FRANKENSTEIN UNBOUND
Towards a legal definition of Artificial Intelligence
Sam N. Lehman-Wilzig

The Frankenstein myth of creature turning on creator is centuries if not millennia old. But only recently under the impact of the cybernetic revolution has this fantasy entered the realm of the possible. This paper explores the legal ramifications of Artificial Intelligence (AI) with specific emphasis on "humanoid" criminality. Following a review of the actual (or theoretically proven) powers of artificially intelligent machine automata and the likely advances to be made in the future, four general categories of AI harmful behaviour are suggested, with illustrations from cybernetic research and science fiction. An analysis is made of the jurisprudential principles underlying several legal categories already existent, upon which future cybernetic law may be based.

"Things are in the saddle and ride mankind."
*Ralph Waldo Emerson

"The function of prediction is not to aid social control, but to widen the spheres of moral choice."
*Daniel Bell

WHILE SOCIETY is abuzz today with the novel and increasingly troublesome problem of human-inspired computer crime, a related yet far more profound issue may be lying ahead—the humanity-created machine as a 'criminal' in its own right. Whether it be Butler's 19th century *Etruscan* machines, Rabbi Loewe of Prague and his 16th century *Golem*, or even the ancient Daedalus, the Frankenstein complex, i.e. the fear of creature turning on its creator, has been and continues to be a major source of man's mythology and literary output. Yet yesteryear's myth may soon turn into likely future reality; we are no longer dealing with irrational nightmares, but with a probable or even inevitable phenomenon for which we seem to be socially quite ill prepared.

Dr. Sam N. Lehman-Wilzig is Lecturer, Department of Political Studies, Bar-Ilan University, Ramat Gan, Israel. This essay is a revised and expanded version of a paper presented at "Apollo Aquinata: The Humanities in a Computerized World", an international symposium sponsored by the Institute for Humanistic Studies, State University of New York at Albany, on 18-21 April 1979.
This essay considers the legal ramifications of Artificial Intelligence (AI). A survey of the literature on computers and the law reveals that AI is not a subject addressed hitherto. While the Anglo-Saxon legal tradition is generally averse to jurisprudential speculation, the widespread interest in such a potential phenomenon should at least justify a preliminary discussion of the parameters and/or directions it—and society’s response to it—could conceivably take.

Of course, one must first be convinced that “humanoid criminality” is a possibility (barring other likelihood) to which reputable experts in the field of cybernetics themselves admit. Consequently, we will briefly review the present state of the technology, arguing that the present powers of AI automata are by themselves enough to warrant this legal analysis. To be sure, what is involved here in an incremental phenomenon for which no real line can be drawn between present and future worries. In order to bridge this gap, examples from science fiction literature will be presented to illustrate the points at hand.

AI power and potential

By any definition the present powers of AI machines are both impressive and widespread. Cyberneticians have already created or proven that AI constructs can do the following:

(2) “Imitate the behavior of any other machine.”

(2) Exhibit curiosity (ie are always moving to investigate their environment; display self-recognition (ie in order to the sight of themselves); and manifest mutual recognition of members of their own machine species.

(3) Learn from their own mistakes.

(4) Be as ‘creative’ and ‘purpose’ as are humans, even to the extent of “looking for purposes which they can fulfill.”

(5) Reproduce themselves, in five fundamentally different modes, of which the fifth—the “probabilistic mode of self-reproduction”—closely parallels biological evolution through mutations (which in the case of M. Sapiens means random changes of elements), so that “highly efficient, complex, powerful automata can evolve from inefficient, simple, weak automata.”

(6) “Can have an unbounded life span” through self-repairing mechanisms.

In short, “a generation of robots is rapidly evolving, a breed that can see, read, talk, learn, and even feel [emotional].” But the essential question remains—can these machines be considered to be “alive?” Kemeny presents six criteria which distinguish living from inanimate matter: metabolism, locomotion, reproducibility, individuality, intelligence, and a “natural” (non-artificial) composition. In all six he concludes, AI servo-mechanisms clearly pass the test. Even a critic of AI such as Weiner admits that computers are sufficiently “complex and autonomous” to be called an ‘organism’ with ‘self-consciousness’ and an ability to be ‘socialized’. He sees “no way to put a bound on the degree of intelligence such an organism could, at least in principle, attain”, although from his critical vantage point not in the “visible future.”

From the opposite perspective, cyberneticians have come to the realization that some sort of automatons/human equivalence is fast becoming reality, due to
the structural limitations of the human brain as compared to machine potential. Clarke notes, "the cells composing our brains are slow-acting, bulky and wasteful of energy—compared with the scarcely more than atom-sized computer elements that are theoretically possible"; an order of efficiency 10 billion times greater for electronic as opposed to protoplasmic cells. Second, the brain suffers from being "an organ that has been developed in evolution as a specialized means to survive," largely dedicated to ensuring body homeostasis (equilibrium) of which abstract thought plays a small role. Homo Speculatrix is really at most Homo Semi-spectulatrix, and so "is not even a priori a good thinking instrument.""

In addition, there may be no such thing as 'true creativity' since "neither man nor machine are able to create information." Given that all 'creative' thought is merely a matter of juxtaposing or combining previously existing information into different configurations (ie recycling 'matter' into different forms of energy), there is consequently no bar in principle to the development of artificial intelligence. In reality, "computers do only what you program them to do in exactly the same sense that humans do only what their genes and their cumulative experiences program them to do." Thus, groundwork has already been laid for the arrival in the not too distant future of artificial intelligent machines—"humanoids"—which will exhibit all the important qualities and traits characteristic of Man. It will be ready to 'serve' us—whether we shall be ready for it is quite another matter.

Definition and categorization of robot 'criminality'

Much of what is written in science fiction seems impossible, and some of it may indeed be just that. Yet as Clarke argues: "the one fact about the Future of which we can be certain is that it will be utterly fantastic." Since it is beyond human capability to distinguish a priori the truly impossible from the merely fantastic, all possibilities must be taken into account. Thus science fiction's utility in outlining the problem.

Essentially, humanoid anti-human activity can be separated into four categories: misplaced benevolence, well-intentioned behaviour causing harm, unintentional injury, and intentional criminality. The first is not strictly criminal but does involve moral and philosophical issues of fundamental import, especially that of human freedom. In political terms, it is akin to the fascist syndrome of the "authoritarian personality," or as Erich Fromm calls it: our "escape from freedom." The theme was already implicit in the Golem who, or which? was used as a super-human Guardian of the Jewish community. It continues in such novels as The Mad Metropolis and Fowl's Hammer wherein, faced with mutual annihilation, Maxis signs a new Hobbesian Social Contract transferring "the burden to a power more capable of reaching the solution than ourselves"; "a common supra-national authority"—a giant computer. The first problem with this, however, as Asimov suggests, is the difficulty of enshrining such a ruler with the human attributes of compassion and mercy. In addition, this artificial 'philosopher-king' would be more king than philosopher, given the logical contradiction (to its mind) in our asserting a difference between natural and positive law.

TUTURES December 1961
However, the greatest problem here is that inevitably such a seemingly omniscient dictator will “err on the side of benevolence” by not allowing humanity to do anything which can logically be construed as harmful. It could censor violence and “non-functional sex” entertainment as being “mentally disturbing”; it could be equipped with an emotional polygraph leading it to ignore orders presented by someone “emotionally overwrought”; it could even act as an exaggerated Jewish mother, soothingly ordering (and forcing) its charge not to smoke, to eat only nutritious food, etc., as hilariously portrayed by Sheckley. But it is really no laughing matter, especially when we turn to Pohl’s vision of a world controlled by computers—without Man being aware of it—through “the systematic biasing of data”. In this, as in the previous cases, the computer’s purpose is to ‘benefit’ man, thus heightening its invidiousness.

When we turn to more obvious forms of ‘criminal’ behaviour we find that Asimov, preeminently, has given serious thought to the problem, attempting to preclude the possibility by positing his “Three Laws of Robotics” which theoretically prohibit a robot from injuring a human being. But this raises more questions than it answers. What could stop humans from creating robots without the “Three Laws” programmed in? In addition, if computers indeed have self-educating capacities, what is to stop their eventual by-passing of such safety circuits “just as man gets by with a strict upbringing”? And insofar as the “Three Laws” are concerned, Asimov himself admits that to accomplish other goals (eg child-rearing) “a certain weakening of the First Law” is necessary—for example, spanning (‘harming’ a child in robot terms) for a greater future human good. But where does such a dilution end? And finally, even granted the immutability of the “Three Laws” (or perhaps because of it) Asimov’s robot stories continually show how so many things can still go wrong. In sum, one cannot give robots the Prometheus fire-gift of intelligence and still hope to keep them shackled.

One way or another, then, robot freedom must lead to some harmful behaviour, even if well intentioned. This is due in part to the literal-mindedness of a computer which “is logical but not reasonable”, and thus may carry out orders ad absurdum. As Kemeny notes: “The trouble with modern computers is the fact that they do precisely what you told them to do, and not what you meant to tell them to do”, Asimov describes how the First Law forces a robot proofreader to distort scholarly criticism because of its ‘harmful’ effect on the recipient, and how it would even commit adultery to raise its master’s self-confidence.

The dangers increase when we move to the third category—unintentional harm. This could occur when a robot performs a programmed task but not in the proper place or time; when a self-teaching robot has not yet advanced past its child-stage and is not cognizant of the consequences of its actions (imagine a superhuman hyperactive brat); or when under First Law programming it becomes disoriented at the sight of a human in the process of being harmed and compounds the trouble.

As we turn to the fourth category—intentional harm—we move from danger to nightmare. The first problem is one of control. As Wiener points out, “the ideal computing machine must . . . be as free as possible from human interference to the very end” for maximum speed and efficiency. This leads to the situation

FUTURES December 1961
wherein the learning machine’s “teacher will often be very largely ignorant of quite what is going on inside”, and thus will not know if and when the computer has learned too much, i.e. that the danger point has been passed. In fact, as Clarke notes, “even machines less intelligent than man might escape from our control by sheer speed of operation.” And if they become more “intelligent” another problem arises, for “the machine with a higher-order programming will do things that we may not have foreseen in detail... and so we are bound to find that the purpose of the machine does not conform to the detailed mode of action which we have chosen for it.”

This latter point is of such importance that it bears some elaboration. The crux of the problem, Weizenbaum argues, is that modern large-scale systems are created by various individuals who have different functions and pursue varying goals. What is the result?

It is simply a matter of fact that almost all very large computer systems in use today for ‘many layers of poorly understood control structure and undocumented encoded knowledge.’ It is simply no longer possible for these systems’ designers—or for anyone else—to understand what these systems have ‘evolved into,’ let alone anticipate into what they will evolve. The ramifications will be felt not only in science but in law as well. Minsky, for one, argued that “it does not follow that the programmer... has full knowledge (and therefore full responsibility and credit) for what will ensue. For certainly the programmer may set up an evolutionary system whose limitations are to him unclear and possibly inescapable.”

It is little wonder, then, that cyberneticists have begun discussing the possibility and attendant problems of AI psychopathology. Minsky goes so far as to “expect the first self-improving AI machines to become ‘psychotic’ in many ways, and it may take generations of theories and experiments to ‘stabilize’ them.” Others have already developed programs which are able to artificially synthesize paranoia in AI computers. Asimov, again one step ahead of the game, includes a robot psychologist in most of his robot stories.

Under such conditions one cannot assume that AI machines will always labor for human benefit. Especially if Neumann’s premise of machine self-reproduction and evolution is correct, the ultimate horror of ‘speciesmen’ may enter the picture. If our own usual criterion for ranking the hierarchy of life is intelligence, M. Sapiens could take us at our word and relate to us as we today relate to ants.

This latter premise is already found in Capek’s classic R.U.R., which describes Man’s annihilation as a result of such a development. Of course, this is rather extreme and melodramatic, but why couldn’t such ‘live’ creatures at least destroy humans who seek to ‘unplug’ them, as does Vulcan 3 or HAL, in Clarke’s 2001? This is far less ‘fantastic’ than similar instincts found in far less intelligent beings which exist today.

On the other hand, one need not anthropomorphize these creatures to accept that a problem exists. As Hoffstadter speculates:

Probably the differences between AI programs and people will be larger than the differences between most people. It is almost impossible to imagine that the “body” in which an AI program is housed would not affect it deeply. So unless it had an

FUTURIST December 1981
amazingly faithful replica of a human body—and why should it?—it would probably have enormously different perspectives on what is important, what is interesting, etc. 

In short, whether humanlike or not, such creatures will probably have purposes and goals which do not jibe with those of their human creators. As the father of cybernetics, Wiener, acknowledged over two decades ago:

It has been supposed . . . that the dangers mentioned by Samuel Butler that the machines may come to assume control humanity are absurd and empty. But now that the machines are stepping up one or more stages in their functions, it is their ability to program a program, this idea of the machine is already insufficient and the difficulties and dangers conceived by Samuel Butler assume a new actuality. 

In the present state of cybernetic uncertainty we would do well, then, to heed Hamlet’s warning: ‘For’tis the sport to have the engineer/ Hoist with his own petar . . .’

Cybernetic law in the future

From a legal perspective it may seem nonsensical to even begin considering computers, robots, or even more advanced humanoids, in any terms but that of inanimate objects, subject to current laws. However, it would have been equally ‘nonsensical’ for an individual living in many ancient civilizations a few millennia ago to think in legal terms of slaves as other than chattel.

Notwithstanding certain obvious biological differences between these two cases,10 for purposes of law those civilizations could hardly have cared less that a slave bred the same as his masters, or their legal definition of ‘humanness’ was based essentially on their conceptions of mind, intelligence and moral understanding—characteristics which the slave supposedly lacked. Similarly, by our present legal definitions robots too must lack such traits, but this may be more a matter of antiquated semantics than (potential) physical reality. Just as the slave gradually assumed a more ‘human’ legal character with rights and duties relative to freemen, so too the AI humanoid may gradually come to be booked on in quasi-human terms as his intellectual powers approach those of human beings in all their variegated forms—moral, aesthetic, creative, and logical.

Thus, in the highly schematic analysis which follows, the legal categories will be presented on a graduated scale, as we move from the AI robot as a piece of property to a fully legally responsible entity in his own right. This preliminary inquiry will attempt only to extract those legal principles which may be relevant for the phenomenon at hand, and is not an attempt to review the large body of cases and precedents within each category.

1. Product liability. As long as robots continue to be merely sophisticated automata, many injuries stemming from their actions would fall into the broad category of product liability. However, not only the manufacturer may be liable for damages. Limited liability may also be ascribed to other sources such as importers, wholesalers, and retailers (and their individual employees if personally negligent);11 repairers, installers, inspectors, and certifiers,12 and even the end user himself.

FUTURES December 1981
While the laws of product liability are fairly clear-cut for traditional property and merchandise, in the case of computers and robots the issue is more complex. First, there are usually at least two distinct manufacturers involved here—one for ‘hardware’ (the physical structure of the machine), the other for ‘software’ (its instructional program). As noted earlier, when things go wrong it is becoming more and more difficult to trace the defect or fault back to any single manufacturing or programming source, and at times there is no one at fault.15

A second difficulty arises with the principle of ‘inherent risk’:16 If there is a risk inherent in the very nature of the product, then liability is assigned only if the manufacturer or dealer do not attach a warning to the product, or if the product has a defect above and beyond the normal inherent risk of the product. While the inherent risk of a lawn mower is clear, not so that of a computer which is capable of a huge number of diverse functions, and the problem will become even more complex once fourth generation computers are given the power of self-programmability. As Waddams concludes: ‘Not every product that causes damage is thus free from liability’.17

(2) Dangerous animals. Once such scientistrachiasms have the ability to independently choose their own behavior patterns and become auto-locomotive, the possible damage that they may inflict increases greatly. Because of this greater inherent risk to society, the issue of responsibility may be transferred from the manufacturers/distributors to the end users/owners and the principles relating to ‘dangerous animals’ become germane. Such a concept, of course, is hardly new. Jewish Talmudic law devotes a substantial amount of attention to the ‘going out’, US law also addresses the issue, although not in a completely consistent fashion: ‘While two or three jurisdictions insist that there is no liability without some negligence in keeping the animal, by far the greater number impose strict liability’.18

However, strict liability is applicable only to geographically dangerous species of animals: eg wolves, monkeys, etc. In the case of ‘usually harmless’ species, ‘it must be shown that the defendants knew, or had reason to know, of a dangerous propensity in the one animal in question’.19 Given the ‘Frankenstein complex’ which large segments of society might develop in a highly roboticized society, it might be prudent to begin with the legal assumption of generic dangerousness, thereby forcing owners to more closely supervise their charges.

One point should be noted here. While the difference in tort responsibility between product liability and dangerous animals is relatively small, the transition does involve a quantum jump from a metaphysical standpoint. As long as AI robotics are considered to be mere machines, no concernative evaluative considerations are placed on their essence—they are inorganic matter pure and simple. However, applying the legal principle of dangerous animals (among others) opens a jurisprudential and definitional Pandora’s Box, for what is the ‘machine’ will have been transformed into a legal entity with properties of consciousness, if not some semblance of free will. Once begun, the legal development towards the ‘higher’ categories will be as inexorable as the physical expansion of robotic powers. In short, the move from the previous legal

FUTURES December 1981
category to the present one is the most critical step; afterwards, further juris-
prudential evolution becomes inevitable.

(3) Slavery. The term 'robot' stems from the Czech word robota, meaning
diligence, servitude, or forced labour. From the beginning, then, its purpose
was to function as humanity's modern slave. The ancient laws of servitude are
particularly relevant to our new slaves since by and large they were legally
perceived as mere chattels. Nevertheless, differences did exist. Jewish law essen-
tially held that, al and k'nd robah—the hand of the slave is like the hand of its
master—but only for purposes of agency. As Cohen notes, "with regard to the
personal liability of a slave there is an old controversy between the Sadducees
and Pharisees"; with the former contending that the master should be answerable
for his slave's injurious actions while the latter (whose position proved decisive)
argued no liability for the owner since slaves have the ability to understand the
consequences of their behaviour. One should note that this particular dis-
agreement is quite relevant to our situation since robotic 'understanding' too is
highly problematic.

Roman law considered the slave in a different light: "a servile act lies
against the dominus, under which he must pay the damages ordinarily due for
such a wrong, or hand over the slave to the injured person." However, the
Roman "system of servile actions applied . . . to cases of civil injury, involving a
liability to money damages; it does not apply to . . . criminal proceedings of any
kind." And even in civil cases, the master was free from personal liability if
there existed a total absence of complicity.

American slave law followed the broad outlines of Roman law. As Cole
noted: "Criminal acts not done by his order, do not create a responsibility upon
the master". This, however, did not mean that the slave could be held
responsible for all harm caused by him, for he was justified in repelling force by
the use of his own force (inm 3i dedefere, annes feges omniique pars permitzit). In
the case of robots this would involve transgressing Asimov's Third Law.

The real difficulty in the slave-robot legal parallelism, however, lies not in
the liability of the owner but rather in the punishment to be meted out to the robot
in cases where no liability can be attached to his modern dominus. In all three
aforementioned legal traditions, it is the slave in certain circumstances who
must bear the brunt of the law's punishment. But how does one 'punish' a
robot?

On the surface the question seems absurd, for if a robot did consciously
commit harm one would immediately suggest 'pulling the plug.' But as was
pointed out above, conscious actions need not entail an intent to commit injury.
Yet even assuming a worst case scenario, the mere fact that at some stage robots
have slave law applied to them means in effect that certain gradations of
punishment will have to be applied as well. The law could hardly relate to
robots as slaves for the purpose of determining owner liability (and in certain
cases finding the owner has none), while at the same time relating to robots as
mere machinery when its own liability is under consideration. What then can be
done? Two broad approaches seem to be most feasible: rehabilitation and
restitution. The first would involve reprogramming the robot (far easier with
robots than with men)—and might even prove of eventual use to psychologists in

FUTURES December 1981
determining the why of criminal psychology and the ways of restoring the human criminal to social functionality. The second is an approach only recently being tried in criminal cases—depriving the criminal to compensate the victim for the harm caused (again, easier with robots than with human criminals).

(4) Diminished capacity. As a result of differences in human capabilities the law has developed several approaches which take into account those individuals who, while legally independent, have a diminished capacity for initiating actions or understanding the consequences of such actions at the time they are being committed. Here the law is concerned with mens rea ("guilty mind"), i.e. the conscious intent to commit a crime. In the case of AI humansoids, the question of intent may become significant in light of the aforementioned possible types of injurious and 'criminal' behaviour.

Within this broad category, two different types of deficient personalities are to be found. The Common Law distinguished between those who are mentally defective (permanently meretricious) and mentally disabled (temporarily disabled). Of the two, the problem of mental 'disorder' is more germane to humansoids, although not any less problematic than when applied to humans. For example, how would one apply 'temporary insanity' in a humansoid? One possibility—humansoids programmed with Asimov's Three Laws, could become temporarily disoriented while observing a human in the process of being harmed, with the possibility of such a creature compounded the injury or causing injury to others. This would fall under most traditional psychological theories of aberrant behaviour. A different, and more likely, possibility would be a temporary malfunctioning of the humansoid brain for physical reasons (short circuits, burnt out 'fuse' etc.). Here again the law only recently has begun to come to grips with the problem in humans—e.g. the XY chromosomal syndrome which may force the individual to become violent on occasion.

Indeed, the entire history controversy over Mentalism versus Behaviourism has received renewed impetus under the impact of recent work in behavioural psychology which seems to be on the ascendant. Its dominance would undermine the use of such terms as 'responsibility', 'intent', the whole mens rea principle. More important for our purposes here, however, is that such a behavioural tendency would greatly narrow the gap between human and humansoid psychology since ultimate both would be grounded on an epiphenomenal basis. One might even go so far as to suggest that here at least principles relating to Advanced humansoids may predote and pave the way for a re-evaluation of the law regarding human mental capacity.

(5) Children. The question of diminished capacity can be addressed from quite a different direction—the law of minors—applicable to humansoids because it deals with a legal entity of relatively high intelligence and low moral responsibility. In other words, the physical consequences of the specific action may be understood, but not the normative ramifications of such an outcome.

Cybernetics consider one of the most promising avenues of AI advancement to be self-education, i.e. learning from one's own mistakes, trial and error. Self-programming computers have already been created, and there is no intrinsic limit to the level of intellectual motivation future humansoids could
attain. Thus, it is quite possible to conceive of such creatures being purchased when still at a relatively primitive level of development (at a 'young age') when they can already perform rudimentary functions (and even as well in the performance of same). The social utility for such an early purchase would be the individualization of its eventual utilities based on the needs and wishes of its owners—no different than parents adopting a child at the earliest possible age so as to inculcate/imprint it with the values of the parents. However, as Prosser notes, with regard to the legal status of children, "the common law, unlike that of the civil law countries, never has made the parent vicariously liable as such for the conduct of the child". Yet even the common law recognizes parental liability under certain conditions. Such is the case if the child's tort is due to the parent's negligent control of his offspring with respect to the act that caused the injury. In addition, as Houston notes, "a child may be his father's servant, so as to bring the father within the rule as to employers' liability". Further elaboration on this point will be presented in the next category.

Underlying these divergent approaches is the philosophical question of how one determines liability. The Anglo-Saxon convention involves culpa liability—the idea that there can be "no liability without fault"; the Continental custom involves causation liability—"one whose interests are injured by the activities of others should be entitled to compensation without regard to the moral or social qualities of the act". Again, in our future case, society will have to strike a balance between a robot's 'parent' (respondeat superior) who may not in any way be guilty and between the need to protect the rights of the equally blameless victim.

(6) Agency. When all is said and done, in almost all circumstances the robot/humanoid acts as the service of some human principal. The law of agency, then, is the most comprehensive and germane with regard to both the essence and function of such a creature.

To begin with, the common law in some respects relates to the agent as a mere instrument. It is immaterial whether the agent himself has any legal capacity, for since he is a sort of tool for his principal he could be a slave, infant, or even insane. As Seavey notes, "it is possible for one not sui juris to exercise agency power". Indeed, the terms 'automaton' and 'human machine' have been used in rulings to describe the agent. Nor must there be any formal acceptance of responsibility on the part of the agent, Seavey argues: "The only element required for authority to do acts or conduct transactions... is the communication by one person to another that the other is to act on his account and subject to his orders. Acceptance by the other is unnecessary". Thus, as Mechem concludes: "Generally speaking, anyone can be an agent who is in fact capable of performing the functions involved". Here, then, is a legal category already tailor-made for such a historical novelty as the humanoid.

There are, however, two classes of people within this overall category: the 'agent', and the 'servant'. Who is a servant? If the situation is such that his employment is in fact closely controlled by his employer, he is to be subject to the control and directions of his employer in respect of the manner in which his work is to be done". However, "control and direction"
must be clarified. Rogers notes that these need not actually be present in any specific case; rather, the possibility of control and direction if the master so wishes, is the determining factor. But what of the aforementioned problem that modern computers, and certainly future robots, are not amenable to strict control or even open to detailed direction due to the incredible speed of intellectual operation with which they carry out functions as well as the programmers' inability after a while to even understand how it "thinks"? This lack is already accounted for in the law of agency, through a number of outstanding exceptions to the rule of "control and direction": such individuals as chefs, doctors, airline pilots, ship captains, etc., are allowed significant autonomy in the performance of their duties because of their expertise and skills which are not amenable to precise instruction on the part of the purchaser of their services.

Thus, in effect, the master is at the mercy of his own servant since the "master" is jointly and severally liable for any tort committed by his servant while acting in the course of his employment . . . based, not on the fiction that he had impliedly commanded his servant to do what he did, but on the safer and simpler ground that it was done in the scope or course of his employment or authority". Indeed, Prosser goes even farther in maintaining that "the master is held liable for any intentional tort committed by the servant where its purpose, however misguided, is wholly or in part to further the master's business." And Houston in the end comes close to applying "strict liability" to the master-servant relationship: "Every express prohibition of the wrongful act is no defense to the master at common law, if that act was merely a mode of doing what the servant was employed to do".

To future masters considering purchasing a humanoid servant one can only suggest—caecus emptor.

(7) Person. While this seventh and last category in practice involves merely an incremental upgrading of the humanoid's legal character, it does obviously mark a quantum emotional and philosophical leap from a human perspective. Even those future diehards who may balk at any suggestion that humanoids are in any way truly "alive", could accept the legal fiction of determining legal responsibility and liability in terms of categories (2)-(6) and principles which herefore have been applied only to sentient beings. But to consider such a creature autonomous and exclusively personally responsible for its (his) actions? Can there be such a thing as AI "free will"? As Hofstadter notes, the question itself makes you pause to think where your sense of having a will comes from. Unless you are a soulist, you'll probably say that it comes from your brain—a piece of hardware which you did not design or choose. And yet that doesn't diminish your sense that you want certain things, and no others. You aren't a "self-programmed object" (whatever that would be), but you still do have a sense of desires, and it springs from the physical substrate of your mentality. Likewise, machines may someday have wills despite the fact that no magic program spontaneously appears in memory from out of nowhere (a "self-programmed program"). They will have wills for much the same reason as you do—by reason of organization and structure on many levels of hardware and software.

Of course, this is hardly the last word on the matter (although it is the most
recent). No definitive answers are possible—yet. The future, though, may bypass the philosophers, theologians, biologists, psychologists, and the like, with a reality that will be difficult to explain away. As an early student of this problem put it:

What is it to be a person? It can hardly be argued that it is to be human... Could an artifact be a person? It seems to me the answer is now clear; and the first R. [Robot] George Washington to answer "Yes" will qualify. A robot might do many of the things we have discussed: moving and reproducing; predicting and choosing; learning; understanding and interpreting; analyzing (translating, abstracting, and indexing); deciding; perceiving; feeling—and not qualify. It could not do them all and be denied the accolade.

Thus, it would be best to leave to future generations the resolution of the ultimate legal challenge presented by the first R. George Washington to stand before the bar and proclaim: "Computo, ergo sum!" But society would do well to begin grappling with the lower-order legal questions inherent in the cybernetic revolution which has already arrived. It is hoped that this exploratory essay provides a first, albeit modest, step in that direction.

Notes and references

12. On the issue of "natural" material he notes that with the recent success of synthetic life in the laboratories, there is no theoretical bar to eventually evolving highly complex beings out of such matter. Ibid, page 13.

FUTURES December 1961
19. Sec. 1. W. Adorno et al., The Authoritarian Personality (New York, Harper and Row, 1950); and
22. Dr. N. S. Sutherland, Professor of Experimental Psychology at the University of Sussex (and a
computer expert) suggests that by the early 21st century human society will be grappling
with the problem of whether AI robots should be allowed to vote. From such enhancement
it is just a small step to AI leadership. See Roskii, op cit., reference 9, pages 47-48.
23. I. Asimov, The Caves of Steel (New York, Fawcett Books, 1953), page 140. See also U. Neisser,
"The imitation of man by machine", in Diebold, op cit., reference 1, page 449, for an
elaboration of this point.
26. Ibid., page 50.
27. Ibid., page 37.
28. R. Shlechter, "Street of dreams, feet of clay", in Jack Dann, ed., Wandering Star (New York,
Pocket Books, 1975). For variations on this theme see also D. F. Jones, Colony (New York,
30. (1) A robot may not injure a human being or through inaction, allow a human being to come
in harm.
(2) A robot must obey the orders given it by human beings except where such orders would conflict with the
First Law.
(3) A robot must protect its own existence as long as such protection does not conflict with
either the First or Second Law.
33. Ibid., page 61.
35. I. Asimov, "Galley slave", in Eight Sonies from The Best of the Robot (New York, Pyramid
36. "Satisfaction guaranteed", ibid., page 84. See also "Galley slave", ibid., page 151 for yet
another example. R. A. Heinlein, The Moon is a Harsh Mistress (London, New English Library,
1977), provides a different illustration of such well intended computer activity — to the
detriment of Earth's inhabitants.
38. "Lenny", ibid., page 123. See also N. Wiener, Cybernetics (Cambridge, MA, MIT Press, 1961),
page 7.
41. A. Turing, "Computing machinery and intelligence", in A. R. Anderson, ed., Minds and
42. Wiener, op cit., reference 30, page 175.
45. J. Weizenbaum, "Once more: the computer revolution", in M. L. Dertouzos and J. Moses,
46. Emphasis mine; M. Minsky, "Steps toward Artificial Intelligence", in Feigenbaum and
Hamm, op cit., reference 2, page 447.
47. M. Minsky, "Computer science and the representation of knowledge", in Dertouzos and
Hoffstadter argues that AI 'emotions' will evolve in a more naturalistic fashion: "Any direct
simulation of emotions ... cannot approach the complexity of human emotions, which arise
indirectly from the organization of our minds. Programs or machines will acquire emotions in

FUTURES December 1981
the same way: as by-products of their structure, of the way in which they are organized." D. R. Horstman, Gold, Escher: Bach. An Eternal Golden Braid (New York, Vintage Books, 1980), page 67. For a brilliant and profound discussion of AI, see especially chapters XVII-XX.


50. Kemeny himself uses the term 'species' to describe this new 'race.' Kemeny, op. cit., reference 20, page 71, in chapter one, entitled "A New Species is Born."


55. In principle, however, this is becoming a distinction without a difference as biochemistry has come to understand the inorganic origins of life. See, for example, R. E. Dickerson, "Chemical evolution and the origin of life." Scientific American, September 1978, 239 (7), pages 62-70.

56. Wiener concur: "The more intelligent the slave the more he will insist on his own way of doing things... To this extent he will cease to be a slave. Similarly, the machine with a higher-order programming..." Wiener, "The robot and the machine," in Hook, op. cit., reference 5, pages 119-120. For a lengthy review of the historical development of slave legal theory, see T. R. C. Cobb, Law of Negro Slavery in the United States of America (New York, Negro Universities Press, 1968), pages 279-293.

57. While the question of creativity is perhaps the most baffling of all, some recent evidence seems to dismiss the very notion. Mozart, for example, published a pamphlet explaining how to compose "as many German waltzes as one pleases" merely by throwing dice! See W. B. E. Brinton, Jr., "How scientific is intelligence?" American Sociologist, November-December, 1977, 65 (6), pages 694-702.

58. S. M. Waddams, Product Liability (Toronto, Canneull, 1974), pages 15-17. For a brief but comprehensive sketch of various aspects of this latter, see R. N. Freed, "A lawyer's guide through the computer maze." The Practical Lawyer, November 1969, 6 (7), pages 34-38. Freed is one of the pre-eminent experts in the field of computer law. For a wide-ranging survey of his work see his Computer and Law: A Reference Work, 4 ed (Boston, Roy N. Freed, 1973). Especially relevant to the above point is his detailed section: "Contracting for the use of computers and related services", pages 170-173.


60. See J. G. Fleming, The Law of Torts, 4 ed (1971), page 444, for a list of such products and relevant case citations; also J. W. Salmond, The Law of Torts, 2nd ed (1973), page 511. The entire legal analysis as presented in this section— with regard to categories (1)-(6)—need not imply only private ownership. One can easily picture a society with a state-owned army of garbage cleaners, soldier policemen, teachers, etc. Indeed, as previously suggested, there is no bar to AI robots comprising part of the government itself. For legal principles which could be atrodicted in the former eventuality see R. Street, Government Liability: A Comparative Study (USA, Federation Books, 1973).

61. supra, page 11.

62. Waddams, op. cit., reference 58, pages 39-42. Other criteria are tests for adequacy and the duty to recall goods when defects are found. But this fails under the broad category of 'due care.' In addition, as Freed notes in op. cit., reference 58, page 38, one cannot avoid legal problems by not using computers. "Once it can be shown that a computer system overcomes the deficiencies of people in detecting hazards and avoiding harm then it is economical and big business proves to work, failure to use such a system would constitute negligence." On this, see also E. B. Levin, "Torts," in R. F. Bigelow, ed., Computer and the Law, 2 ed (Chicago, Commerce Clearing House, Inc., 1969), page 155.

63. supra, page 11.


67. Ibid.

FUTURES December 1981
although even here there are some notable exceptions, *Ibid*, pages 614-621.


91. Rogers, *op cit*, reference 81, page 518. Possible here adds: "or is subject to a right of control, by the other". Pozner, *op cit*, reference 65, page 660.


94. Pozner, *op cit*, reference 63, page 464. The only time the master is not liable is when the act of his servant is committed during "foolish and dastard" (p 461) which hardly applies to the situation at hand.


97. For an excellent dialectical interplay of ideas from the various disciplines on this issue see the selections in *Hook, op cit*, reference 48.


99. The problem here is not merely how does our relate to the humanoid if it transgresses the law; even more delicate is the question of how the law will deal with those humans who injure a humanoid. Is shooting one to be considered murder?

100. Agashe, Alaino's "Biennennial man" op cit., pages 179-180, provides a glimpse of what that historic scene may look like.