2013

The application of traditional tort theory to embodied machine intelligence

Curtis E.A. Karnow

The Robotics and the Law Conference, Center for Internet and Society (Stanford Law School, April 2013)

The application of traditional tort theory to embodied machine intelligence

By

Curtis E.A. Karnow[1]

> ….ask a humanoid robot to reach his right hand to touch his left ear. In most of the cases I saw - the robot tried to get to the left ear through the head…[2]

> The future is already here - it's just unevenly distributed.[3]

Introduction

I assume interesting robots embody machine learning, the product of e.g., genetic algorithms, neural nets, or other sorts of feedback loops which generate unpredictable behavior. That is, these robots are given instructions as to ultimate goals and determine for themselves the means of accomplishing these goals. The means are not predictable by either the operator-owner or by the original programmers. Rather, the software teaches itself by running experiments or making other sorts of real or virtual attempts at a solution, corrects for error and approximates a result which it then implements. I call these "autonomous" robots.

This note discusses the traditional tort theories of liability such as negligence and strict liability and suggests these are likely insufficient to impose liability on legal entities (people and companies) selling or employing autonomous robots. (Because robots are not themselves legal entities and do not hold property, it is pointless to hold robots liable for their actions.) Commentators have previously made conflicting suggestions on the application of traditional tort law to the acts of robots and machine intelligence.[4]

---

[1] Judge of the Superior Court, County of San Francisco. This note is meant only to further discussion, and I express no opinion as to how I or any other judge might rule on any issue, nor do I express any opinion on any pending or impending case.

[2] http://robotics.stackexchange.com/questions/342/what-are-some-common-mistakes-that-robots-make.

[3] Attributed (in some form) to William Gibson. See http://quoteinvestigator.com/2012/01/24/future-has-arrived/; http://www.economist.com/node/811961

[4] Some suggest that there are no rules that govern robot actions, at least under European law. C. Leroux, "A Green Paper on Legal Issues in Robotics," http://www.jura.uni-wuerzburg.de/uploads/media/A_green_paper_on_legal_issues_in_robotics_-_Leroux_01.pdf (2/11/2012). Others suggest the opposite, that there are so many laws hampering developments in robotics that "we can and should similarly immunize robotics manufacturers for many of the uses to which their products are put," R. Calo, "Robotics & The Law: Liability For Personal Robots," http://ftp.documation.com/references/ABA10a/PDfs/2_1.pdf; see also Danielle Citron, "Bright ideas: Talking About Robotics With Ryan Calo," http://www.concurringopinions.com/archives/2010/05/bright-ideas-talking-about-robotics-with-ryan-calo.html. Another commentary arguing that existing law is a barrier to developments in machine intelligence is found at

In this note, I provide the essential working definitions of 'autonomous' as well as the legal notion of 'foreseeability' which lies at the heart of tort liability. The note is not concerned with the policy, ethics, or other issues arising from the use of robots including armed and unarmed drones, because those, as I define them, are not currently autonomous, and do not implicate the legal issues I discuss.

I conclude with some speculation on how the actions of robotic intelligence may become susceptible to traditional tort law, and in doing so change the way in which those legal tests are applied.

This note is written for both legal and nonlegal audiences, i.e., those involved in the development of robots. It must follow that both audiences will be disappointed-- the lawyers in my relatively superficial legal discussion, and the engineers and programmers who will sadly shake their heads at my simplistic view of developments in their field. I plead guilty to both indictments.

I. *Robots And Autonomy*

The terms *robot* is used indiscriminately to refer to a wide variety of machines which exhibit, or are said to exhibit, some semblance of 'intelligence'. Fantasy, fiction and reality are the amalgam from which we draw the meaning of the term *robot*, ascribing it to Schwarzenegger's Terminator, unmanned but wholly guided submarines and Mars rovers, missiles partially guided in real time, as well the Roomba floor cleaner and Sony's robot dog[5] which may not be humanly guided in real time but operate from previously fixed code. Many commentators do not spend much time distinguishing these sorts of robots as they address the difficulty of applying law to their effects; this is a mistake because the interesting legal issues only pertain to a small (but growing) set of them. Interesting robots, for purposes of this paper, are those which are not simply autonomous in the sense of not being under real time control of a human, but autonomous in the sense that the methods selected by the robots to accomplish the human-generated goal are not predicable by the human. Uninteresting robots, therefore, are such things as industrial factory robots, tethered robots,[6] missiles and drones.

Steven J. Frank, "Tort Adjudication and the Emergence of Artificial Intelligence Software," 21 SUFFOLK U. L. REV. 623, 639 (1987). As noted by others, I raised the issue some years earlier. Bert-Jaap Koops, et al., *Bridging the Accountability Gap: Rights for New Entities in the Information Society?,* 11 MINN. J.L. SCI. & TECH. 497, 539-40 (2010)("In 1996, Karnow investigated the issue of legal solutions for harm caused by distributed artificial intelligences. His major point is that, at this moment, we see emergent AIs that operate in the real world with decision programs, making 'decisions unforeseen by humans.'")(notes omitted). *See* Curtis Karnow, "Liability for Distributed Artificial Intelligences," 11 BERKELEY TECHNOLOGY LAW JOURNAL 147 (1996), reprinted in Curtis Karnow, FUTURE CODES: ESSAYS IN ADVANCED COMPUTER TECHNOLOGY AND THE LAW (1997). As with other commentators, however, I did not distinguish between or discuss autonomous and non-autonomous machine intelligence, nor did I distinguish embodied from disembodied software, and so I failed to explain specifically why certain types of robots posed problems for traditional tort law. I hope that is remedied to some extent in this paper.

[5] Up to about 2006 Sony made the Aibo robot dog which could move, play ball (at least with the special ball that came with the machine), and exhibit certain other 'dog' behaviors such as barking, chasing, and reacting to certain speech.

[6] I include unmanned submersibles and Mars rovers, where the tether is not only long but delayed by up to about 21 minutes (the maximum time signals take to travel between Earth to Mars).

A good example of the confusion in this area is generated by the Aegis Combat system, deployed on naval ships of many nations and currently made by Lockheed Martin.[7] One commentator writes:

> Machines entirely capable of replacing humans are not yet on the market, but robotic systems capable of using lethal force *without a human in the loop* do already exist. The U.S. Navy's Aegis Combat System, which is capable of autonomously tracking enemy aircraft and guiding weapons onto them, is an example.[8]

While it is true that Aegis can automatically detect and track targets, as well as provide missile guidance functions for over 100 simultaneous targets, it is not true that it autonomously determines which targets to attack. Its 'autonomous' attack functions are limited to midflight and terminal guidance.[9] It is, at best a "human supervised autonomous system"[10] and the nightmare error made in 1988 shooting down an unarmed civilian airliner was human targeting error, including communications to the commanding officer that were at odds with the data generated by Aegis.[11] These weapons are designed to allow humans to decide what to attack.[12]

The notion of intelligence as applied to machines is often just shorthand for "I don't know how they do that so quickly," an amazement borne of ignorance. We might in that way ascribe intelligence to Apple's Siri, which can respond to basic voice commands with vaguely contextually correct responses, missile defense systems which distinguish hostile intruders, and stock market programs which in fractions of a second calculate the best price and execute trades. The apparent magic of these advanced technologies is generally a function of speed outside the human scale, and of the observer's ignorance of the programs used.

But ignorance of the method used to generate an output is not the same as an unforeseeable output, and we can use this distinction to pursue a rough working definition of *autonomous*.

Most owners of the Roomba floor cleaner or Sony's Aibo dog robot know nothing of the program that in each case allows the unit to return to a charging station when battery life is low. This behavior occurs without concurrent human supervision, and appears autonomous in some vague sense. Farther up the line we have Google's (and others') driverless cars, which operate with considerably more information about the environment and make a wide variety of choices

---

[7] http://www.lockheedmartin.com/us/products/aegis.html.

[8] Paul Robinson, "Who Will Be Accountable for Military Technology?" (Nov. 15, 2012), http://www.slate.com/articles/technology/future_tense/2012/11/lethal_autonomous_robots_drones_tms_and_other_military_technologies_raise.html (emphasis supplied). Other descriptions, too, blur the role of human supervision. For example in Ronald Arkin, GOVERNING LETHAL BEHAVIORS IN AUTONOMOUS ROBOTS (2009), the author quotes a Navy description of Aegis as "capable of autonomously performing its own search, detect, evaluation, track, engage and kill assessment functions," *id*. at 7-8.

[9] The "SM-6 [missile] receives midcourse flight control from the Aegis combat system via ship's radar; terminal flight control is autonomous via the missile's active seeker or supported by the Aegis combat system via the ship's illuminator." U.S. Navy Standard SM Missile, http://www.dote.osd.mil/pub/reports/FY2012/pdf/navy/2012sm-6.pdf. I suggest that by "autonomous" the writer means "automated".

[10] Michael N. Schmitt, "Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics" *Harvard National Security Journal Features* (2013) http://ssrn.com/abstract=2184826.

[11] *See generally*, http://en.wikipedia.org/wiki/Aegis_Combat_System.

[12] Schmitt, *supra* note 10 at note 19.

without concurrent human control.[13] But "much research work is necessary to make more cognitive systems and reach a higher grade of autonomy." The requirements for this "higher grade" of autonomy provide an important clue for the meaning of true autonomy:

> An important step towards this goal is learning, e. g. how other vehicles look like and move, and using this knowledge for perception, navigation and interaction purposes. …. [we also may need] imitation learning, meaning to deduce the vehicle's behavior from the observed behavior of other traffic participants or from manually operated runs. However, there will always be unexpected situations never seen and "learned" before, which the vehicle nevertheless has to cope with. The difficult interaction between autonomous vehicles and other traffic participants, both robots and humans, will also need much work. For this purpose the behavior and intentions of other traffic participants have to be gathered.[14]

The point, I suggest, is that true autonomy involves self-learning: where the program does not simply apply a human-made heuristic (as with Roomba and Aibo), but generates its own heuristic. Thus, *autonomous* obviously does not refer to the behavior of tethered robots (such as unmanned submersibles and drones) nor (less obviously) to the behavior of robots operating without concurrent human control but in an entirely preprogrammed or automated way (such as AAMRAM missiles and industrial robots). Sophisticated sensors, extensive programming, and fast execution can *simulate* autonomy through automation. Some functions allow for leisurely human decision-making; others, such as Aegis, in the heat of battle, tempt human abdication and so spawn the illusion of autonomous decisions making. But this is not autonomy as I mean it.

Even with this working definition—programs that generate their own heuristic—autonomy is a matter of degree, a theme I will repeat. Nevertheless, this note focuses on functions which are in fact autonomous, in that sense.

A. Self Taught Programs

Some forty years ago John Holland and others wrote of software which made software: genetic algorithms.[15] The work was inspired by the evolutionary process of discarding inefficient modules and having the efficient ones survive. Holland took a series of code fragments, and through a series of iterations generated code that no human would ever have written. Later, one experimenter won a prize for his genetic algorithm's design of an antenna for a spacecraft, which bested the human efforts.[16]

Physical analogues of this disembodied programming are now tested in laboratories. We have, for example, physical modules each of which has a simple capability and together, in an endless variety of combinations, can achieve tasks uninstructed by an overarching piece of software. These modules may have capabilities such as locomotion, ground based or airborne; grippers,

---

[13] Thorsten Luettel, et al., "Autonomous Ground Vehicles: Concepts and a Path to the Future," Proceedings of the IEEE (May 2012), http://www.mucar3.de/bib/thlu/luettel2012ieeeproc.pdf.
[14] *Id*.
[15] John H. Holland, ADAPTATION IN NATURAL AND ARTIFICIAL SYSTEMS (1993)(first published 1975). *See also*, M. Mitchell, COMPLEXITY: A GUIDED TOUR 127 *et seq.* (2009).
[16] COMPLEXITY: A GUIDED TOUR at 142. I wonder if the human winner in accepting the prize really gave credit where credit was due?

climbing capabilities, various sensors, and the like. The modules can separate and reconfigure to make new overall combinations in response to the constraints of the physical environment.[17] They will combine locomotive power to climb hills, and use airborne sensors to enable communications among other modules and to scan the environment. Relatively old technologies such as neural nets are used to test for potential solutions; increasingly concentrated computing power allows small robots to distinguish the environments and produce reasonable results. For robots operating in various environments, "neuro-evolutionary navigation provided better overall behavior than rule-based navigation…."[18] Others creating adaptive robotic behaviors expressly term their work "evolutionary robotics,"[19] which is meant to suggest code designed by code in response to constraints such as task allocation and the environment. Here's the link between genetic algorithms and unpredictable 'emergent' behavior:

> The paper describes the results of the evolutionary development of a real, neural-network driven mobile robot. The evolutionary approach to the development of neural controllers for autonomous agents has been successfully used by many researchers, but most -if not all- studies have been carried out with computer simulations. Instead, in this research the whole evolutionary process takes places entirely on a real robot without human intervention. Although the experiments described here tackle a simple task of navigation and obstacle avoidance, we show a number of emergent phenomena that are characteristic of autonomous agents. ….
>
> We have neither pre-designed the behaviors of the robot, nor have intervened during evolution. The robot itself and alone has developed -starting from a sort of *tabula rasa* - a set of strategies and behaviors as a result of the adaptation to the environment and its own body. Despite its simple components and the simple survival criterion, it is difficult to control and predict the robot behavior, due to the non-linearities and feedback connections exploited for optimal navigation and obstacle avoidance.[20]

Robot autonomy of course exists across a spectrum. Less autonomous robots have been mentioned above. More autonomous robots rearrange logical or physical modules on the fly, as it were, to solve the assigned task. Even more autonomous robots create the modules needed from smaller units, programming themselves; and they do so in response to different types of

---

[17] N. Mathews, et al., "Spatially Targeted Communication and Self-Assembly," 2012 IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (2012), http://code.ulb.ac.be/dbfiles/MatChrOgrDor2012iros.pdf; Yan Meng, "Autonomous Self-Reconfiguration of Modular Robots by Evolving a Hierarchical Mechanochemical Model" http://www.soft-computing.de/CIM_MR.pdf ("self-reconfiguration of modular robots under changing environments … inspired by the embryonic development of multi-cellular organisms and chemical morphogenesis …. Self-reconfigurable modular robots are autonomous robots with a variable morphology, where they are able to deliberately change their own shapes by reorganizing the connectivity of their modules to adapt to new environments, perform new tasks, or recover from damages"); Marco Dorigo, "Swarmanoid: A Novel Concept For The Study Of Heterogeneous Robotic Swarms," IEEE ROBOTICS & AUTOMATION MAGAZINE (2012), http://www.idsia.ch/~gianni/Papers/Swarmanoid-techrep.pdf.

[18] Rodney A. Brooks, "A Robot that Walks; Emergent Behaviors from a Carefully Evolved Network," NEURAL COMPUTATION (1989, posted online March 13, 2008), http://dspace.mit.edu/handle/1721.1/6500.

[19] D. Floreano et al., "Evolution of Adaptive Behaviour in Robots by Means of Darwinian Selection," http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.10002922010.

[20] D. Floreano, et al., "Automatic Creation Of An Autonomous Agent: Genetic Evolution Of A Neural-Network Driven Robot," *FROM ANIMALS TO ANIMATS* 421-430 (1994), http://infoscience.epfl.ch/record/63866/files/floreano.sab94.pdf?version=2.

constraints, i.e., moving not only in response to proximity to objects, but also in response to heat, color, letters (e.g. highway signs) and so on; and able to react with *varying* responses (go around, go over, push obstacle, destroy obstacle; enlist help of other modules, reconfigure self to accommodate obstacle, etc.).[21] Some researchers have described these sort of autonomous robots as "underspecified."[22]

Increasing autonomy (we might call this 'machine IQ') is a function of the ability to rearrange ever smaller units or modules in response to an ever increasing series and types of constraints to generate an ever increasing series of responses.

B. Environments

The goal in all this work is simple: to produce robots which can make real time decisions in unpredictable environments all with the end of attaining some set task. Obviously, the more unpredictable the environment the greater the need for 'machine IQ,' the ability to adapt efficiently.

To be sure, software which integrates only with software has, we might say, a *sort* of environment, which is of course the other programs (including operating systems, etc.) with which it interacts. These are not trivial places, and the vast unpredictability of the software environment is commonly noted. We know that large programs cannot be wholly understood by a single human, and the range of output can never be wholly predicted even in the relatively organized and understood software environment of a 2013 personal computer, not to speak of the interactions of multiple machines.[23] But there are limitations to the complexity of the

[21] "The Self-Organising Incremental Neural Network (SOINN) is an algorithm that allows robots to use their knowledge – or what they already know – to infer how to complete tasks they have been told to do. In a laboratory demonstration, the machine begins to break down the task into a series of skills that it has been taught: holding a cup, holding a bottle, pouring water from a bottle, placing a cup down. Without special programmes for water-serving, the robot works out the order of the actions required to complete the task. In a separate experiment, SOINN is used to power machines to search the internet for information on what something looks like, or what a particular word might mean." http://www.timesofmalta.com/articles/view/20111017/world/Japanese-scientist-unveils-thinking-self-teaching-robot.389515

[22] "Unlike an industrial robot, the tasks of a service robot are frequently underspecified, ie, not predefined completely, because users usually provide underspecified descriptions about their intentions (eg, tasks) and the environments are typically unpredictable and dynamic. Of course, one can choose to develop service robots of which the tasks are defined completely in advance. But this choice means that the robots have no suffient [sic] capability to response/adapt to their unpredictable and dynamic environments, as well as the users." Xiaoping Chen, et al., "Developing High-Level Cognitive Functions for Service Robots," PROC. OF 9TH INT.CONF. ON AUTONOMOUS AGENTS AND MULTIAGENT SYSTEMS (AAMAS 2010)(notes omitted).

[23] A nice example is the "flash crash" of May 6, 2010 on the New York Stock Market when a trading program interacted in unpredictable ways with other such programs to create stunning losses (most of which were recovered within minutes). http://www.npr.org/blogs/money/2010/10/01/130272516/the-flash-crash-explained.
"U.S.-based equity products experienced an extraordinarily rapid decline and recovery. That afternoon, major equity indices in both the futures and securities markets, each already down over 4% from their prior-day close, suddenly plummeted a further 5-6% in a matter of minutes before rebounding almost as quickly. Many of the almost 8,000 individual equity securities and exchange traded funds ("ETFs") traded that day suffered similar price declines and reversals within a short period of time, falling 5%, 10% or even 15% before recovering most, if not all, of their losses. However, some equities experienced even more severe price moves, both up and down. Over 20,000 trades across more than 300 securities were executed at prices more than 60% away from their values just moments before. Moreover, many of these trades were executed at prices of a penny or less, or as high as $100,000, before prices of those securities returned to their "pre-crash" levels." http://www.sec.gov/news/studies/2010/marketevents-report.pdf

interactions among programs. Data usually come in types, and are often accepted only if the timing is right; there are built-in expectations of what is acceptable, and software efforts outside those constraints are often sidelined in order to protect the integrity of the software. Of course every program has bugs and anything can eventually be exploited, but the normal functioning of software interactions is measured, and controlled, and the sources of unpredictability are limited. Disembodied software is exposed to limited types of teachable moments.

Robots (embodied software), on the other hand, operate in the physical environment, and may be exposed to a more highly varied types of inputs which must be accounted for as they occur- the pace of the world cannot be dismissed as inconvenient. It is this rich set of unpredictable real time data which presents the challenges for robots and creates the desire for autonomy.[24]

The line between robot and environment, though, is in the abstract as vague as the line between one program and another. It is a matter of convenience and convention where we draw the line between program and its 'external' constraints, between the program on the one hand, and the sources of inputs and destination of output on the other hand. Modules make up modules, not just in software but for recombinant modular robots as well. (One man's program is another's subroutine.) In the purely software context, we might think of a *system*, which is an arbitrarily composed of a group of algorithms, and is so distinguished from every other source of input or direction for output. The larger and more complex the system the more likely it is to have the tools to solve a problem and the more likely it is that one might call it "intelligent." Small modules, neurons, subroutines won't qualify, but many of them interacting might.[25]

Turning to robots, the greater the porousness between a robot and its environment—which is a measure of the degree of its ability to interrelate with and affect the environment—the more likely we might term it intelligent, the greater the likelihood it and its environment together can be thought of as a system. In this way, names are just theories of systems, whether those names be subroutine labels, human family and clan names, or other sets.

What counts as an environment is arbitrary. Highly porous robots in effect create larger systems.[26] As the environment changes, so does the robot's neural activities; indeed, one way to "poison" a robot is to interfere with its on-the-job training as it seeks to make patterns from

---

[24] Saddek Bensalem, et al., "Toward a More Dependable Software Architecture for Autonomous Robots," http://homepages.laas.fr/felix/publis-pdf/ieee-ram-ser08.pdf.

[25] *See generally*, Marvin Minsky, THE SOCIETY OF MIND (1985).

[26] An important perspective on the unpredictable results arising from the complex feedback loops we expect from autonomous robots is provided by Charles Perrow. Focusing on systems such as nuclear reactors, maritime accidents and others, he suggests that when unpredictable interactions (including feedback loops) among subsystems is permitted we may see so-called complex interactions. Charles Perrow, NORMAL ACCIDENTS 75 (1999). In systems where the couplings among subsystems is also tight, i.e. where output of one results directly in a change in another, the risk of so-called normal accidents may be very high. I do not suggest that dynamic complex tightly coupled systems are tantamount to autonomous machine intelligence—there is nothing intelligent about the nuclear reactor accidents and bizarre ship collisions Perrow describes—but complexity does in both situations underlie unpredictability. As Perrow notes, inserting buffers in these systems such that they are more loosely coupled helps prevent accidents. A similar functionality may be useful in avoiding unpleasant unpredictable effects of autonomous robots. Below in § III, I term this buffering "common sense."

instances in the environment by substituting in misleading training data—that is, faking the environment.[27]

A recent article notes the intimate connection between true autonomy, embodiment and the physical environment:

> The ER [Evolutionary Robotics] approach emphasizes agent's embodiment, which means that an emerging behavior is not only dependent on various properties of the actual robot such as its size, speed, degrees of freedom, sensors and actuators, but also on the environment with which a robot interacts . ER is an excellent technique that allows us to create artificial control systems that autonomously develop their skills in close interaction with the environment and that exploit very simple, but extremely powerful sensory-motor coordination.[28]

That larger system (of robot and environment) is not more predictable than the unpredictable environment that fathered it; it must be less predictable. These porous, adaptive robots—autonomous robots, in short—are the subject of the next sections. These robots by definition will take unpredictable actions in the physical world that it shares with us humans in the pursuit of the tasks that we assign, and sometimes (in the context of larger systems) in the context of the tasks that it has assigned to itself.

II. *Law*

People and companies can sue each other for money when one is 'liable' to the other, having breached some sort of norm. The norms are generally found in generally three areas of the law: contract, tort, and statutes. These areas overlap. Contract law finds its norms in the promises people make directly to one another. Tort law generates norms which apply to all, strangers or not. In this way tort law is like criminal law: Joe murdering or hitting Bob violates criminal law, regardless of their prior interactions, and similarly Bob (or his personal representative if he is dead) can sue Bob in tort for money damages. Most states have, in addition, statutes which allow people to sue to fulfill important state policies. Thus for example, California has its Business and Professions Code § 17200 allowing one person, sometimes on behalf of many people, to sue for unfair or fraudulent business practices, such as false advertising.

As a result, most property damage and injuries done to people with whom one does not have a preexisting relationship are subject to suit under tort law. Even injuries done to one with whom there is a contract, such as medial malpractice, can be the subject of tort law, and tort law applies to most cases involving some kind of injury even between people who have other types of preexisting relationships, such as employers and employees, family members, and consumers and sellers of products. The interesting legal issues with respect to addressing injuries and damage caused by robots, then, are likely to arise under tort law.

---

[27] Alex Armstrong "Poison Attacks Against Machine Learning," (19 July 2012), http://www.i-programmer.info/news/105-artificial-intelligence/4526-poison-attacks-against-machine-learning.html.

[28] Martin Peniak, "Active Vision For Navigating Unknown Environments: An Evolutionary Robotics Approach For Space Research" (2012)(available via psu.edu)(notes omitted). The irony is not lost on me that the actual subject of this paper is a virtual (i.e. disembodied) robot. The environment and robot are literally one. (But do however note the use in this system of genetic algorithms.)

Tort law is, in turn, commonly divided into two areas- negligence and strict liability.[29] These two areas are part of our common law tradition, inherited from England, by which judges over the ages slowly modify the rules, sometimes expressly and sometimes not, in responses to economic, cultural, and other developments,[30] as well as in response to tugs from the Legislature which steps in from time to time to brake or push movement in the law.[31]

A negligent action is something a reasonably prudent person would not do. For example, failing to brake while distracted by a cell phone qualifies as negligence.

> Negligence is the failure to use reasonable care to prevent harm to oneself or to others. A person can be negligent by acting or by failing to act. A person is negligent if he or she does something that a reasonably careful person would not do in the same situation or fails to do something that a reasonably careful person would do in the same situation.[32]

Strict liability is sometimes thought of as liability 'no matter what,' although this is, as we will see, not quite right. But it is liability without negligence or other "fault," and usually applies only to products (things like cars and pills and metal hip joints). If these are defective then, regardless of whether or not the maker exercised 'due care' or was negligent, the consumer generally can sue.

A. *Negligence*

Offhand, most lawyers probably believe that negligence belongs to the distant past, with strict liability the recent development to account for the results of industrialization and the ability to share risk through insurance and pricing across a large number of sales. Strict liability actually came first, albeit in a different guise. Common laws suits had earlier been based on the status, or relationship, of the parties, and a violation of the concomitant duty that defined the status. So for example sheriffs had by virtue of their status duties to restrain and arrest malefactors, and having let one go (no matter if negligent or not) were liable for the resultant damages.[33] Only in the 19th century did the standard of carelessness erupt as the core concept behind negligence.[34] This correlates with the increasing incidents of strangers coming (literally) into collision, as opposed to disputes erupting between people who had a previously defined relationship through status. So, for example, we had the 'running down' cases, in effect car collisions, which could no longer be decided by the pre-existing status or relationship of the parties (master-servant etc.).[35] Not until Oliver Wendell Holmes in the mid-19th century do we use *prediction* to analyze the scope of tort liability.[36] Prediction theory is now central to negligence. Carelessness or negligence is

---

[29] *See generally, Merrill v. Navegar, Inc.,* 26 Cal. 4th 465, 478 (2001).
[30] Morton J. Horwitz, THE TRANSFORMATION OF AMERICAN LAW 1780-1860 (1977); Morton J. Horwitz, THE TRANSFORMATION OF AMERICAN LAW 1870-1960 (1992).
[31] For example, statutes can be used as evidence of negligence "per se," that is, the breach of the duty is established by violating the statute. E.g., *Norman v. Life Care Centers of Am., Inc.,* 107 Cal.App.4th 1233, 1240 (2003).
[32] California Civil Jury Instructions ("CACI") 401.
[33] Morton J. Horwitz, THE TRANSFORMATION OF AMERICAN LAW 1780-1860 at 87 *et seq.* (1977).
[34] Patrick J. Kelley," Proximate Cause in Negligence Law: History, Theory, and the Present Darkness," 69 WASH. U. L.Q. 49, 55 (1991) (as of "mid-nineteenth century .. the newly-developed negligence cause of action").
[35] Horwitz, note 33 at 94 *et seq.*
[36] Morton J. Horwitz, THE TRANSFORMATION OF AMERICAN LAW 1870-1960 at 56 (1992).

made out when a reasonably prudent person "ought to have known" that injury would result from the action.

> "Foreseeability ' "is not to be measured by what is more probable than not, but includes whatever is likely enough in the setting of modern life that a reasonably thoughtful [person] would take account of it in guiding practical conduct." [Citation.] One may be held accountable for creating even " 'the risk of a slight possibility of injury if a reasonably prudent [person] would not do so.'"[37]

There have long been arguments on the what *sort* of injury need be predictable for negligence to apply. For example, one might anticipate a hurt foot if one leaves a rusty nail on the floor, and so be liable for injury to a stranger's foot when he steps on the nail. But one might not have reasonably predicted that having tripped over a rusty nail a person might fall, arms flung out, ripping out a pipe which, flung across the room, ignites a fire which burns the barn down. The legal fight had been, that is, over whether the actual bad result had to be foreseeable, or whether, having done something which foreseeably could result in *some* sort of harm, one might be liable for any and all harm that as a matter of fact resulted.[38] The fight isn't quite over. One classic authority suggests a sort of balancing test: as the gravity of harm increases, the likelihood of its happening may decrease and yet be sufficiently foreseeable as to make the defendant liable.[39] Every act presents *some* risk: the question is always, in some fashion, whether it presents an "unreasonable" risk.[40] The answer is that foreseeability of the *actual type of harm* is usually a predicate for negligence liability:

> Some negligence cases impose liability only where the type of risk that was foreseeable to the defendant actually occurred. If the defendant's negligence causes harm by fire, he is liable if he could foresee the risk of fire, but not otherwise. Some courts do this by using foresight as the test of proximate cause. Others use a duty analysis, reasoning that there is no duty to protect a foreseeable plaintiff from an unforeseeable risk of harm. The result is the same under either approach.[41]

So, while one need not have foreseen the precise manner in which harm occurred, one must have been able to foresee at least the "kind of harm" that occurred.[42]

---

[37] *Constance B. v. State of California*, 178 Cal.App.3d 200, 206 (1986).

[38] No note on negligence is complete without a cite to *Palsgraf*. Here it is: *Palsgraf v. Long Island R. Co.,* 248 N.Y. 339, 162 N.E. 99 (1928).

[39] Wm. Prosser, LAW OF TORTS 146 (4th ed. 1971).

[40] *Id*. at 146-47.

[41] David A. Fischer, "Products Liability-Proximate Cause, Intervening Cause, and Duty," 52 MO. L. REV. 547, 550-51 (1987)(notes omitted). *See also*, 4 F. Harper, et al., THE LAW OF TORTS § 20.5 (2d ed. 1986)(test of foreseeability); *Robison v. Six Flags Theme Parks Inc.,* 64 Cal.App.4th 1294, 1297 (1998)(proper focus is on the foreseeability of a harmful event of at least the general type that occurred); *Brewer v. Teano,* 40 Cal App.4th 1024, 1030 (1995)(test is whether the "category of negligent conduct at issue is sufficiently likely to result in the *kind* of harm experienced")(emphasis supplied).

[42] *Bryant v. Glastetter*, 32 Cal.App.4th 770, 780 (1995). Foreseeability is more intricately connected with negligence than I make out here; it's probably wrapped up in related issues of duty, proximate cause, and so on. *E.g.,* D. Owens, "Figuring Foreseeability," 44 WAKE FOREST L.REV. 1277 (2009). It is enough for my purposes here to show that foreseeability, in the way discussed in the text, is essential to negligence liability.

B. *Strict Liability*

There are roughly four theories of strict liability: ultrahazardous activity, and three brands of products liability (failure to warn, design defect, and manufacturing defect).[43] A superficial reading might suggest that foreseeability is not pertinent here: after all, the liability is "strict" and imposed just if a product is defective; it matters not if the defect was intentionally or negligently built. But the truth is that foreseeability plays an important role even in strict liability cases.[44]

Modern strict liability law arose in the last century, in response to the fact that contract law—itself then a recent invention[45]-- was unable to provide a remedy for the proliferation of dangerous products sold through multiple layers of a distribution chain. A car, for example, is not sold by the manufacturer directly to the consumer- it is sold to a distributor first. Many distribution chains are even longer, and pretermit the existence of a contract between the entity responsible for the dangerous item and the person injured. Negligence might be available, but would present some difficulty when applied to most of the chain of distribution: after all, what should a car dealer know of the intricacies of car design, or an importer of the chemical makeup and risks of foreign candy? In an age of mass markets and long distribution chains, costs could be allocated across a large number of sales, and manufacturers were in a position accordingly to spread costs including by purchasing insurance. Why not similarly spread the costs of injury? Surely the innocent consumer should not bear the cost of a bad product. So the courts devised the doctrines of strict liability, imposing it on all members of the distribution chain, whether or not they were at "fault" in the sense of being negligent, and whether or not a contract existed between them and the ultimate consumer who was injured. The essence of the move to strict products liability was to focus on the condition of *the product itself*—was it dangerous?—and away from an evaluation of the defendants' *conduct* in making the product—the realm of negligence—because the innocent victim of a dangerous product should be compensated, even if the defendants were not negligent in making it.[46]

Below, I first briefly outline the three types of products liability (failure to warn, design defect, manufacturing defect).[47] I will then discuss and dispose of ultrahazardous activity before returning to the three types of products liability to determine their application to the unpredictable actions of robots.

---

[43] *See generally*, *Barker v. Lull Eng'g Co.,* 20 Cal.3d 413, 418 (1978); *Anderson v. Owens-Corning Fiberglas Corp.,* 53 Cal.3d 987, 995 (1991).

[44] *E.g*. 3 Harper et al., THE LAW OF TORTS § 14.15 at 329 *et seq.* (2d. ed. 1986).

[45] Modern contract law dates back to the late 19th century, arguably devised by Dean Langdell at Harvard Law School and developed by Oliver Wendell Holmes, Jr. Grant Gilmore, THE DEATH OF CONTRACT (1974). Holmes developed the notion that the exchange of consideration ('consideration' is the thing you give to, or do for, the other person involved in the contract) had to be the *product of a bargain* to create an enforceable promise. "Before 1871contract law was a loose confederation of subspecialties such as negotiable instruments and sales that was not yet a "systematically organized, sharply differentiated, body of law."" Robert Gordon, "Book Review of The Death Of Contract," Faculty Scholarship Series, Paper 1376, http://digitalcommons.law.yale.edu/fss_papers/1376.

[46] *See generally, Barker v. Lull Eng'g Co.,* 20 Cal.3d 413, 434 (1978).

[47] *See generally*, *Barker v. Lull Eng'g Co.,* 20 Cal.3d 413, 418 (1978); *Perez v. VAS S.p.A.,* 188 Cal.App.4th 658, 676 (2010); *Anderson v. Owens-Corning Fiberglas Corp.,* 53 Cal.3d 987, 995 (1991).

The first sort of product liability is *manufacturing defect:* that is, when the product departs from intended design. An error in the factory causes a cog to be defectively made, or a part to be missing—if someone is hurt as a result, this theory will provide relief.[48]

Second, we have *design defect*- foreseeable harm could have been avoided by alternative design. Readers may recall the Ford Pinto, designed to have a relatively fragile gas tank in the rear just where it would explode when hit from behind. Every one of those cars was made just as designed- but the design itself was bad. How do we tell if a design is bad? After all, every product can be made in many ways, and manufacturers have to balance a wide variety of factors as they make their products—they cannot always use titanium or diamond to make the product invulnerable to breakage; and there is such a thing as an overdesigned product—which no one can afford. The analysis is often based on risk/benefit analysis: "a product is defective in design either (1) if the product has failed to perform as safely as an ordinary consumer would expect when used in an intended or reasonably foreseeable manner, or (2) if, in light of the relevant factors …, the benefits of the challenged design do not outweigh the risk of danger inherent in such design."[49]

Thirdly, we have a *failure to warn*. Here, the distribution chain is liable if the consumer is not adequately warned on the uses or risk of the product. The case might focus on the directions for use, user manuals, labeling, advertising, and so on. Many very ordinary products, such as chain saws, lawn mowers, drugs, and hot tub drains can cause injury if not used correctly, and manufacturers might be liable if they did not adequately warn of those dangers.

Liability for ultrahazardous activity is in a way quite the opposite of product liability, and does not really focus on the defect of a product, but rather, as the name implies, a course of conduct. The classic ultrahazardous conducts are blasting, transporting nuclear materials, and the like. These activities and others are considered so inherently dangerous that the law makes those engaging in them in effect insurers to others who are hurt, without requiring proof of negligence or other types of fault. The point is not to focus on mass market items or actions, but to isolate rare activities of dubious social utility. An activity may be deemed ultrahazardous if it is uncommon, poses a high risk of harm, and creates a high degree of injury when injury does occur.[50] What counts as ultrahazardous changes over time as we become accustomed to the activity, as it spreads, and presumably as it becomes safer. So, for example, we have this wonderful observation from the 1930s:

---

[48] For example, the one soda bottle in ten thousand that explodes without explanation. *Escola v. Coca Cola Bottling Co.*, 24 Cal.2d 453 (1944).

[49] *Barker v. Lull Eng'g Co.*, 20 Cal. 3d 413, 418 (1978)("This dual standard for design defect assures an injured plaintiff protection from products that either fall below ordinary consumer expectations as to safety, or that, on balance, are not as safely designed as they should be. At the same time, the standard permits a manufacturer who has marketed a product which satisfies ordinary consumer expectations to demonstrate the relative complexity of design decisions and the trade-offs that are frequently required in the adoption of alternative designs. Finally, this test reflects our continued adherence to the principle that, in a product liability action, the trier of fact must focus on the product, not on the manufacturer's conduct, and that the plaintiff need not prove that the manufacturer acted unreasonably or negligently in order to prevail in such an action."

[50] RESTATEMENT (SECOND) OF TORTS § 520; RESTATEMENT (THIRD) OF TORTS: Phys. & Emot. Harm § 20 (2010); 6 Witkin, SUMMARY OF CALIFORNIA LAW, Torts, § 1416 at 841 (10th ed. 2005).

> Thus, aviation in its present stage of development is ultrahazardous because even the best constructed and maintained aeroplane is so incapable of complete control that flying creates a risk that the plane even though carefully constructed, maintained and operated, may crash to the injury of persons, structures and chattels on the land over which the flight is made.[51]

While it is possible that the use of robots might be considered ultrahazardous, especially at their introduction, ultrahazardous theory is not well suited to the imposition of liability. Many of the injuries may not be serious, and robots are not likely *routinely* to pose hazards. As we will see with the other three types of strict liability, foreseeability runs a line through this doctrine of ultrahazardous as well: the essence of the liability is that harm is especially likely to happen, harm that the defendant knows about, or should have known about;[52] that is, predictable. Unless we are willing to dub all robotic actions foreseeably dangerous because *some* are unforeseeably dangerous, this doctrine will not assist.

Let us now turn back to the first three theories, the three types of product liability. These too have foreseeability as an essential characteristic.

> Strict products liability cases employ all three types of limitations on liability. That is, courts commonly require the plaintiff, the type of harm, and the manner of harm to be foreseeable. Some courts may use all of these limits, while others may formally use only one or two. …[T]here is such a close relationship between the three limits that there may be little practical difference in the scope of liability between a court that uses all three of these limitations and a court that uses only two of them.[53]

> Strict products liability is predicated on the existence of an unreasonably dangerous product whose *foreseeable* use has caused injury.[54]

Design defect is, as noted above, based on balancing factors, the pros and cons of the design, as those were understood at the time the product was made. "The relevant factors include "the gravity of the danger posed by the challenged design, the likelihood that such danger would occur, the mechanical feasibility of a safer alternative design, the financial cost of an improved design, and the adverse consequences to the product and to the consumer that would result from an alternative design."[55] Inhering in each of these factors is foreseeability: of the harm, and of the benefits.

Similarly with manufacturer defect. Again, we do not focus on the reasonableness of the defendants' conduct—that's negligence talk—but rather the risks posed by the products as actually manufactured. And here too there is an underlying predicate of foreseeability, because whether the entire line of products is dangerous (design defect) or one of them is (manufacturing

---

[51] RESTATEMENT (FIRST) OF TORTS § 520 (1938), comment on clause a.
[52] RESTATEMENT (THIRD) OF TORTS: Phys. & Emot. Harm § 20 (2010), comment i (foreseeability).
[53] David A. Fischer, "Products Liability-Proximate Cause, Intervening Cause, and Duty," 52 MO. L. REV. 547, 553 (1987).
[54] Steven J. Frank, "Tort Adjudication and the Emergence of Artificial Intelligence Software," 21 SUFFOLK U. L. REV. 623, 637 (1987)(emphasis supplied).
[55] *Barker v. Lull Engineering Co., supra,* at 431; *Torres v. Xomox Corp.,* 49 Cal.App.4th 1, 15-16 (1996).

defect), the issue is whether there is risk when used as intended, that is, when foreseeably used.[56]

But we must be a little careful: design and manufacturing defect are not the same thing in the sense that the balancing act done as of the time of manufacture does not, of course, apply in the manufacturing defect scenario, and the foreseeability element attendant to that is not part of the manufacturing defect test. But it is also true that, as with ultrahazardous activity, it would be entirely illogical to apply manufacturing defect liability to autonomous robots which, as they come off the assembly line (as it were) are all exactly the same and for purposes of my argument here, conform to design—at least as they are when delivered into the hands of the consumer. True, they will rapidly depart from a standard as they conduct their on-the-job learning, but strict liability cases do not impose liability on the manufacturer and other parts of the distribution chain for changes made to the product after delivery to the consumer unless those changes were foreseeable;[57] which, because we speak of autonomous robots here, they are not.

So we turn to the final strict liability tort theory of *failure to warn.* The doctrine requires fair warning of the risk posed by the product. The theory is premised on unequal access to information, i.e., that the manufacturer (for example) knows more about the risks than a relatively unsuspecting consumer.[58] This is equivalent to a premise that the manufacturer is able to predict, in a way that the consumer is not, the effects of the product at issue. But where the product is by definition acting in an unpredictable way, the application of this theory of liability may be impossible. Defendants probably are not liable for "unknowable" risks.[59] Liability depends on what the defendant knew or should have known at the time the product was sold,[60] and so the 'failure to warn' doctrine depends on the foreseeability of the harm which the instructions would have avoided. It is illogical to impose strict liability on the manufacturer for unknowable hazards when a core rationale for the imposition of liability is the ability to account in advance for the costs of injury and obtain insurance.[61]

C. *Summary on foreseeability*

The basic theories of tort law predicate liability on foreseeability, by which I do not mean foreseeability in the general sense but rather a type of predictable harm to a predictable group of potential victims. This is true whether one looks at negligence or strict liability. Those theories of strict liability which depend the least on foreseeability—ultrahazardous activity and manufacturing defect—are for other reasons poorly suited to injury caused by autonomous robots.

---

[56] *Cronin v. J.B.E. Olson Corp.,* 8 Cal. 3d 121, 130 (1972), citing *Greenman v. Yuba Power Products, Inc.*, 59 Cal. 2d 57 (1963).

[57] For example, if a machine is deigned for use with asbestos—if asbestos use is inevitable—then the machine's maker may be liable for causing asbestos-related disease even though the asbestos is added after the sale of the machine. *Shields v. Hennessy Indus., Inc.,* 205 Cal.App.4th 782, 797 (2012).

[58] *E.g., Johnson v. Am. Standard, Inc.,* 43 Cal.4th 56, 65 (2008).

[59] *Anderson v. Owens-Corning Fiberglas Corp.,* 53 Cal.3d 987, 1000 (1991); *Carlin v. Superior Court*, 13 Cal.4th 1104, 1118 & n.8 (1996); *Oakes v. E.I. Du Pont de Nemours & Co., Inc.*, 272 Cal.App.2d 645, 650–651 (1969).

[60] *Vermeulen v. Superior Court,* 204 Cal.App.3d 1192, 1203 (1988).

[61] *Taylor v. Elliott Turbomachinery Co., Inc.,* 171 Cal.App.4th 564, 596 (2009)(citing *Anderson, supra* n.59).

Foreseeability can mean many things. Sometimes we have enough information to predict an event exactly, roughly, or within certain bounds, or to forecast the odds of an event. Forecasting might predict the likelihood (e.g., 50%) of a thunder storm or coin flip, and Bayesian equations might tell us that, given certain assumptions, the likelihood of having a disease given a test for it (that has some known error rate).[62]

Some predictions have a low likelihood, others a higher likelihood; in common parlance, even a very, very unlikely event might be "predictable" in the sense that we know that, eventually, the event will occur, such as thirty coin flips in a row that all come up heads (given enough coin flips). The law does not mean this when it uses 'foreseeable'; it refers to the sort of future events we think it's reasonable to have people guard against. The law is prophylactic. It is moral for this reason (among other things): it punishes only transgressions which (it presumes) people are actually capable of avoiding. When we refer to risks the defendant "knew or should have known" about, we mean to insist that the defendant should know certain things before operating in the arena he did. You cannot drive a car unless you know the basics of braking and acceleration; you cannot act as a doctor or lawyer or architect unless you know the basics of those professions, and having failed to do so, you will be punished even if, as a matter of fact, you did not have the knowledge that would have allowed you to do so.[63]

Negligence and strict liability were born and raised in a Newtonian universe, the universe of billiard balls hitting billiard balls, car hitting cars; force, mass and reaction; and machinery executing one step at a time. The risks discernible in this world are the consequences of Newtonian mechanics, which is linear: A causing B causing C. In this world, the more knowledge we have, the more we can predict; and if we knew every last detail of every last particle in the universe we could mechanistically unravel time backwards to see where everything was in the past, and ravel time forward—here's the *foreseeability*—to see where everything will be at any arbitrary time. The old legal fights—which still continue—are usually about how far to take foreseeability in this sense: how remote or trivial do we want to push it; how low the odds before the event is no longer reasonably foreseeable, how willing we are to have burdensome preventative measures, and so on.[64]

In this note, however, we don't care about those disputes, or whether our confidence in a predicted result is 5% or 95%, or whether our forecast of the odds of an event should be 90% or 45%. With autonomous robots which are complex machines, ever more complex as they interact seamless, porously, with the larger environment, linear causation gives way to complex, nonlinear interactions. These robots interact with an environment "that is much larger and more complex than the system [here, an autonomous robot] itself,"[65] In these interactions we find as a

---

[62] For more on Bayesian probability, some examples, and a series of associated fallacies, see my "Statistics In Law: Bad Inferences & Uncommon Sense" (2011), http://works.bepress.com/curtis_karnow/

[63] *Howard v. Omni Hotels Mgmt. Corp.,* 203 Cal. App. 4th 403, 429 (2012)(industry and professional standards on what counts as foreseeable). *See also* the 'sophisticated user' defense in which a plaintiff who as a result of his work knows, or ought to know, of the dangers of a product, may be hampered or defeated in his suit against the distributor of the product. *E.g., Chavez v. Glock, Inc*., 207 Cal. App. 4th 1283, 1301 (2012).

[64] D. Owens, "Figuring Foreseeability," 44 WAKE FOREST L.REV. 1277 (2009).

[65] Sidney Dekker, "In The System View of Human Factors, Who Is Accountable for Failure and Success?," Human Factors And Ergonomics Society Europe Chapter Annual Meeting (2009). Dekker and others pointedly contrast our usual (i.e., Newtonian) insistence on a "root cause" of events and accidents with the truth about complex systems:

matter of course "combinatorial explosions [which] can outwit people's best efforts at predicting and mitigating trouble."[66] All the knowledge in the universe about all the agents and subsystems is not enough to fix the future behavior of these systems. The problem is not ignorance; the problem is the limits of knowledge:

> When, however, organizations are recognized as Complex Adaptive Systems (CAS), surprise is not necessarily the result of bounded rationality, limited information or systems design, but often is the result of the fundamental nature of the system in question. Complexity theory suggests that much surprise is inevitable because it is part of the natural order of things and cannot be avoided, eliminated, or controlled.[67]

So the actions of autonomous robots devising their own means to attain a task may not be subject to liability under any of theories of tort, to the extent robots are adaptive, i.e., able to interact with the world, physical and otherwise, through a wide variety of input [e.g., sensors] and output [e.g. manipulation and locomotion] means. And these are just the circumstances in which autonomous robots are most likely to pose danger.

Contrary to the suggestions of some commentators,[68] this problem is not a general one of robots or machine intelligences, because most robots are not autonomous in the sense I have used the term. Most of these products do what they are told to do, in the way they are told to do it, or the variation of means is itself programmed and utterly predictable, which is just another way of saying the same thing. Unintended injuries are often just the result of human error[69] and poor workplace design.[70]

"Post-accident attribution accident to a 'root cause' is fundamentally wrong. Because overt failure requires multiple faults, there is no isolated 'cause' of an accident. There are multiple contributors to accidents. Each of these is necessary insufficient in itself to create an accident. Only jointly are these causes sufficient to create an accident. Indeed, it is the linking of these causes together that creates the circumstances required for the accident. Thus, no isolation of the 'root cause' of an accident is possible. The evaluations based on such reasoning as 'root cause' do not reflect a technical understanding of the nature of failure but rather the social, cultural need to blame specific, localized forces or events for outcomes." Richard I. Cook, "How Complex Systems Fail," http://www.ctlab.org/documents/How%20Complex%20Systems%20Fail.pdf (referred to me by Bruce Schneier's CRYPTO-GRAM, a fecund source of insight on risk, security, and related issues <http://www.schneier.com/crypto-gram.html>).

[66] S. Dekker, *ibid*.

[67] Reuben R. McDaniel, Jr., et al., eds., UNCERTAINTY AND SURPRISE IN COMPLEX SYSTEMS § 1.5 (2005).

[68] See note 4 above. *See generally*, Gary E. Marchant, et al.," International Governance of Autonomous Military Robots," 12 COLUM. SCI. & TECH. L. REV. 272, 281, 283-84 (2011)(does not appear to clearly distinguish unpredictability as a function of (i) large complex software systems or (ii) autonomy); Steven J. Frank, "Tort Adjudication and the Emergence of Artificial Intelligence Software," 21 SUFFOLK U. L. REV. 623, 639 (1987).

[69] *See* above § I, discussion of Aegis missile strike on civilian aircraft.

[70] Jiang, B.C. and Gainer, C.A., Jr., "A cause-and-effect analysis of robot accidents," 9 JOURNAL OF OCCUPATIONAL ACCIDENTS 27-45 (1987)("Pinch-point accidents accounted for 56% of all accidents while impact accidents accounted for 44%. Most accidents were caused by poor workplace design (20 of 32 accidents) and human error (13 of 32 accidents)".) *See also,* "Lack of expert data sinks Pa. Suit over 'robotic' surgery," 6 ANDREWS EXPERT & SCI. EVIDENCE LITIG. REP. 11 (2009)(injury from da Vinci surgical robot).

III. *Future Reciprocal Accommodation*

I suggest two types of developments are likely to ameliorate the problem of using traditional tort law to examine the acts of some autonomous robots. The first has to do with robots, the second, with humans.

To account for the unexpected, robots will do unexpected things, but their actions, like ours, need to be constrained by common sense, just as the ideal "reasonable person" provides the vantage point from which we determine the issue of negligence. Common sense is the knowledge we spend a lifetime learning (and unlearning); put simply, it is the general body of knowledge we have about the world and the way its pieces interact. Common sense critically includes awareness of the physical world and its mechanisms, but common sense is also famously wrong; to be generally reliable, narratives and theories about how the world works must be able to be tested empirically and so modifiable. That is to say, we want robots to believe things, stories, series of facts, and so on, even if (like us) they have no direct access to those facts (such as the existence of atoms and bacteria, the temperature of a star, or the fact that on a hot day I would prefer a cold drink to hot tea). This is what it means to be able to act on inadequate, or underspecified, information; we humans do it all the time. And we want robots to update their common sense when they appear to be in error (perhaps they may do this better than we can[71]). The minimum requirements, then, seem to be a very large database as well as highly porous interactions with the physical environment to permit learning. A number of researchers are working on exactly these databases.[72]

Common sense may provide the buffer that dynamic complex systems need in order to avoid disaster. As suggested above at note 26, tightly coupled systems can be dangerous:

> Interactive complexity refers to the presence of unfamiliar or unplanned and unexpected sequences of events in a system that are either not visible or not immediately comprehensible. A tightly coupled system is one that is highly interdependent: Each part of the system is tightly linked to many other parts and therefore a change in one part can rapidly affect the status of other parts. Tightly coupled systems respond quickly to perturbations, but this response may be disastrous. Loosely coupled or decoupled systems have fewer or less tight links between parts and therefore are able to absorb failures or unplanned behavior without destabilization.[73]

---

[71] I refer to the persistent cognitive fallacies under which we operate. E.g., D. Kahneman's indispensable THINKING, FAST AND SLOW (2011); L. Mlodinow, THE DRUNKARD'S WALK (2008); D. Ariely, PREDICTABLY IRRATIONAL (rev. ed. 2010).

[72] C. Havasi, et al, "Digital Intuition: Applying Common Sense Using Dimensionality Reduction." 24 INTELLIGENT SYSTEMS, IEEE 24-35 (2009)(re Open Mind Common Sense [OMCS] project); D. Lenat, "CYC: A Large-Scale Investment in Knowledge Infrastructure," 38 COMMUNICATIONS OF THE ACM 33-38 (No. 11, 1995); Niket Tandon et al., "Deriving A Web-Scale Common Sense fact Database," (2011), https://www.mpi-inf.mpg.de/~ntandon/papers/aaai11.pdf.

[73] Karen Marais, et al, "Beyond Normal Accidents and High Reliability Organizations: The Need for an Alternative Approach to Safety in Complex Systems," http://sunnyday.mit.edu/papers/hro.pdf (discussing Charles Perrow's work).

Common sense as the buffer for machine intelligence has the effect of loosening the couplings. It may be that robotic common sense will make the acts of autonomous robots more predictable, and so make it easier to engage classic tort theories.

The second development has to do with humans becoming accustomed to the work of robots. Predictability and foreseeability are in practice vague and peculiar notions, and people with different experiences and beliefs about how the world works will treat different things as 'predictable.' In any event humans are poor at predicting odds,[74] and generally are not accurate estimating the likelihood of future events. Perhaps we may get better at predicting the behavior of autonomous robots as we interact with them; actions which appear at first random may begin to cluster in their frequencies, revealing theretofore unanticipated patterns which will help future prediction.

There are other areas of law in which developing technology itself modifies the application of the legal rules, such as with respect to privacy rights. In a test that sounds eerily reminiscent of negligence, the scope of our privacy rights depend on our reasonable expectations of privacy,[75] so that, for example, we may have such rights in our bedroom but not our front lawn, with respect to data on our home computers but not those at work, or perhaps not even our own computers when passing through customs. Obviously expectations evolve,[76] and so does the right itself (perhaps involving its own feedback loop as court opinions affect the public's expectations). In like manner, the behavior of autonomous robots is likely to affect our sense of what is predictable, and we may fold their actions into the realm of reasonable expectation. We may be aided in this by yet another cognitive fallacy, *hindsight bias*,[77] by which we overestimate the past odds of an event which has already occurred. That is, after an unlikely event takes place, we tend to believe it was more predicable than it was.[78]

So it is that through fiction we might become more comfortable with the unknown future—telling ourselves that we knew it, all along.

---

[74] *E.g.*, Charles Seife, PROOFINESS: THE DARK ARTS OF MATHEMATICAL DECEPTION 67 *et seq*. (2010); Kahneman, *supra* note 71 at 144 *et passim* (probability neglect).

[75] *E.g.,* David S. Barnhill, "Cloud Computing and Stored Communications: Another Look at Quon v. Arch Wireless," 25 BERKELEY TECH. L.J. 621 (2010).

[76] Orin S. Kerr, "The Fourth Amendment and New Technologies: Constitutional Myths and the Case for Caution," 102 MICH. L. REV. 801, 805 (2004).

[77] The fallacy is significant, and can interfere with a jury's evaluation of whether a defendant was negligent—whether the defendant should have been able to foresee the harm, for after all, the case only gets to the jury if the assertedly unpredictable result (a bad reaction to a drug, or an attack by a psychiatric patient, for example) actually takes place and injures someone. *See e.g.,* Susan LaBine et al., "Determinations of Negligence and the Hindsight Bias," 20 LAW AND HUMAN BEHAVIOR 501 (1996); Hal Arkes, et al., "Medical Malpractice v. the Business Judgement Rule: Differences in Hindsight Bias," 73 OR. L. REV. 587 (1994).

[78] *E.g.,* Nassim Taleb, FOOLED BY RANDOMNESS 56, 192 (2d ed. 2004).