.g., H. Laurence Ross, Settled Out of Court (2d ed. 1980) (discussing the routinization of negligence and insurance compensation formulas for auto accidents).
by leaving it to juries to pick the right number after an injury has already occurred. Effective
deterrence in robots would, therefore, require accurate predictions about how juries might
assess specific harmful events, not to mention a host of other computationally complex
considerations. Scholars already find these types of predictions difficult, if not impossible, in the
human context.\(^{170}\) And we know virtually nothing of how juries will react to harmful events
caused by robots, particularly those exhibiting behaviors they can’t understand because the
algorithm is inscrutable.\(^{171}\) As we discuss below,\(^{172}\) this reality on the ground may even lead to
feedback loops, in which the very act of trying to price harms in a decision-making algorithm
changes the jury’s view of the robot’s responsibility.\(^{173}\)

The problem is even more complex than that, though, because robots don’t necessarily
care about money. They will maximize whatever they are programmed to. If we want them to
internalize the costs of their behavior, we will need to put those costs in terms robots can
understand—for example, as weights that go into a decision-making algorithm. That’s all well
and good for robots already designed to maximize profit in purely monetary terms—say, a day-
trading AI. But lots of robots will be designed with something other than money in mind. A
policing or parole algorithm might minimize the likelihood that a released offender commits
another crime. A weather-prediction system may maximize successful prediction outcomes. A

\(^{170}\) See LAYCOCK, supra note ___ at 165-6 (discussing multiple studies showing disagreement among juries
“over how to convert severity of injury into dollars”).

\(^{171}\) See infra notes ___ - ___ and accompanying text.

\(^{172}\) See infra notes ___ - ___ an accompanying text.

\(^{173}\) See, e.g., Malcom Gladwell, The Engineer’s Lament: Two Ways of Thinking About Automotive Safety,
THE NEW YORKER (May 4, 2015), https://www.newyorker.com/magazine/2015/05/04/the-engineers-
lament (describing jurors’ horror at internal memos that seemed to callously weigh the value of human
lives against business considerations).
surgery robot might maximize success in the surgery without considering certain side effects down the road. And a self-driving car might minimize time to destination subject to various constraints like generally obeying traffic laws and reducing the risk of accidents. But to build deterrence into those algorithms, we must convert certain divergent values into a common metric, whether it be money or something else.

A final complexity involving *rabota economicus* emerges for economic costs that are not directly reflected by legal remedies. The cost of any given decision, after all, is not just a function of the legal system. In many instances, extralegal forces such as ethical consumerism, corporate social responsibility, perception bias, and reputational costs will provide powerful checks on profit maximizing behaviors that might, otherwise, be expected to produce negative societal externalities. By pricing socially unacceptable behavior through the threat of public backlash, these and other market forces help to fill some of the gaps left by existing remedies regimes. But they may open up other holes, creating rather than internalizing externalities. In fact, in certain circumstances, these factors may end up utterly swamping the costs of actual legal liability. For instance, if I make it clear that my car will kill its driver rather than run over a pedestrian if the issue arises, people might not buy my car. The economic cost of lost sales may swamp the costs of liability from a contrary choice. [In the other direction, car companies could run into PR problems if their cars run over kids]. Put simply, it is aggregate profits—not just profits related to legal sanctions—that will drive robot decision making.

---

174 See Casey, *Amoral Machines*, supra note __ at fn 69 (discussing “the warped incentive signals conceivably sent [to robots] by transaction costs, first- and third-party insurance intermediaries, administrative costs, technical limitations, agency costs, information costs, detection costs, judgement proofing, human error and incompetence, consumer psychology, potential media backlash, and judicial and regulatory uncertainty”).
Further, even when a profit maximizing corporation is wholly responsible for the conduct of a robot, incentives may misalign for other reasons. Corporations might want robots that maximize the long-term value of their brand even if doing so imposes unnecessary hidden costs. Or, conversely, they may task their robots with creating content that goes viral and, therefore, maximizes short-term visibility—even if it is divisive and potentially contrary to the corporation’s long-term interest. Corporations may also decide that first-mover advantages are worth the risk of causing some injury in order to capture a long-term market. “Move fast and break things” is a slogan in Silicon Valley, one that has served many disruptive tech companies well. But this same slogan can take on somewhat more sinister cast when it is self-driving cars that are literally moving fast and breaking things.

Corporations are also likely to be siloed in ways that interfere with effective cost-internalization. Machine learning is a specialized programming skill, and programmers aren’t economists. Even those who are employed by profit maximizing companies interested in effectively internalizing their legal costs may see no reason to take the law into account, or may not be very good at it even if they try to. They may resent constant interference from the legal department in their design decisions. And agency costs mean that different subgroups within companies may be motivated by different incentives—as when sales divisions, manufacturing

---

175 See Johnathan Taplin, Move Fast and Break Things: How Facebook, Google, and Amazon Cornered Culture and Undermined Democracy (Little, Brown, and Co. 2017).

176 At least not most of them.
divisions, and service departments all get compensated based on different and potentially conflicting metrics.\textsuperscript{177}

Further, designers aren’t the only people whose motivations we need to worry about. What a self-learning robot will maximize depends not only on what it is designed to do—the default optimizing function or functions that it starts with—but also how it learns. To efficiently deter behavior we must be able to predict it. But if we don’t know how the robot will behave because it might discover novel ways of achieving the goals we specify, simply pricing in the cost of bad outcomes might have unpredictable effects. And even if it doesn’t, we once again have to confront the possibility that not all engineers will design their robots to maximize profit. Even if the designer of my self-driving car defaults to an algorithm that appropriately balances the risks to everyone associated with driving, I might personally prefer a car that protects its passengers at the expense of pedestrians. And if I (or, more realistically, a car company that wants to market to me) instructs the car accordingly, simply pricing the social cost of accidents into the algorithm won’t modify behavior in the way we hope.

This complex relationship between deterrence, responsibility, and financial liability does not, alone, differentiate robots from corporations or people. Deterrence is imperfect among humans, too, because humans aren’t motivated entirely by money and because they can’t always pay for the harm they cause. But what is different here is that the possibility of deterrence working \textit{at all} will depend entirely on the robot’s code. A robot programmed to be indifferent to...

money won’t be deterred by any level of legal sanction. And while making the responsible legal party pay\footnote{178} might encourage that party to design robots that do take adequate care, the division of responsibility between component makers, software designers, manufacturers, users, owners, and third-parties means that the law must be careful about who exactly it holds accountable.\footnote{179}

\section*{C. Deterrence Without Rational Actors: Is There Still a Role for Morality and Social Opprobrium in Robot Remedies?}

\subsection*{1. Equitable Monetary Relief and Punishment}

So far, we have focused on internalizing the costs of accidents or other injuries that result from otherwise socially desirable activities, such as driving cars. But we also need to worry about genuinely “bad” behavior by robots that may merit prohibition. Many of our equitable monetary remedies are aimed at this sort of conduct. Their goal is not to make defendants internalize costs—to put a price on socially valuable behavior because of the costs it imposes—but to prevent the behavior. If you steal my car, the law says that you don’t get to keep it even if you value it more than me. Rather, you hold it in constructive trust for me.\footnote{180} If you make profits by infringing my copyright or trade secret (but not my patent), the law will require you to disgorge those profits, paying me the money you made even if I never would have made it myself.\footnote{181} We

\begin{footnotesize}
\begin{enumerate}
\item Or face time behind bars.
\item As it gets easier to design AIs, these entities will be increasingly judgement proof. That will make us want to look upstream past the owner/user to the manufacturer. A second and more significant category of circumstances where a robot might depart from purely profit maximizing behavior involves instances where the chain of legal responsibility running from the robot to the manufacturer is intermediated by a downstream user.
\item See LAYCOCK, supra note ___ at 699-711.
\item Id. at 655-63.
\end{enumerate}
\end{footnotesize}
require defendants to give up such “unjust enrichment,” not because we think we need to do so to compensate the plaintiff, but because we don’t want the defendant to have the money.\footnote{182} \footnote{Id.}

These equitable rules share some similarities with the cost-internalization measures discussed in the last section. But there are two key differences: (1) the money a defendant must pay is not limited to what is needed to compensate the plaintiff, and (2) the defendant must give up all gains, making the entire activity unprofitable. The focus here is not on the plaintiff’s rightful position but on the defendant’s rightful position. And in the class of cases in which we often use these remedies, the defendant’s rightful position is one in which she didn’t engage in the activity at all.\footnote{183} \footnote{Id.}

From an economic perspective, depriving defendants of their gains is simply a matter of coming up with a number. It might be greater than, equal to, or less than the damages we would otherwise impose to internalize the costs of unlawful conduct or to restore the plaintiff’s rightful position. But there is something psychologically effective about taking away a defendant’s gains altogether. Indeed, in certain contexts, it might be a better means of deterring humans than the threat of paying compensatory damages, even if those damages turn out to be higher than a disgorgement remedy would. When it comes to robots, however, there is little reason to think that the notion of taking “all your profits” will have the same psychological effects. True, if you set “profit = 0,” a profit maximizing AI would not engage in the conduct. But that same logic would apply with equal force if the damages award made the activity unprofitable too.
Remedies focused on the defendant’s rightful position do have one significant economic advantage over damages remedies intended strictly as ex ante deterrence: we can calculate them after the fact once we have all the necessary information. If we want to use the threat of damages to deter conduct, we need to predict the likelihood and severity of the harm that the conduct will cause. But if we care only about depriving the defendant of benefits on the theory that that doing so will deter her, we just need to wait to set the number until the parties get to court and figure out how much the defendant actually gained. That often won’t be trivial. The benefit of stealing a trade secret, for example, can be as amorphous as a “quicker time to market” or a “more competitive product.” But it’s still likely to be easier than predicting in advance who will be injured and by how much.

This same calculus doesn’t work for injuries that are the byproduct of productive behavior. It doesn’t make sense to say that a self-driving car that hits a pedestrian should disgorge its profits. It likely didn’t profit from hitting the pedestrian. And we don’t want to force defendants to disgorge all the value they make from driving. But defendant-focused equitable monetary remedies, like disgorgement or constructive trust, may have advantages for robot torts for which our goal is to stop the conduct altogether, not to simply to price it efficiently.

---

184 See supra notes __ - __ and accompanying text.


For examples, see K-2 Ski Co. v. Head Ski Co., 506 F.2d 471, 474 (9th Cir. 1974) (“We are satisfied that the appropriate duration for the injunction should be the period of time it would have taken Head, either by reverse engineering or by independent development, to develop its ski legitimately without use of the K-2 trade secrets.”); Winston Research Corp. v. Minn. Mining & Mfg. Co., 350 F.2d 134, 145–46 (9th Cir. 1965) (discussing injunction protection for a machine company); Verigy US, Inc. v. Mayder, No. C-07-04330 RMW, 2008 WL 564634, at *9, *11 (N.D. Cal. Feb. 29, 2008) (granting a five-month injunction to account for the lag time defendant would have faced in getting to market absent misappropriation).
2. Detection, Deterrence, and Punitive Damages

The fact that robots won’t be affected by the psychological impact of certain remedies also has consequences for how we should think about the threat of detection. For a robot to be optimally deterred by remedies like disgorgement—which rely on human psychology to maximize their effects—we must also detect and sanction the misconduct 100% of the time. That, in turn, leads us to the problem of robots (or their masters) that hide misconduct.

To be sure, many robot harms will be well publicized. The spate of autonomous vehicle accidents covered by media in recent years provides one stark example. But countless robot harms will be of far subtler, so-called “blackbox,” varieties and will, therefore, be much harder to detect.

Makers and trainers of robots may have incentives to hide their behavior, particularly when it is profitable but illegal. If a company’s parole algorithm concludes (whether on the merits of the data or not) that black people should be denied parole more often than similarly situated white people, it might not want the world to know. And if you, as an owner, tweaked the algorithm on your car to run over pedestrians rather than put your own life at risk, you might

---


187 This term refers to algorithms that are inscrutable to outsiders, either by virtue of complexity, lack of technical fluency, or trade secrets protection.

seek to hide that too. We have already seen remarkable efforts by companies conspiring to cover up wrongdoing, many of which succeeded for years.\textsuperscript{189} Often such conspiracies are brought down by sheer virtue of their scale—i.e. the fact that many people know about and participate in the wrongdoing. This same property may be less true of future robotics firms, which may require fewer people to participate and cover up unlawful acts.\textsuperscript{190}

Further, robots that teach themselves certain behaviors might not know they are doing anything wrong. And if their algorithms are sophisticated enough, neither may anyone else for that matter.\textsuperscript{191} Deterrence will work on a robot only if the cost of the legal penalty is encoded in the algorithm. A robot that doesn't know it will be required to disgorge its profits from certain types of conduct will not accurately price those costs and so will optimize for the wrong behaviors.

The economic theory of deterrence responds to the improbability of getting caught by ratcheting up the sanctions when you are caught, setting the probability of detection times the penalty imposed equal to the harms actually caused.\textsuperscript{192} Proportionality of punishment makes


\textsuperscript{190} Desai and Kroll argue for protections for whistleblowers who identify flaws in robotic design in an effort to reduce the risk of such cover-ups. Desai & Kroll, \textit{supra} note __, at 3.

\textsuperscript{191} Pricing algorithms may effectively replicate the anticompetitive effects of a cartel by predicting the behavior of their rivals, for instance. See Kellie Lerner & David Rochelson, \textit{How Do you Solve a Problem Like Algorithmic Price Fixing}, ANTITRUST & TRADE REG. DAILY (BNA), Feb. 7, 2018.

sense here. As the chance of detection goes down we want the damage award to go up. And machines can do this math far better than humans can. Indeed, this idea may be tailor-made for robots. Becker’s “high sanctions infrequently applied” approach seems unfair in many humans contexts because it can have widely varied interpersonal effects: even if we get equal deterrence from a 100% chance of a year in prison or a 10% chance of 10 years in prison, the lottery system that punishes a few very harshly seems intuitively unfair. We want our laws to protect both victims and wrongdoers against some forms of moral bad luck (whereas Becker’s approach exacerbates it). But robots will internalize the probability of punishment as well as its magnitude, so we may be able to encourage efficient behavior without worrying about treating all robots equitably. Further, we are unlikely to feel bad for harshly-punished robots in the ways that we might for human beings.

Even if we decide to heed Becker’s advice, getting the numbers right presumes that we have a good estimate of the proportion of torts committed by robots that go undetected. That’s tough to do, especially for newly introduced technologies. And it also requires programmers to predict the multiplier and feed those calculations into the algorithm, something that might not be a straightforward undertaking for any of the variety of reasons covered in the last section (not to mention the possibility that we get the numbers wrong, which will either over- or under-deter certain behaviors).


193 High sanctions, for example, “may lead juries to be less likely to convict defendants, or may induce individuals to engage in greater efforts to avoid detection.” Polinsky & Shavell, supra note __ at __ (citing James Andreoni, Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?, 22 RAND J. of Econ. 385 (1991); Arun Malik, Avoidance, Screening and Optimum Enforcement, 21 R. and J. of Econ. 341 (1990)).
Maybe society will, instead, be able to force corporations to internalize their costs through non-legal mechanisms—e.g. by voting with their wallets when a company’s robots engage in misconduct. But this, too, is easier said than done, particularly for the types of “systemic harms” described in Part I. In the era of big data and even bigger trade secrets, structural asymmetries often prevent meaningful public engagement with the data and software critical to measuring and understanding the behavior of complex machines. Because private companies retain almost exclusive control over both the proprietary software running the machines and their resultant data,194 barriers to accessing the information necessary to understand the reasons behind particular machine decisions can often be insurmountable. What’s more, even in circumstances where the information is available, evidence of unlawful decision making can still be notoriously difficult to detect. As the AI Now Institute notes, “[u]nintended consequences and inequalities [of sophisticated computational systems] are by nature collective, relative and contextual, making measurement and baseline comparisons difficult” and creating the “potential for both over- and under-counting biases in measurement of distributions given the limits on observable circumstances for individuals, problems with population gaps and possible measurement errors.”195

Current trends in AI appear likely to only exacerbate this problem. As Bryce Goodman and Seth Flaxman observe, even after “putting aside any barriers arising from technical fluency [and]  

ignoring the importance of the training model,” modern machine learning techniques pose significant “tradeoff[s] between the representational capacity of a model and its interpretability.”196 Systems capable of achieving the richest predictive results tend to do so through the use of aggregation, averaging, or multilayered techniques which, in turn, make it difficult to determine the exact features that play the largest predictive role.197 Thus, even more so than with the past generation of algorithms governing machines, understanding how modern robots arrive at a given decision can be prohibitively difficult, if not technically impossible—even for the designers themselves.198 As a result, potentially unlawful or defective decision making within such systems can often only be demonstrated in hindsight, after measuring the unevenly distributed outcomes once they have already occurred. And as systems get more complex, maybe not even then.

The risk presented by this combination of factors is not so much that corporations will intentionally build bad robots in order to eke out extra profits, but that bad “effects [will] simply happen, without public understanding or deliberation, led by technology companies and governments that are yet to understand the broader implications of their technologies once they are released into complex social systems.”199 Indeed, much of the misconduct that tomorrow’s designers, policymakers, and watchdogs must guard against might not be intentional at all. Self-learning machines may develop algorithms that take into account factors we may not want them

196 See Bryce Goodman & Seth Flaxman, European Union Regulations on Algorithmic Decision-Making and a “Right To Explanation”, ICML WORKSHOP ON HUMAN INTERPRETABILITY IN MACHINE LEARNING (2016).

197 Id.

198 Id.

to, like race or economic status. But on some occasions, taking precisely those factors into account will actually get us to the ultimate result of interest.

For this reason, we think AI transparency is no panacea. Transparency is a desirable goal in the abstract. But it may inherently be at odds with the benefits of certain robotics applications. We may be able to find out what an AI system did. But, increasingly, we may not be able to understand why it did what it did. Calls for transparency are useful to the extent that they identify bad behavior, defective designs, or rogue algorithms. But mostly what people want when they talk about transparency is an explanation they can understand. Why was my loan application denied? Why did the car swerve in the way it did? For some robots, we simply won’t know the answer. Even if we see how the algorithm comes to a conclusion, we won’t necessarily be able to understand how it derived a relationship between, say, butterfly populations in Mongolia and thunderstorms in Ethiopia, or why it thinks the precise time of day and year should affect the speed at which it proceeds through an intersection.

Are we right to be bothered by this? Should we have a right to understand the mens rea of robots? Or to impute explanations so we can appropriately channel opprobrium? Our punitive and deterrence remedies are based on identifying and weeding out bad behavior. The search for

---

200 See, e.g., FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION. (Cambridge, MA: Harvard University Press 2015); Katyal, supra note ___.

201 See Deven R. Desai & Joshua A. Kroll, TRUST BUT VERIFY: A GUIDE TO ALGORITHMS AND THE LAW, 31 HARV. J. L. & TECH. 1 (2016) (arguing that the push for transparency is misguided because it misunderstands the nature of the algorithms at stake).

202 We made these examples up. The real ones are likely to be even weirder. The whole point is that they are inexplicable to humans. Even today, AI is making decisions humans struggle to understand. Dave Gershgorn, AI IS NOW SO COMPLEX ITS CREATORS CAN’T TRUST WHY IT MAKES DECISIONS, QUARTZ, Dec. 7, 2017. Some companies are studying the decisions of their own AIs to try to unpack how they are made. See Cade Metz, GOOGLE RESEARCHERS SAY THEY’RE LEARNING HOW MACHINES LEARN, N.Y. TIMES, Mar. 7, 2018, at B3.
that bad behavior is much of what drives the “intuitive appeal of explainable machines.” But our intuitions may not always serve us well. The question is whether the demand for an explanation is actually serving legitimate purposes (Preventing Skynet? Stopping discrimination?) or just making us feel that we’re the ones in charge. The punitive and equitable monetary side of remedies law wants to understand the “why” question because we want to assign blame. But that might not be a meaningful question when applied to a robot. More on this later. 

3. Inhuman, All Too Inhuman

a. Punishing Robots for Responding to Punishment

Even economic forms of deterrence—both legal and extralegal—will look different than they currently do when people or corporations are being deterred. Deterrence of people often takes advantage of cognitive biases and risk aversion. People don’t want to go to jail, for instance, so they will avoid conduct that might lead to that result. But robots can be deterred only to the extent that their algorithms are modified to include external sanctions as part of the risk-reward calculus. Once more, we might view this as a good thing—the ultimate triumph of a rational law and economics calculus of decision making. But humans who interact with robots may

---


204 See infra note __-__ and accompanying text.

205 See Peter M. Asaro, Punishment, Reinforcement Learning, and Machine Agency, http://peterasaro.org/writing/cosmopolis.globalist.it%20%20Punishment,%20Reinforcement%20Learning%20Machine%20Agency.htm (“a key intuitive difference between humans . . . and machines is that when a human misbehaves, you punish it, whereas when a machine does, you fix it. On our present theory, however, it becomes clear that punishing and fixing are essentially the same: punishing is a clumsy, external way of modifying the utility function.”).
demand a non-economic form of moral justice even from entities that lack the human capacity to understand the wrongfulness of their actions (a fact that anyone who has ever hit a malfunctioning device in frustration can understand).

Indeed, the sheer rationality of robot decision making may itself provoke the ire of humans. Any economist will tell you that the optimal number of deaths from many socially beneficial activities is more than zero. Were it otherwise, our cars would never go more than five miles an hour. Indeed, we would rarely leave our homes at all.

Effective deterrence of robots requires that we calculate the costs of harm caused by the robots interacting with the world. If we want a robot to take optimal care, we need it to figure out not just how likely a particular harm is but how it should weight the occurrence of that harm. The social cost of running over a child in a crosswalk is high. But it isn’t infinite.

Even today, we deal with those costs in remedies law unevenly. The effective statistical price of a human life in court decisions is all over the map. The calculation is generally done ad hoc and after the fact. That allows us to avoid explicitly discussing politically fraught concepts.

---


208 “Global variation in estimates of the value of life range from $70,000 to $16.3 million.” See Deborah L. Rhode et al., Legal Ethics 645 (U. Casebook Series, 7th ed. 2016) (citing Eric A. Posner & Cass R. Sunstein, Dollars and Death, 72 U. CHI. L. REV. 537 (2005); Binyamin Appelbaum, As U.S. Agencies Put More Value on a Life, Businesses Fret, N.Y. TIMES (Feb. 17, 2011)). “In the United States, federal agencies operate with figures generally ranging from roughly $6 to $9 million—but tort awards for wrongful death are typically a fraction of that, and even agency estimates tend to shift with the political winds.” See Rhode et al., supra note ___ at 645.
that can lead to accusations of “trading lives for cash.” And it may work acceptably for humans, because we have instinctive reactions against injuring others that make deterrence less important. But, in many instances, robots will need to quantify the value we put on a life if they are to modify their behavior at all. Accordingly, the companies that make robots will have to figure out how much they value human life, and they will have to write it down in the algorithm for all to see (at least after extensive discovery).

The problem is that people strongly resist the idea of actually making this calculus explicit. They oppose the seemingly callous idea of putting a monetary value on a human life, and juries punish companies that make explicit the very cost-benefit calculations that economists want them to. Human instincts in this direction help explain why we punish intentional conduct more harshly than negligent conduct. A deliberate decision to run over a pedestrian strikes us as worse than hitting one by accident because you weren’t paying attention. Our assumption is that, if you acted deliberately, you could have chosen not to cause the harm, thereby making you a bad actor who needs to modify your behavior. But that assumption often operates even when causing that harm was the socially responsible thing to do, or at least was justified from cost-benefit perspective.

Things are more complicated, of course. We do try to create justifications and excuses in the law, even for intentional conduct that we think is socially acceptable. But juries often have a

---


visceral desire to hold someone responsible when bad things happen. And they are inclined to treat killing or injuring a human being as a bad act even if it was (statistically) inevitable. They will rebel against treating it as a mere cost of doing business. Thinking about it in such terms offends many people’s sense of human decency.

b. Punishment as Catharsis: Punching Robots

Punishment may serve other, non-monetary purposes as well. We punish, for instance, to channel social opprobrium. That can set norms by sending a message about the sorts of things we won’t tolerate as a society. And it may also make us feel better. We have victim allocation in court for good reason, after all. It may provide useful information to courts. But it also helps people to grieve and to feel their story has been heard.

Our instinct to punish is likely to extend to robots. We may want, as Christina Mulligan puts it, to punch a robot that has done us wrong.\(^\text{212}\) Certainly people punch or smash inanimate objects all the time.\(^\text{213}\) Juries might similarly want to punish a robot, not to create optimal cost internalization but because it makes the jury and the victim feel better. It’s already quite easy to think of robots as humans.\(^\text{214}\) We naturally anthropomorphize.\(^\text{215}\) That instinct is likely to get stronger over time, as companies increasingly deploy “social robots” that intentionally pull on

---

\(^\text{212}\) See Mulligan, supra note __.

\(^\text{213}\) See supra note __ and accompanying text.

\(^\text{214}\) See also, e.g., Robbi Gonzalez, Hey Alexa, What Are You Doing to My Kid’s Brain?, WIRED (May 11, 2018), https://www.wired.com/story/hey-alexa-what-are-you-doing-to-my-kids-brain/ (describing the tendency for children to anthropomorphize chat bots like Amazon’s Alexa).

\(^\text{215}\) See Calo, Robotics and the Lessons of Cyberlaw supra note __ at 538 (terming this phenomenon “social valence”).
Humans will expect human-like robots to act, well, human. And we may be surprised, even angry, when they don’t. Our instinct may increasingly be to punish humanoid robots as we would a person—even if, from an economic perspective, it’s silly. Making us feel better may be an end unto itself. But hopefully there is a way to do it that doesn’t involve wanton destruction of or damage to robots.

D. Ordering Robots to Behave

All these problems with monetary remedies as deterrents seem to point in the direction of using injunctive relief more with robots than we currently do with people. Rather than trying to encourage robot designers to build in correctly priced algorithms to induce efficient care, wouldn’t it be easier just to tell the robot what to do—and what not to do?

1. Be Careful What You Wish For

First, the good news: injunctions against robots might be simpler than against people or corporations, because they can be enforced with code. A court can order a robot, say, not to take race into account in a processing an algorithm. Likewise, it can order a self-driving car not to exceed the speed limit. Someone will have to translate that injunction, written in legalese, into code the robot can understand. But once they do, the robot will obey the injunction. This virtual

---

216 Like Asimov’s fiction, Westworld’s days as pure fantasy may be numbered.

217 It’s an open question whether we will react differently to a self-learning AI that isn’t in corporeal form and doesn’t act in human-like ways.

guarantee of compliance seems like a significant advantage over existing injunctions. It is often much harder to coerce people (and especially groups of people in corporations) to comply with similar court orders—even when the consequences are dire.

But, once again, not so fast. As the adage goes (and as legions of genies in bottles have taught us): be careful what you wish for. Automatic, unthinking compliance with an injunction is a good idea only if we’re quite confident that the injunction itself is a good idea. Now, obviously the court thinks the injunction improves the world. Otherwise, it wouldn’t issue it. But the fact that injunctions against people aren’t self-enforcing offers some potential breathing room for parties and courts to add a dose of common sense when circumstances change. This is a common problem in law. It’s a major reason we have standards rather than rules in many cases. 219 And it’s the reason that even when we do have rules, we don’t enforce them perfectly. To a person (and even to a police officer), “don’t exceed the speed limit” implicitly means “don’t exceed the speed limit unless you’re rushing someone to the emergency room or it would be unsafe not to.” “Don’t cross the double yellow line” implicitly means “don’t cross the double yellow line unless you need to swerve out of the lane to avoid running over a kid.” No cop is going to ticket you for such a maneuver. Similarly, even if an injunction says “don’t cut lumber on this property,” a court isn’t going to hold you in contempt for taking down the one rotten tree that’s about to fall on your neighbor’s house. That’s because people understand that rules and injunctions come with the implied catchall “unless you have sufficient justification for departing from the rule” exception.

219 See supra note ___ - ___ and accompanying text.
Try telling that to a robot, though. Machines, unlike at least some humans, lack common sense. They operate according to their instructions—no more, no less. If you mean “don’t cross the double yellow line unless you need to swerve out of the lane to avoid running over a kid” you need to say that. Meanwhile, Av should probably avoid adults too, so better put that in the algorithm. . . And maybe dogs. . . And deer and squirrels, too. Or maybe not: crossing into oncoming traffic is dangerous, so while we might do it to avoid hitting a kid even if it raises the risk of a head-on collision we shouldn’t do it to avoid a squirrel unless the risk of a head-on collision seems low. (Sorry, squirrels). If you want the self-driving car to do all that, you need to tell it exactly when to swerve and when not to swerve. That’s hard. It’s more plausible to give each outcome weights—killing squirrels is bad, but head-on collisions are much worse, and killing a kid is (Probably? Maybe?) worse still. But then we’re back to deterrence and cost internalization, not injunctions.

Further, even if we can specify the outcome we want with sufficient precision in an injunction, we need to be extremely careful about the permissible means a robot can use to achieve that result. Think back to our example from the introduction. The drone did exactly what we told it to. The problem is that we weren’t sufficiently clear in communicating what we wanted it to do. We wanted it to head to the center of the circle without shutting down and without human intervention. But we didn’t say that, because we didn’t anticipate the possibility of the drone doing what it did.220

220 Former Secretary of Defense Donald Rumsfeld famously described these types of foreseeability concerns:

There are known knowns. These are things that we know we know. There are known unknowns. That is to say, there are things that we know we don’t know. But there are also unknown unknowns. There are things we don’t know we don’t know. And if one looks
The “be careful what you wish for” problem is a major one for robotics and AI. Tim Urban of Wait But Why tells the hypothetical story of Turry, a self-learning AI that is designed to mimic handwritten greeting cards. If you don’t specify the things it can’t do, or at least impose cost weights, an AI could literally take over all the resources of the world and devote them to producing handwritten greeting cards. Computer programmers will, we hope, be aware of this problem and be extremely careful about phrasing their instructions to a robot in just the right way, with precise caveats and limiting conditions to prevent them turning into Skynet or Turry. But judges aren’t computer programmers, and they are unlikely to be as knowledgeable or as careful in what they order robots to do or not do. And even if we could do it, an injunction of this sort represents a pretty significant intrusion into the product design process, something courts have been unwilling to do in other circumstances. Whether or not courts are right to shy away throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones.


222 This is a variation on Eliezer Yudkowsky’s and Nick Bostrom’s famous “paper clip maximizer” thought experiment.

223 Search King, Inc. v. Google Technology, Inc., No. CIV-02-1457-M, 2003 WL 21464568, at *4 (W.D. Okla. May 27, 2003) (ruling that Google’s page rankings were “subjective result[s]” that constituted “constitutionally protected opinions . . . entitled to full constitutional protection”); Langdon v. Google, Inc., 474 F. Supp. 2d 622, 629–30 (D. Del. 2007) (refusing to affirmatively order Google and Microsoft to rank certain search results prominently); United States v. Microsoft Corp., 253 F.3d 34 (D.C. Cir. 2001) (applying balancing test to judge whether new product is predatory); United States v. Microsoft Corp., 147 F.3d 935 (D.C. Cir. 1998) (deferring to a company’s claims of product improvement to avoid enmeshing the court in design decisions); Allied Orthopedic Appliances, Inc. v. Tyco Health Care Grp., 592 F.3d 991, 998–99 (9th Cir. 2010) (permitting companies to introduce any product that constitutes “improvement” over predecessors).
from telling companies how to design products generally, we think that’s a good instinct when it comes to robotics, at least in the early stages of the industry.

To issue an effective injunction that causes a robot to do what we want it to do (and nothing else) requires both extreme foresight and extreme precision in drafting it. If injunctions are to work at all, courts will have to spend a lot more time thinking about exactly what they want to happen and all the possible circumstances that could arise. If past experience is any indication, they are unlikely to do very well at it. That’s not a knock on courts. Rather, the problem is twofold: Words are notoriously bad at conveying our intended meaning, and people are notoriously bad at predicting the future. And if we fall into either of these traps, the consequences of drafting the injunction incompletely may be quite severe.

2. “What Do You Mean You Can’t?!”

Courts that nonetheless persist in ordering robots not to do something may run into a second, more surprising problem: it may not be simple or even possible to comply with the

---


225 These two facts combine to provide the plot line of virtually every Isaac Asimov novel. ISAAC ASIMOV, THE REST OF THE ROBOTS 43 (1964) (remarking “[t]here was just enough ambiguity in the Three Laws [of robotics found in his works] to provide the conflicts and uncertainties required for new stories, and, to my great relief, it seemed always to be possible to think up a new angle out of the sixty-one words of the Three Laws”).
injunction. Just as robots don’t have money, they also don’t read and implement court opinions. And they aren’t likely to be a party to the case in any event. Enjoining a robot, in other words, really means ordering someone else to implement code that changes the behavior of the robot.

The most likely party to face such an injunction is the owner of the robot. They are the ones who will likely have been determined to have violated the law, say by using a discriminatory algorithm in a police profiling decision or operating a self-driving car that has behaved unsafely. But most owners won’t have the technical ability, and perhaps not even the right, to modify the algorithm their robot runs. The most a court could order may be that they ask the vendor who supplied the robot to make the change, or perhaps to take the robot off the market as long as it doesn’t comply with the injunction.

Even if the developer is a party to the case, perhaps on a design defect theory, the self-learning nature of many modern robots makes simply changing the algorithm more complicated still. A court may, for instance, order the designer of a robot that makes predictions about recidivism for parole boards not to take race into account. But that assumes that the robot is

---


227 More on this below when we consider the robot death penalty. See infra notes ___-___ and accompanying text.

228 Far from hypothetical, courts have considered these types of arguments on multiple occasions in recent years. See, e.g., State v. Loomis, 881 N.W.2d 749, 767 (Wis. 2016); Malenchik v. State, 928 N.E.2d 564 (Ind. 2010).
simply doing what it was originally programmed to do. That may be less and less common as machine learning proliferates. Ordering a robot to “unlearn” something it has learned through a learning algorithm is much less straightforward than ordering it to include or not include a particular function in its algorithm. Depending on how the robot learns it might not even be possible.

Life gets easier if the courts can control what training information is fed to robots in the first place. At the extremes, a court might order a company to take badly-trained robots out of service and to train new ones from scratch. But as the example in the introduction indicates, the effects of training material on robots are not always predictable. And the results of training are themselves unpredictable, so even controlling the training dataset is no guarantee that a robot, once trained, will behave as the court wants it to.

Further, the future may bring robots that are not only trained in complicated ways but that train themselves in ways we do not understand and cannot replicate. Ordering such a robot to produce or not produce a particular result, or even to consider or not consider a particular factor, may be futile. If we don’t understand how the robot makes decisions we can’t effectively guide those decisions. It is one thing to look at a transparent algorithm written by programmers and see whether it includes the race of the parolee as a factor. It is quite another to try to untangle whether a robot has learned that race matters by looking at the data and how that learning is implemented in an always-changing algorithm that doesn’t itself explicitly include race. An algorithm that is simply told to minimize the risk of recidivism but not to take race directly into account might end up generating proxies that are correlated with race instead. That’s fine if those proxies are in fact the variable of interest. If, say, the fact that members of a
minority group commit disproportionately more crimes results from the fact that they are poorer than average, an algorithm that gets to the same result by considering family poverty instead of race may solve the problem.\textsuperscript{229} But if the algorithm has really just found a proxy for race (say, the street you grew up on in a segregated neighborhood) we aren’t any better off. And it is much harder to tell a robot not to consider “race or anything that serves as a proxy for race.”

Courts are used to telling people to do something and having them do it. They may have little patience for the uncertainties of machine learning systems. And they are quite likely to have even less patience with lawyers who tell them their “client” can’t comply with the court’s order.

3. Unintended Consequences

Even when the injunction is simple and clearly identifies who should change the algorithm and how, ordering a robot to change how it “thinks” is likely to have unintended consequences. Consider two examples.

(1) We don’t want self-driving cars to hit pedestrians. But just brute-forcing that result might lead to other problems, from taking crowded freeways instead of less-crowded surface streets to running into other cars. Some of those consequences could be worse, either because a head-on collision kills more people than running over the pedestrian would or, more likely, because instructing the car to act in a certain way may cause it to avoid a very small chance of killing a pedestrian by avoiding surface streets altogether (even though the collective cost of traffic jams might be quite great). This is a version of the same problem we saw in damages: we

\textsuperscript{229} Whether we want to disproportionately punish poor people is another matter, of course, but doing so isn’t race discrimination.
need to assign a cost to various outcomes if we want an algorithm to weigh the alternatives. But here the injunction effectively sets the cost as infinite. That’s fine if there really is nothing to balance on the other side. But that will rarely be true.

(2) The case against algorithmic bias seems one of the strongest, and easiest to enjoin, cases. And if that bias results simply from a bad training set, it may be straightforward to fix. But if an algorithm takes account of a prohibited variable like race, gender, or religion because that variable matters in the data, simply prohibiting consideration of that relevant information can have unanticipated consequences. One possible consequence is that we make the algorithm worse at its job. We might be fine as a society with a certain amount of that in exchange for the moral clarity that comes with not risking discriminating against minorities. But where people are in fact different, insisting on treating them alike can itself be a form of discrimination. Being male, for example, is an extremely strong predictor of criminality. Men commit many more crimes than women, and male offenders are much more likely to reoffend. We suspect police and judges know this and take it into account, consciously or unconsciously, in their arrest, charging, and sentencing decisions, though they would never say so out loud. But a robot won’t conceal what it’s doing. A court that confronts such an robot is likely to order the it not to take gender into

\[\text{230} \text{ It is possible a company will simply factor the cost of contempt into the algorithm, but that seems unlikely. And if they do, courts will probably not be happy about it.}\]

\[\text{231} \text{ To the extent that the algorithms are transparent to third-parties, of course. Yet, even detecting bias within a system can be less straightforward than may initially appear. See Corbett-Davies et al., Algorithmic Decision Making and the Cost of Fairness, https://arxiv.org/pdf/1701.08230.pdf (pushing back on Julia Angwin’s claim that the COMPAS criminal sentencing algorithm was biased).}\]

\[\text{232} \text{ See supra notes } \text{ and accompanying text. See also, e.g., Reuters Staff, New Zealand Passport Robot Tells Applicant of Asian Descent to Open Eyes, REUTERS (Dec. 7, 2016) (reporting on facial recognition software failure that resulted from an evidently unrepresentative training set).}\]
account, since doing so seems a rather obvious constitutional violation. But it turns out that if you order pretrial sentencing algorithms to ignore gender entirely, you end up discriminating against women, since they get lumped in with the heightened risks of recidivism that men pose.233

Ordering a robot not to violate the law can lead to additional legal difficulties when injunctions are directed against discrete subsystems within larger robotics systems. These types of injunctions seem likeliest to be granted against newly introduced subsystems within a tried and true application—given that older systems will, by definition, have a longer track record of success. Not only could targeting one component of a larger system change it in unpredictable and often undesirable ways, doing so could also discourage innovation. With the field of AI improving by leaps and bounds, maybe we should be less protective of tried-and-true approaches and more willing to experiment. Even though some of those experiments will fail, the overall arc is likely to bend towards better systems than we have now. But we won’t get there if courts are too quick to shut down new systems while leaving established but imperfect procedures in place. If the alternative to a flawed predictive policing algorithm is the gut instincts of a large number of cops, some of whom are overtly racist and others of whom are subconsciously biased, we might be better off with the robots after all.

III. Rethinking Remedies for Robots

We’ve seen that robots and AI pose a number of challenges to the law of remedies as it is currently applied. In this section, we offer some preliminary thoughts about how we might redesign the law for the world that is fast approaching. We don’t intend this to be the last word on how to design remedies for robots. Much more remains to be done. Rather, we hope it marks the beginning of a conversation on these issues. The suggestions we outline below will help align the law of remedies with what we know about the behavior of robots.

A. Compensation, Fault, and the Plaintiff’s “Rightful” Position

Compensation is the easiest remedy to translate to robots because its focus is on the (presumably human or corporate) plaintiff. The harm is to plaintiffs, not robots, and the same valuation measurement problems arise here that always do in calculating damages. But as we have seen, robot defendants do introduce some complications. Who is responsible when a robot misbehaves? The designer? The manufacturer? The owner? Under current tort law the answer may depend on whether the harm resulted from a design defect, a problem in training, or an

234 A different issue arises when the robot is itself the injured party. What would it mean to put a robot in its rightful position? What that likely means, at least until we recognize robot rights, is putting the robot’s owner or operator in its rightful position. While there are issues here, we think they are likely to be more straightforward than most of the ones we have discussed. If a robot is damaged or destroyed through negligence or vandalism, we will normally treat that as we would damage to any other property. It’s easy enough to replace parts for pre-programmed bots, but if the algorithms learned from unique, one-off interactions and cannot be recovered, robots might not be so easy to replace. Hopefully emergent AI will be backed up regularly, though, so it could still be replaced.

We can imagine deliberately unique robots, though. Technologies like block chain are now being used to impose scarcity (e.g. crypto kittens, digital cats you can raise and trade via the blockchain. Because cats. And blockchain). We could see this same phenomenon transposed to robot personalities, so as to artificially impose scarcity. Tay, for instance, was a unique chat bot deployed by Microsoft. Like too many people, when exposed to the Internet, Tay quickly became a fascist. See supra note __. When Microsoft shut her down, her “learning” was gone and could not be replaced. Few would lament that in this specific case, but we can imagine valuation difficulties if a tortious or malicious act destroys a unique AI personality.
error in operation. But learning AIs will blur this line; the designer might not be the one training the AI in ways that caused it to subsequently do harm.

Many (though not all) of the problems with compensating plaintiffs for robot injury come from tort law’s focus on fault as a prerequisite to responsibility. We generally hold people responsible for accidental injuries only if they are negligent. The focus on fault may make sense where people are concerned, but it is much less meaningful as applied to a robot. What does it mean for a robot to be negligent? This isn’t really a remedies question, though it may be a causation question. But it’s not obvious that it is a question worth asking. True, we might want to single out certain design or implementation choices that we think are problematic and discourage them. But in many environments in which robots operate there are more direct regulatory means to do so. NHTSA, for instance, approves or mandates the introduction of many vehicle safety technologies. So, too, does the FAA for aircraft. If we think a particular design shouldn’t be on the market at all, some regulatory bodies will be able to simply prohibit it.

Tort law does, as noted above, also serve to raise the cost of products that cause harm and therefore deter the deployment of inefficient ones. In theory, tort law makes that calculus directly by setting $B < PL$ or demanding some other risk-utility test.$^{235}$ But in doing so, the law makes a threshold judgment as to whether there should be any liability for costs imposed on others. An alternative formulation would require an actor to pay for any harm it causes, negligent or not. That shifts the focus of deciding whether $B$ is less than $PL$ to the company that makes the product rather than to the courts.

---

Perhaps we just want someone to pay the costs of any harm robots cause, even if that occurred without a wrongful or illegal act. We often use negligence as a proxy for whether the defendant’s conduct was justified despite the costs it imposes, but that will be harder to do with robots. And maybe we don’t want to ask a jury to decide who was at fault if programmers can actually code in a standard of care that internalizes the harm the robot imposes on others.

Existing remedies laws might get us there, though not without modification. We do impose strict liability in some circumstances.\(^\text{236}\) That’s easier when the plaintiff is a passive victim like someone injured by pollution from a factory or from a product that unexpectedly exploded. It’s more problematic when both the plaintiff and the defendant might have contributed to the cause of the injury. When two cars collide, one reason we try to decide who was at fault (or whether both were in part) is to fairly allocate the cost of injury to the party who was best positioned to avoid it. Allocating that fault will raise new challenges when a robot-driven car gets into an accident because its driving capabilities and the sorts of evidence it can provide will be different than human drivers. We can’t cross-examine the robot to interrogate its state of mind. On the other hand, AVs are likely to record every aspect of the accident, giving us a better record than fallible human memory currently does. A second reason we focus on blame is that we need to worry that the parties might lie about what happened. But self-driving cars are likely to keep clear records and video that may make it easier to figure out what happened.\(^\text{237}\) And it may make

\(^{236}\) Albeit, not in the case of defective products.

\(^{237}\) See, e.g., Francesca M. Favaro, Examining Accident Reports Involving Autonomous Vehicles in California, PLoS ONE. https://doi.org/10.1371/journal.pone.0184952 (reconstructing autonomous vehicle accidents through the data collected by onboard recording devices).
less sense to try to assess fault when two robotic cars collide, though we expect that will be a much rarer occurrence.238

Yet another reason we assess fault against people is that blame for wrongdoing can encourage more careful behavior. As we discussed in Part II, that isn’t likely to work, or at least to work in the same way, with robots. Without the element of moral culpability that underlies much remedies law, we might be better off looking to insurance schemes or no-fault liability regimes to internalize the costs robots impose rather than using existing legal rules in a fruitless quest to get robots to act morally. As robots and AI take on more responsibility in our society, the law should move away from efforts to assess blame and towards a system that internalizes the costs those machines impose on those around them. Doing so will make the problem of coding effective care easier. And it may increasingly mean tort cases involving robots don’t show up in the legal system at all, but in some sort of regulatory compensation system.239

That might incline us towards some sort of a no-fault system as self-driving cars and self-flying planes increasingly share space with their human-operated counterparts.240 While we


239 See, e.g., Kenneth Abraham & Robert Rabin, A New Legal Regime For a New Era (arguing for a no-fault accident compensation regime once autonomous vehicles have reached sufficient market penetration); Katherine Wallis, New Zealand’s 2005 ‘No-Fault’ Compensation Reforms and Medical Professional Accountability For Harm, 126 NEW ZEALAND MED. J. 33 (2013) (detailing New Zealand’s “taxpayer funded accident compensation scheme to provide compensation for medical injury”).

could assess the overall safety of an autonomous vehicle and—assuming it was safer than the human standard—deny liability altogether in crashes, we think the determination of what AV behavior falls below the standard of care is likely to be hard for the foreseeable future. Adding in the metric of fault doesn’t make much sense, and depriving injured parties of any remedy might not make sense either. The simplest way to train AVs to avoid doing unnecessary harm is to make them responsible for the harm they cause whether or not they were “negligent.”

That doesn’t solve all problems with AVs, particularly when they interact with humans, because we still must decide when an AV “causes” an accident with a human driver. While occasional fatal crashes have dominated the headlines, most AV-human car accidents involve humans running into AVs, often because the AV did something legal and presumably safe but unexpected, like driving the speed limit or coming to a complete stop at an intersection. While that may suggest that we want to program AVs to behave in a more predictable way, it’s hard to fault the AV for being rear-ended because it came to a complete stop at an intersection. Without the addition of a contributory negligence defense (which functions a lot like plain old B<PL from a fault perspective), innovators would end up disproportionately bearing costs, human drivers wouldn’t be priced off the roads as quickly as they should, and companies would also be apt to

---

241 For a suggestion along these lines, see Mark A. Geistfeld, A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation, 105 CALIF. L. REV. 1611, 1612 (2017). Geistfeld would leave an exception for cars that were designed or manufactured defectively and for those that were hacked.

spend less on safety from a competitive perspective, since no amount of investment could get them off the liability hook when people, themselves, created the hazards.\footnote{For a more detailed discussion of these issues, see Bryan Casey, working paper.}

Thus, while we think moral fault makes little sense in accidents involving AVs, and perhaps any consideration of blame is problematic when considering accidents between two AVs,\footnote{Even then we might want to assess liability against the AV that is using an outdated or less-safe algorithm, to encourage the development of better safety technology in AVs.} we will still need to compare the behavior of humans and AVs in order to make sure that we give proper incentives to human drivers. Comparative negligence may still matter for robot drivers, therefore. But it is the idealized cost-internalization vision of negligence reflected in Learned Hand’s formula, not consciousness of fault or state of mind, that we should care about.

\section{B. Punishment, Deterrence, and the Human Id}

Deterrence, unlike compensation, is forward-looking. We want robots to internalize the costs of their actions even apart from compensation of particular victims. The good news is that cost internalization has the potential to work better with robots than it does with people.\footnote{See, e.g., Aaron Chalfin & Justin McCrary, \textit{Criminal Deterrence: A Review of the Literature}, 55 J. OF \textit{ECON.} \textit{LIT.} 5 (2017) (finding limited evidence “that crime responds to the severity of criminal sanctions”); Menusch Khadjavi, \textit{On the Interaction of Deterrence and Emotions}, 31 J. L., \textit{ECON.}, AND \textit{ORG.} 287 (2015) (examining the influence of human emotion on different deterrent effects).} Robot algorithms may allow us to internalize costs further down the causal chain than tort law normally does, for example by accounting for the social cost of pollution or other nebulous injuries to society as a whole. But these must be priced, again requiring fraught social tradeoffs to be made explicit. And the pricing should be cost-based. We should minimize the
psychologically-driven aspects of deterrence (jail, disgorgement of ill-gotten gains) and replace them with more rational measures of cost.

Doing so is at odds with many of the mechanisms we have for deterrence, however. Often those mechanisms are directed at showing moral opprobrium or at punishing people in ways we expect them to react to psychologically. Christina Mulligan’s idea of punching robots who wrong us sounds silly, but there is a serious idea behind it. Much of our law of remedies, including our search for fault (but also the way in which we punish), is designed not to compensate plaintiffs or even to internalize costs for defendants but to make us feel better. This sometimes involves “sending a message,” but often the defendant isn’t the target of the message. Perhaps it is society as a whole; large punitive damage awards or harsh criminal penalties can signal the things we won’t tolerate as a society, and overly lenient sentences can do the opposite. That is a broader social conversation, albeit one usually carried out in the context of legal remedies. But often, remedies are purely cathartic: we want someone to blame to make ourselves feel better for the bad thing that happened to us. When there is no obvious candidate for blame, we go to considerable lengths to find one. Punishment in this sense is a form of psychological compensation—the very act of punishing the defendant is the compensation.

246 Mulligan, supra note ___ at ___.

247 The recent controversy that erupted over a Stanford University swimmer’s six-month sentence for sexual assault provides just one examples. See The judge Who Sentenced Brock Turner to Six Months in Stanford Rape Case is Fighting a Recall, ASS’D PRESS (May 18, 2018), http://www.latimes.com/local/lanow/la-me-persky-recall-20180518-story.html.

248 For instance, we have relaxed the rules of causation in remedies law in order to compensate indirect victims of large oil spills. Oil Pollution Act, 33 U.S.C. § 2701 (2006).
This seems socially wasteful. Punishing robots, not to make them behave better but just to punish them, is kind of like kicking a puppy that can’t understand why it’s being hurt. The same might be true of punishing people to make us feel better, but with robots the punishment is stripped of any pretense that it is sending a message to make the robot understand the wrongness of its actions.

We don’t deny that there is a real phenomenon at work here, or even that it may benefit the victim psychologically. But it might not make sense to serve those goals when suing robots. Is there a way to make us stop? To channel that instinct into other areas than the legal system where it might be more productive? Should we just abandon the signaling function of remedies altogether? Perhaps, but we probably won’t, human nature being what it is.

Rather, if we want to rationalize remedies for robots, we might need to take human decision makers (especially untrained ones like juries) out of the remedies equation in some cases (or at least closely constrain the remedies they can order and the reasons that justify those remedies).\(^{249}\) Juries are likely to have an instinct to punish bad behavior by robots. But punishment makes sense only if we think compensation for damages is inadequate and so defendants will take insufficient precautions or engage in socially harmful behavior we want them to stop.\(^{250}\) A robot that calculates the cost of its various decisions accurately will make bad

\(^{249}\) One day, we may even want to go further by putting robots in charge of remedies decisions. See, e.g., Eugene Volokh, *Chief Justice Robots*, PULSE LUNCH TALK (JAN. 24, 2018), https://law.ucla.edu/news-and-events/4096/2018/1/24/pulse-lunch-talk-c--professor-eugene-volokh/ (discussing possibility of automating such legal decisions).

\(^{250}\) It might be, for various reasons. We cut off liability with proximate cause before we have traced all the harm from wrongful acts. See Pruitt v. Allied Chem. Corp., 523 F. Supp. 975 (E.D. Va. 1981) (denying relief for indirect injury from pollution); Lemley, *Fruit of the Poisonous Tree*, supra note __. We are bad at valuing pain and suffering and do so in idiosyncratic ways that will sometimes undercompensate plaintiffs. And we have imposed caps on liability in many circumstances that undercompensate for actual injuries.
decisions if we add in data on the likelihood of punitive damages that exceed those costs. And if the robot is being punished precisely because it is calculating how many people it’s ok to kill, the problem becomes recursive and we will undo the purpose of optimal deterrence and cost internalization.

C. Reeducating Robots

Injunctions, as we have seen, are both important and problematic remedies for robots. Can courts order a robot to do better—to change its programming? Perhaps we can require changes in design, or we might compel some sorts of modifications to learning algorithms.

Courts in general favor injunctions that preserve the status quo and prohibit parties from changing things (so-called prohibitory injunctions). They are traditionally more reluctant to order parties to do affirmative things to change the state of affairs (mandatory injunctions). It does happen, particularly in impact litigation after a final finding of liability. But courts tend to shy away from involving themselves in the details of running a business or designing a product if they can avoid it. With robots, though, there’s no avoiding it—whether the injunction is mandatory or prohibitory. An order for a robot to do something and an order for it to not do something both

---

*See Michael Kang, Don’t Tell Juries About Statutory Damage Caps: The Merits of Nondisclosure, 66 U. Chi. L. Rev. 469 (1999) (noting “[i]t has become increasingly common for Congress and state legislatures to enact statutory limits on the amount of money damages that a plaintiff can recover in a jury trial). But if we are not compensating plaintiffs properly, the solution is to compensate them properly, not to add a damages multiplier to awards whether or not they are actually compensatory.

251 *See supra* notes ___-___ and accompanying text (discussing this problem).

252 *See supra* notes ___-___ and accompanying text (discussing this phenomenon in antitrust context).
require redesigning the product. Courts should take care when and how they grant those injunctions.

In light of this reality, what exactly will courts order robots to do? One likely compromise is not to order the code to be written in a specific way, but rather to order the company to find a way to achieve a specific result. As we saw in Part II, that by no means solves the problems with injunctions against robots. But it does offer some flexibility to the company that needs to rewrite their code, ideally without introducing other problems in the process.

One way to increase that flexibility is to give companies time to comply. Courts generally expect their orders to be obeyed quickly. But writing quick code often means writing bad code, particularly in an ever-changing, complex machine learning system. Courts and regulators should be patient. Self-driving cars go through years of testing before we are comfortable that they will drive safely. We shouldn’t just rewrite that code and put it on the streets without testing. So courts should delay implementation of their orders against robots to enable the defendant to develop and test a solution that doesn’t cause more problems than it solves. Regulators have so far shown admirable restraint in not rushing to mandate particular rules for AVs.253

Turning that results-oriented goal into an injunction runs into legal problems, though. Obviously we don’t want cars to run over kids, but a judge can’t simply order that. Court orders can’t just say “obey the law”;254 they must give clear notice of what the defendant must do. So

---


254 FED. R. Civ. P. 65(b) (“Every order granting an injunction and every restraining order shall set forth the reasons for its issuance; shall be specific in terms; [and] shall describe in reasonable detail . . . the act or acts sought to be restrained.”); Burton v. City of Belle Glade, 178 F.3d 1175, 1201 (11th Cir. 1999) (rejecting
an injunction might say “stop the car if the likelihood that a pedestrian will imminently enter the intersection is greater than 0.2%.”

In some cases, orders might require robots to make their algorithms worse. An injunction preventing the police from taking gender into account in predicting criminality may make it harder to predict who will commit crimes. We might nonetheless want to order it, either to counteract existing bias reflected in the training data or simply because recognizing gender differences in criminality violates a constitutional norm even if the differences are real. But in doing so we are departing from the real world, ordering companies to train their robots to make decisions based on the society we would like to have rather than the one we do have.

One compensating advantage to robot injunctions is that the orders involve rewriting code, and in a connected world these changes can often be shipped out retroactively. Tesla updates the software periodically in cars it has already sold. Unlike traditional products, where an injunction is generally limited to the sale of products in the future, court orders against robots can affect existing robots already in the hands of consumers. That makes the injunction much

_________________________

an injunction that simply forbade future discrimination); Hughey v. JMS Dev. Corp., 78 F.3d 1523 (11th Cir. 1996) (overturning injunction that forbade discharge of waste in violation of the Clean Water Act).

more effective, though it also may raise due process concerns on the part of owners not a party to the case whose robot suddenly behaves differently or stops working altogether.256

D. The Robot Death Penalty?

One area of remedies that becomes easier when the defendant is a robot is criminal law. We worry about the consequences of depriving people of liberty even when they have done something wrong. We worry even more about depriving them of life. It is an adage that we put a heavy thumb on the scale in favor of innocence, allowing the guilty to go free before punishing the innocent.257 We require guilt to be proven beyond a reasonable doubt, and we have special protections before imposing the death penalty.258 Some states and most countries have in fact abolished the death penalty altogether.

But robots aren’t people, and we might worry less about robot liberty.259 True, robots will be entitled to due process, if for no other reason than that they are owned by people or companies that would lose valuable property if their robots disappeared. But one new and significant form of remedy becomes available against robots that isn’t available against people in most circumstances: the robot death penalty. If a robot is causing unjustified harm and we can’t

256 Cf. Hassell v. Bird, ___ Cal.4th ___ (2018) (stressing that “the courts’ power to order people to do (or to refrain from doing) things is generally limited to the parties in the case”).


259 For a suggestion that robots can be held liable for crimes just as people can, see Gabriel Hallevy, Dangerous Robots—Artificial Intelligence vs. Human Intelligence, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3121905.
stop it, either because we don’t understand how it works or because the harm is inextricably bound up with its programming, we might simply shut it down.\footnote{We distinguish this from the case where humans use robots to commit crimes. A human can use a drone to fire missiles, for instance, or to spy on people. See Amanda McAllister, \textit{Stranger than Science Fiction: The Rise of A.I. Interrogation in the Dawn of Autonomous Robots and the Need for an Additional Protocol to the U.N. Convention Against Torture}, 101 \textsc{Minn. L. Rev.} 2527 (2017). If the robot is the instrument of the crime but not its cause it is the human, not the robot, that should face criminal penalties.} Turning off malfunctioning robots is a simple and effective, if blunt, instrument to enforce an injunction. And removing the robot from commercial deployment may allow us to figure out what went wrong by engaging in the sorts of testing we couldn’t do without jeopardizing operational function.

\textit{Should} we shut down misbehaving robots? In some cases the answer is yes. Corporations do it all the time.\footnote{See, e.g., Perez, , \textit{supra} note __.} And essentially any time you change the code you are changing the robot, replacing it with a new and (hopefully) improved one.

Whether courts can order a robot shut down over the objections of its owner is a slightly harder question, but the answer is still probably yes. Courts kill pets that repeatedly attack others and can order other types of machines shut down if they are unreasonably dangerous.\footnote{See Safia Hussain, \textit{Attacking the Dog-Bite Epidemic: Why Breed Specific Legislation Won’t Solve the Dangerous Dog Dilemma}, 74 \textsc{Fordham L. Rev.} 2847 (2006); \textit{see also supra} notes __ - __ and accompanying text.} If robot can be replaced by others with competing algorithms, we probably want to shut it down if it is operating below the standard of care. One way learning algorithms improve is through natural selection,\footnote{These often go by the name “genetic algorithms.”} and shutting down the bad ones is just a form of that process. But if an AI has developed unique attributes as a result of its own learning, we have the problem of dual-use
technologies. A self-learning AI may behave differently in both good and bad ways, and those differences may be related. The robot death penalty kills off the good as well as the bad. So we want to do it only if we think the harm the robot is causing is sufficiently great and the unique benefit of its approach sufficiently low that the cost of losing the benefit is worth it.

For this reason, the use of the robot death penalty should probably be rare. Shutting down a robot, especially a self-learning one, means shutting down an avenue of innovation. We should do that only if there is strong evidence that the AI does more harm than good and that there isn’t a less intrusive way to solve the problem. Just as courts should be reluctant to tell robots to change how they behave, they should be reluctant to turn the robots off altogether.

Further, the robot death penalty presents more serious due process issues with respect to the existing stock of robots in the hands of people other than the defendant. Courts generally can’t reach out and take away property in the hands of non-parties without due process, even if those products cause problems and even if the court can order the company to stop selling new copies of the product. But the malleability of software presents some grey areas here. It is OK to order a defendant to push out changes to the product, though it’s an easier case if the recipient has the choice of whether to accept those changes. The company can probably stop supporting

---


the product remotely. But a software “upgrade” that is really just an effort to “brick” an existing product seems a reach too far.266

Finally, there is the possibility that the law will recognize robots as sentient entities with their own rights.267 That isn’t as far-fetched as it sounds. Corporations aren’t people either, but they get legal rights (in some instances more rights than people).268 Animals also have some rights, though fewer than humans or corporations.269 Charles Stross has called corporations the first AIs.270 Like AIs, corporations are created by people, designed to serve ends dictated by people, but over time come to serve their own purposes.271 It’s not impossible that in the future we will extend at least some legal rights to robots as well, particularly unique robots with learned

266 See, e.g., Universal City Studios Productions LLLP v. TickBox TV LLC, No. CV 17-7496-MWF (ASx) (C.D. Cal. Jan. 30, 2018). It appears that the court is poised to order a device maker to use its software update mechanism to remove functionality and content from users’ devices.


269 See Hussain, supra note __; Christopher Sepes, Animal Law Evolution: Treating Pets as Persons in Tort and Custody Disputes 2010 U. ILL. L. REV. 1339 (2010). Animals have only limited standing to bring cases, but they sometimes can. See, e.g., Naruto v. Slater, 2018 WL 1902414 (9th Cir. April 23, 2018) (finding that a crested macaque alleged facts sufficient to establish Article III standing because it was the apparent author and owner of selfies it took and may have suffered legally cognizable harms); Cetacean Cmty. v. Bush, 386 F.3d 1169, 1175 (9th Cir. 2004) (stating that mere fact that plaintiffs were animals did not rule out possibility of standing). But the law also refuses to treat animals as anything other than property in many instances. See, e.g., Johnson v. Douglas, 723 N.Y.S.2d 627, 628 (Sup. Ct. 2001) (refusing to allow emotional distress damages because dog was considered personal property).


271 Id.
behavior. And one of those rights may well be the right not to be shut down without due process.\textsuperscript{272}

\textbf{E. What Robots Can Teach Us about Remedies for Humans}

Robots present a number of challenges to courts imposing remedies on robotic and AI defendants. Working through those challenges is valuable and important in its own right. But doing so can also teach us some things about the law of remedies as it currently applies to people and corporations.

First, much of remedies, like much of law, is preoccupied with fault—identifying wrongdoers and treating them differently. There may be good reasons for that, both within the legal system and in society as a whole. But it works better in some types of cases than in others. Our preoccupation with blame motivates many remedies, particularly monetary equitable relief. It distorts damage awards, particularly when something really bad happens and there is not an obvious culprit. It also applies poorly to corporations, which don’t really have a unitary purpose in the way a person might.\textsuperscript{273} As importantly, it is also costly, requiring us to assess blame in traffic cases that could otherwise be resolved more easily if we didn’t have to evaluate witness credibility. A fault-based legal system doesn’t work particularly well in a world of robots. But perhaps the problem is bigger than that: it might not work well in a world of multinational

\textsuperscript{272} Cf. Asimov’s three laws of robotics, \textit{supra} note \_, which would allow any person to kill a robot for any reason. Isaac Asimov clearly never anticipated Reddit. Trying to implement the three laws of robotics would leave the world strewn with the carcasses of robots killed by griefers.

\textsuperscript{273} For an argument that current methods of punishing corporations are ineffective and that corporations should face organizational remedies—the equivalent of rewriting their “code”—see Mihailis E. Diamantis, \textit{How to Punish a Corporation} (working paper 2018).
corporations either.\footnote{We are by no means the first to have advanced this line of argument.} We should look for opportunities to avoid deciding fault, particularly when human behavior is not the primary issue in a legal case.\footnote{Shavell, supra note __. This does not mean, however, that we don’t need laws. Some have suggested that we won’t need rules or standards in the future because we can just rely on machine judgment to decide what the right thing to do is in any specific situation. See Anthony J. Casey & Anthony Niblett, The Death of Rules and Standards, 92 IND. L.J. 1401 (2017). For the reasons we explained in Part I, we think that highly unlikely. Robots will cause all sorts of harm the legal system will want to remedy. Cf. Dan L. Burk, Algorithmic Fair Use, __ U. CHI. L. REV. __ (forthcoming 2018) (explaining why algorithms won’t effectively replace standards in many cases).} We should look for opportunities to avoid deciding fault, particularly when human behavior is not the primary issue in a legal case.\footnote{Lemley, not Casey.}

A second lesson is the extent to which our legal remedies, while nominally about compensation, actually serve other purposes, particularly retribution. We described remedies law at the outset of the paper as being about “what you get when you win.” But decades of personal experience litigating cases\footnote{Lemley, not Casey.} have reinforced the important lesson that what plaintiffs want is quite often something the legal system isn’t prepared to give. They may want to be heard, they may want justice to be done, or they may want to send a message to the defendant or to others. Often what they want—closure, or for the wrong to be undone—is something the system not only can’t give them but that the process of a lawsuit actually makes worse. The disconnect between what plaintiffs want and what the law can give them skews remedies law in various ways. Some do no harm: awards of nominal damages or injunctions that vindicate a position while not really changing the status quo. But we often do the legal equivalent of punching robots—punishing people to make ourselves feel better, even as we frequently deny compensation for real injuries. It’s just that it’s easier to see when it’s a robot you’re punching.
A final lesson is that our legal system sweeps some hard problems under the rug. We don’t tell the world how much a human life is worth. We make judgments on that issue every day, but we do them haphazardly and indirectly, often while denying we are doing any such thing. We make compromises and bargains in the jury room, awarding damages that don’t reflect the actual injury the law is intended to redress but some other, perhaps impermissible consideration. And we make judgments about people and situations in and outside of court without articulating a reason for it, and often in circumstances where we either couldn’t articulate that decisionmaking process or where doing so would make it clear we were violating the law. We swerve our car on reflex or instinct, sometimes avoiding danger but sometimes making things worse. We don’t do that because of a rational cost-benefit calculus, but in a split-second judgment based on imperfect information. Police decide whether to stop a car, and judges whether to grant bail, based on experience, instinct, and bias as much as on cold, hard data.

Robots expose those hidden aspects of our legal system and our society. A robot can’t make an instinctive judgment about the value of a human life, or about the safety of swerving to avoid a squirrel, or about the likelihood of female convicts reoffending compared to their male counterparts. If robots have to make those decisions—and they will, just as people do—they will have to show their work. And showing that work will, at times, expose the tolerances and affordances our legal system currently ignores. That might be a good thing, ferreting out our

---

racism, unequal treatment, and sloppy economic thinking in the valuation of life and property. Or it might be a bad thing, particularly if we have to confront our failings but can’t actually do away with them. It’s probably both. But whatever one thinks about it, robots make explicit many decisions our legal system and our society have long decided not to think or talk about. For that, if nothing else, remedies for robots deserve serious attention.

IV. Conclusion

Robots and AI systems will do bad things. When they do, our legal system will step in to try to make things right. But how it does so matters. Our remedies rules, not surprisingly, aren’t written with robots in mind. Adapting those rules to deal with bad robots will require a nuanced understanding of how robots and AI work, but also some fundamental rethinking of what remedies we award and why. That rethinking, in turn, will expose some issues that affect people, not just robots. We need a law of remedies for robots. But in the final analysis, remedies for robots may also end up being remedies for all of us.