THE COMMERCIAL DIFFUSION OF FRIENDLY ROBOTS INTO SOCIETY: OUT OF THE UNCANNY VALLEY

Kenneth Anderson
Washington College of Law, American University
Washington DC

CONTENTS

I. INTRODUCTION: NORMATIVE ENGINEERING, ROBOT REGULATION, AND THE PRESSURES OF RETURN ON INVESTMENT IN ROBOTICS

The WeRobot2014 conference is addressed to the many issues raised by the diffusion of robots of all kinds into broader society. This means the spread of robotic technologies from the spaces in which they have been most present – the assembly or factory floor, for example – and into spaces in which they are not found or expected now, places where ordinary humans do their business, on the street and highways and airspace, in shops and stores, hospitals and nursing care facilities, schools, and homes. This is partly a matter of technological development - the capability, for example, of a complex machine to operate in spaces where ordinary people, un-expert in understanding machines and technologies, are present and going about their own business – to be safely present, on the one hand, but also able to perform some useful function reasonably effectively, on the other. It is also a matter of policy and regulation of design and usage, to have technologies that can perform safely and effectively.

The necessary policy and legal regulation will be able best to perform this role, however, if it is a part of the planning and design processes for developing technologies in the first place, rather than a potentially unpleasant and unwelcome add-on at the end. Military robotics, such as drone and automated weapon systems, offer a good example of the gradual integration of what might be called "normative engineering" – law, regulation, and ethics – into the design requirements of the system from the very beginning. This integration is made easier, in the case of military weapon systems, because the requirement of legal and ethical review of weapon systems is already a strong normative feature of the law of armed conflict, so the idea that an expensive weapon system needs to have legally required normative input from the outset (and all the way through the process to the fielding and deployment stages) has not been a difficult idea for the military to embrace. But the same idea will broadly hold for other areas of robotics in which the aim is to bring them into social life and social spaces – among the civilians, so to speak. The mainstreaming of robots into ordinary social life benefits from normative engineering from the outset and throughout the design, development, and deployment process, and beyond.

Technological development and law, regulation, and policy in order to ensure that the machines are safe for diffusion into wider society are essential to this process, of course, but law, regulation, and policy are not limited simply to these roles of ensuring safety and perhaps utility of robotic systems. The process of development and diffusion of these technologies will be through commercialization, marketing and sales of these machines and machine systems, by private enterprises. They will have business plans, ways in which to secure capital and financing for their activities, and estimations of who will buy how many of their systems. The business models will raise a host of conceptual issues; in the self-driving car field, for example, Bryant Walker Smith (among others) has asked whether a self-driving car or a car linked by sophisticated communications to other cars should really be thought of as buying a car or instead leasing a hardware platform for software than needs constant updating – and should the business model and the legal model attached to it, sale or lease, reflect that? This particular question could be asked about a host of other robotic technologies aimed at wide commercialization, of course, not just cars.

Business models for commercializing these new technologies depend upon many assumptions that entail many risks. More risks, arguably, than the development of personal computers, operating systems and software, and the Internet and web. On the technological development supply side, for example, investments into the telecommunications industry that have enabled the global Internet have been expensive, and one does not want to underestimate the costs of development of hardware and software alike in the spread of computers. But on the demand side, the user side - while the risks and vulnerabilities created by a world in which so much physical infrastructure now depends on the Internet are well known – overall, in a world in which people used software programs over decades on personal computers, the problems of physical machines "harming" people have not been a central issue. Particularly once a clear rule was established that suppliers would not be liable for economic losses (including for loss or destruction of data) legal risk has been strikingly low, considering how many activities these technologies touch. It's not entirely clear, frankly, that the culture of the tech industry, trained by experience in these protected spaces, is quite prepared for the world of accident and injury lawsuits that might well await. But it would be astonishing if future legal protections and liability safe-harbors, if any, for new technologies such as self-driving cars were anything like those afforded software and its license agreement regime.

Ryan Calo's path-breaking law review article from several years ago, "Open Robotics," posed a basic question that remains central to the diffusion of robots into social spaces. It is central to the technological path that both makes diffusion possible and carries it forward, potentially to reshape whole activities and spaces in society, possibly no less than the Internet has done. It is a question familiar to everyone at this conference, I'm sure: Is technological development in robotics, and its diffusion and commercialization (and then the *next* round of development, diffusion, and commercialization of technology, and so on) on a path toward general-purpose robots, on the one hand, or more sophisticated, but essentially dedicated-purpose appliances, on the other? "Social" robots or "super-toasters"? "Open Robotics" pointed out that although there would of course be some of both, technological development in one direction or the other would be strongly path-dependent, and the main technological path, in turn, would be strongly dependent on law, regulation, and policy. Regimes of product liability, for example, would have a strong influence on what kinds of machines could gain early recognition of their safety. Early endorsement of safety might exert a powerful push toward one path or another, and product liability law in the United States would, broadly speaking, push toward special-purpose appliances.

Although we are right to focus on products liability and other regimes of accident and design liability through litigation, the path forward as between the general-purpose or super-toaster robot is also influenced by other areas of law and regulation. As the next section of this article notes, in some of the most important areas of robotics in the sense of diffusion and commercialization into broader society – assistive care robots for care of the elderly, disabled, and infirm, for example – it is quite possible that the most important regulatory and legal pressures will come from regulatory agencies. Insofar as this means that "normative engineering" ought to enter the design process early on (rather than being a potentially expensive after-the-fact blow-up in tort and product liability litigation), overall this will be a good thing. But regulatory pressures at the front end will also likely have their own influences upon the path of technological development, and arguably these, too, will be a factor driving toward the super-toaster model.

Not all the pressures for the super-toaster path necessarily arise from regulation or law. Important ones arise for investment reasons in the commercialization of robotic technologies. Not only from the standpoint of safety and product liability law, but also from the standpoint of being able to perform primary design functions and to be developed at a cost consistent with a company's business model, general-purpose robots are likely to be more difficult to develop than special-purpose robotic appliances. The risks of lost investment for a technology that never quite succeeds, over safer investment in a machine that is less flexible but which is easier for that same reason to design (alongside the a more predictable and more certain outcome with respect to design flaws and product liability litigation) perhaps provide a powerful investment incentive not to go for general-purpose designs. Moreover (as the next section also notes), a major source of demand for "assistive care" machines of all types is likely to be government in one way or another – on behalf of elderly Baby Boomers - pressure to keep costs down, by taking the least risky design paradigm in the short to medium term, will likely push for appliance toaster paradigms over general purpose robots.

The broader point is simply that the diffusion of these machines into across areas of activity and places society will tend to follow paths laid down by policy, law, and regulation. Which is neither surprising nor bad: social robots are for social purposes, and these include the costs, benefits, risks, and opportunity costs of one technological path over another. Including, however (taking into account the recent spate of books and articles about impacts of robotic technologies on employment and labor), the possibility of not automating or roboticizing some functions at all.

Considered as the business of robotics and the commercialization of technologies, the safer, less risky, more certain strategy of social investment into robots probably lies in the super-toaster paradigm. Whereas the riskier, less certain investment into general-purpose robots – including what this paper will loosely call "social" robots and "friendly" robots – would likely, in my view, produce a greater social return, but only over a much longer run. It's not an either/or proposition for social investment, of course; but it will tend to be weighted one direction or another, at least for the medium term and perhaps much longer. In the long run, of course, we are all dead – at least if you are a Baby Boomer, and so the question of how long to see a return is not irrelevant. It is possible, of course, that the "shoot for the moon" approach taken by Google and other deep-pocketed tech corporations offers a source of investment capital that makes this question moot, from an investor point of view; Google will fund research and development of these "friendly" robots without the pressures of more near-term return on investment. But I think it would be imprudent to assume this. Normative engineering will be both influenced by and have an influence on investment into the commercialization and diffusion of robots

*　　*　　*

This paper offers a short discussion of an aspect of this commercialization and diffusion that is not precisely about law, regulation, or policy, though it is about "normative" impacts of path dependency and the "normative engineering" of robots. Nor is it about product liability law and doctrine, or whether assistive care robots should be treated as medical devices under FDA regulation. Nothing so narrowly legal as that.

This paper asks, instead, an underlying conceptual question. What are the implications for the technological paths for developing "social" and "friendly" robots arising from how humans affectively perceive them – feelings, emotions, mood, and so on? – what we will call "human affective response" and how the social, humanoid, friendly robots we might ideally like to diffuse into many aspects of social life are perceived by ordinary people? What implications might "human affective response," whatever exactly that might be, have with respect to the reception of "social" and "friendly" robots in society, and also to their design? We'll start with a slightly different one, however: what are the implications for robot design when they are not perceived as friendly and social? The Uncanny Valley and beyond, in so many words.

II. The Diffusion of Robots into Society, Commercialization, and Business Models

A. Some Terms

First, however, a brief discussion of a few terms used in this paper and how they are used, followed by a discussion of the conditions for the commercialization of robots generally in social life. The paper sketches out a few of the fundamental questions that a market study in support of a business plan for assistive care robots for the elderly, disabled, or infirm would have to answer. These fundamental elements spell out certain considerations that become relevant to the question of the paths that the business of robots aimed at ordinary situations where people live and work would have to answer. The point of discussing these business considerations is that the many people in the robotics field – at least those who think about how they can be more widely utilized in society – have long since come to the conclusion that diffusion and commercialization of robots means robots that are more humanoid, social, interactive, smarter, not scary or threatening, and safe and effective in their range of tasks when operating in proximity to humans. These fundamental business considerations suggest two conclusions, one already noted in the Introduction: first, the economic facts of the commercialization process likely favor super-toasters, rather than general-purpose robots, even though more flexible and general robots would be more useful over time. Second, robots that are designed to be something more than appliances will do better in general social usage if they are "friendly."

But we pause briefly to mention what this paper means by its various robot terms. Terminology in this discussion is deliberately loose; many terms covering roughly similar concepts are used in the literature today, and too great a focus on definitions risks not engaging the substance, irrespective of what exact terms are used. In general, this paper follows the conceptualization of robots that Ryan Calo offers in his paper for this conference, "Robotics and the New Cyberlaw." His account of what makes robots different from cyber and software in terms of normative engineering seems to me correct and the best discussion of it available. This paper takes essentially by assumption Calo's framing of robots as specially marked out by the features of "embodiment" in the physical world; "emergence" in the sense of increasing analytical and decision-making capabilities rooted in self-learning and related advances in computing; and "social meaning." The terms used in this paper are looser, deliberately, in order to be able to engage with slightly different terms and concepts used in discussions across the field, but the concepts roughly correspond.

So, *robot* for this paper's purposes refers to machines with some form of the capabilities Calo specifies; in this discussion, the focus is particularly on "social meaning," which is amplified by capabilities in the other two. A robot is, for some purposes, a subset of *automation*. Not all automation counts as "robotic" for our purposes, however. A high frequency trading system on a stock exchange, for example, is automated but does not have, in the sense of this paper, "extension" into the physical world. Robots for this paper are machines that have sensors to take in information about the physical world; processing and computing functions to allow it to analyze and make decisions based on that information and its programming; and both mobility (being able to move about and change its location) and motion or movement of its own parts, such as arms, to be able to physically alter the world around it.

This is intended to be a loose definition and a flexible one. What Calo calls "emergent" capabilities, we'll refer to in talking about a *complex* or *probabilistic* robot - one that has advanced (relative to some point in time) processing and analytic capabilities, including capacities for self-learning, probabilistic decision-making, and the possibility of decisions and actions that cannot be fully predicted. We'll use complex rather than the more obvious *AI* in order not to commit to a view on AI itself. A *social* robot is a robot, for our purposes, that is intended for use in close proximity to human beings, ordinary people in the street, workplace, or home; it requires programming to ensure that it has capacities for movement and mobility without harm to humans, communications skills to ensure that it can receive instructions and further additional capacities include the ability to communicate with people or things outside of itself, and advanced probabilistic programming as well.

A *humanoid* robot is one that has been designed to make it look, sound, or appear more, or especially, human. Finally, a *friendly* robot is one that, for purposes of this paper, combines the features of social and complex with a friendly, non-threatening appearance; it might or might not have humanoid characteristics, though most of those that are relevant to our discussion would be likely to have that as well. But these terms are used flexibly and loosely in the discussion.

B. A BUSINESS MODEL FOR COMMERCIALIZATION OF ASSISTIVE CARE ROBOTS? (NOT QUITE)

The discussion in the Introduction suggested that the path of robotics might easily lead toward specialized appliances, rather than the social, complex, friendly, more general-purpose robots that, arguably, are likely to be the most socially useful form of robots over the long run. If we assume, however, that we are able to press down the path of "friendly" robots, as broadly described above, what are the elements of commercialization as a necessary step in diffusion of these systems? We have the experience of a handful of robotics entrepreneurs – people like Rodney Brooks and others, from whom I have learned much about business models in this business. Some of these robotics businesses involve more "friendly" robots than others, and we also have the recent business events of acquisitions of a number of robot companies by large tech firms. What are essential conditions of commercialization and, for the specific purposes of this paper, what might they imply about the way in which people see these robots?

I want to sketch out in brief terms some basic considerations that a marketing study for a robotics business plan would have to take into account, in a specific market for robots – assistive, personal care robots for use by the elderly, infirm, or disabled. The reason for picking this particular area is simply that it is an area in which there is enormous demand (and which I have already been researching and conducting interviews). Indeed, potential demand is so large that if we stop to consider the fields of robotics in which growth is the most likely across society, one might well conclude that the whole field is essentially about old people.

Indeed, one might say this with some caricature about wide swathes of tech. It might look like something about cool, hip young people, but when it comes down to it, what? Amazon is about the delivery of goods, entertainment, everything, directly to people in their homes, so they never have to leave. Apple supplies your computer needs in your house with fairly idiot-proof plug and play. Google makes the web painless – but more importantly, the self-driving car will enable people who don't or can't or shouldn't drive to get around. It is hard not to think that aging Baby Boomers are the drivers of the tech industries consumer business model. Assistive care robots for the elderly is integral to that, along with the infirm and disabled of any age.

A second reason for looking at this particular area of robotics business is machines in this social area are, almost by definition, up close and personal with ordinary people. Assistive care machines in this context is deliberately a loose term; it might refer to friendly robots, but it also in practice means many mechanical and motorized devices that can assist the disabled. Machines that can help with dressing, bathing, use of toilets, etc., need not be smart machines, and clever mechanical design – but it is dealing directly with people who require assistance or with people who are providing that assistance through or with the machine. For this particular discussion, we will exclude machines used specifically in hospitals by nurses or others, in order to focus on machines intended for use either in a person's own home (either by the person or by some home health care aide, perhaps on a visiting basis), or in a residential care facility for the elderly, infirm, or disabled, with some nursing and attendant staff available, but not a fully staffed medical facility.

Robots used in these settings will have to be "social" in some sense and, by preference, "friendly" robots. In the short term, that means being able to carry out intended tasks without knocking people over by accident. Over the longer term, that means being able to communicate reasonably well and understand essentials of what a person tells it, make some range of "emergent" decisions, and be able to perform well mechanical tasks that might involve touching the person – such as helping to dress or to draw blood.

"Interactions with people" does not simply mean interactions with the "clients" – the elderly, infirm, or disabled person, however. Consistent with the experience gradually being found in military robotics, much of the model of friendly robots revolves around human-robot dyads, "tag-teaming" robot with human in order to achieve the best and most efficient division of labor. Tyler Cowen describes this fundamental idea of human-robot interaction in his book, *Average Is Over*, using the example, not of robots, but of computer-human chess teams, but as he points out, the basic dynamic is the same in many functions in which some part of the task is most efficiently done by a robot in tandem with a human providing supervision, override, monitoring, but also aspects of judgment and direction.

In the assistive care context, one can already see emerging parts of the robotics diffusion consisting of assistance machines for elderly persons who are still in their own homes – an aid to independence. But much of the model is also likely to come from assistive care machines used by residential care facility staff as they perform their duties – such as drawing blood, distributing medications, meals, and other things, aspects of cleaning, and many other things. Equally, periodic home visits by human aides might be accompanied by robots with one function or another.

The human-robot dyad, however, has very important implications for normative engineering, as the military robotics field has already found. A 2012 DOD Directive, for example, makes clear that a core issue of automation and autonomy of weapon systems is to ensure that humans are able to play the role envisioned for them. The issue is less what the robot can do, but rather what its human team-members can do, whether in-the-loop, on-the-loop, or in any other kind of capacity. The most important questions about robot performance might easily turn on whether human beings, when they are part of the robot-human team rather than simply the passive recipient of robotic services on command, can perform as anticipated in the system design or whether, for example, humans might be cognitively overwhelmed by speed or other factors that would create failure for the robotic system as a dyad, even if the robot performed according to its design. This will be important in assistive care and other robots for ordinary society; it already is.

C. FOUR (OR FIVE) FACTORS IN THE MARKET MODEL FOR ASSISTIVE CARE FRIENDLY ROBOTS

The amount of concrete, “numbers” information on the market for assistive care robots for the elderly and disabled in the United States is surprisingly thin. This is true on both the demand and the supply sides, and so it is premature to try and talk about the “market” for assistive care robots. However, we can at least talk about the core elements of what a market study would require and what, anecdotally, can be said about it. (I draw here, with thanks, on preliminary materials from research project currently in progress by Renee S. Anderson, a Rice University undergraduate and my daughter.) The most essential elements that a market study would need to establish are:

- Demand side for new assistive care robots on the part of elderly and disabled consumers or residential care facilities;
- Supply side by technology firms and what assistive care robots they can develop and make available;
- Demand side for paying for these robots by, presumably, mostly government provision for the elderly and disabled;
- Supply side for providing investment capital for research and development funds to bring devices to market;
- Demand side for law and regulation to ensure that the robots and safe and useful; and finally,
- Supply side of normative engineering to ensure that safety and utility are built into the systems from the outset.

This amounts to saying that the market for assistive care robots for this population has a demand side that consists of users of the technology (individuals and institutions) and a demand side consisting of the ultimate payer (presumably the government through programs for the elderly or disabled). There is a further demand side consideration – demand for safe and effective robots – that create demand side regulation.

Ultimately, the question of demand goes to consideration of the needs of this population and what technologies would be useful to them in daily life. Drawing on available literature and anecdotal interviews with people who work with these populations, in home aide care or institutional residential facilities, the most useful machines (in the real world and without sci-fi imagining) would be ones that can help either the elderly or disabled person to perform daily tasks in the residential setting, get out of bed, dress, bathe, use the toilet, cook, and so on.

Contacts in residential care facilities noted a need particularly for these kinds of machines in order that residents could be more independent of staff. But they also added the need for better, smarter, more functional devices to assist nurses and staff with such things as moving obese patients or turning them in bed, some important but routine nursing tasks including administering meds and injections, drawing blood, taking blood pressure, and such tasks. They also noted a need for more automated and smarter robots that could assist staff with cleaning, disinfecting, and related tasks. Only some of these things appear to be most usefully performed by friendly robots in the sense noted above; many of them are much more specialized electro-mechanical devices, perhaps with some version of programming to enable them to be used by staff in a quasi-unsupervised mode, even when in motion, such as cleaning devices.

It is striking that the available US literature is quite sparse on what either consumers or, perhaps more usefully, the management staff of residential care facilities believe they would find most useful in the way of machines and devices of all kinds to make their work more efficient, better, and easier. One thing that stands out, however, is that a core demand by the consumer or residential care staff is for devices of all kinds that can help the person remain as independent as possible for as long as possible. This means ways that mechanization of all kinds – not just robots, friendly or appliance-like, but also including much broader mechanical conceptions including “smart houses” for the disabled, designed for the needs of wheel chairs, etc., and incorporating many of the technological features into the living space itself – can help people remain in their homes as long as possible. This includes, as well, ways in which robots can help make cost effective home health care aides for these people, reducing the human staff time and visits required; in turn, this conception pushes toward machines in the home that can do a vastly better job of determining and reacting to a medical emergency (a fall, for example) than current technologies do.

The notion of preserving independence, and with it human dignity, seems (anecdotally) to be very strong – so much so that it forms a demand all its own. That is, although there are important values and considerations to ensuring that this population have meaningful human contact, many of these people, and the management of facilities that currently care for them, insist that independence and dignity matter most to them in the performance of intimate and personal functions, and for that reason want machines to be able to assist them. The machine becomes an instrument that extends the human self, while having to rely on a human being for those functions means dependence and loss of dignity through loss of privacy and intimacy. The important roles of human interaction are best served in areas that don’t violate a person’s sense of privacy and intimate function.

If we turn to the supply side, however, the question is whether the engineers and technologists, the firms that would be likely to supply new machines, are actually on a track to supply the kinds of machines of all kinds, from useful but not really robotic assistive mechanical devices, to genuinely friendly robots. To what extent does the supply side path of technology look like it is moving toward the demand sketched out above (and assuming that, in an area of remarkably little public data, the description of demand is approximately correct?

People present at this conference probably have more information on these questions than I have so far been able to ascertain, and I (and Renee Anderson) would be grateful to talk with technologists here and after the meetings in order to try and find better information. Initial discussions – again, all this is purely anecdotal – suggests that the technologists are somewhat guessing as to what kinds of actual robots are desired. They have clear ideas in areas such as driverless cars or civil aviation drones, by contrast, but when it comes to homes with elderly people or residential care facilities, the designers seem to be far more limited in their knowledge or vision. There are some important institutional exceptions, particularly in parts of the academic world – the University of Texas center in Dallas-Ft Worth, for example, which has been bridging the gap between understanding demand and engineering supply, particularly with attention to disabled veterans and their living needs. But although more interviewing and broader literature reviews might provide a different answer, the engineering and design efforts are not actively seeking out information about what end users want in the way of products or capabilities.

It should be said, however, that a reason for this is that the field of social, complex, friendly robots is still so much in its engineering infancy that the basics of navigation for motion and movement, manipulation of objects, and so on, are the problems in front of the engineers. Drones and even self-driving cars are easier in some respects. So it is premature to expect the engineers to be focused on what the consumers want, even without imagining sci-fi pie-in-the-sky. The field still has to advance on basics before worrying about what actual robots are wanted for what functions.

What this means, however, is that it is premature to talk about any genuinely respectable market model in which the core question can be answered with any real meaning: what is size of the market and what are its price points? There are too many questions and unknowns on both the supply and demand side to address that question. Nonetheless, it is an important exercise to engage in over time, in part because the answers will help reveal the advance of the field.

But the business of commercializing and diffusing into society friendly robots has a second layer of demand and supply. These are the capital aspects, the finance aspects. Who is going to pay for these robots – if not the consumers or users themselves, then who will pay? And on the supply side, where is the supply of capital for research, development, and commercialization coming from now and in the future?

The answers to who will pay are vital in part because they have a large impact on the direction of technology. It seems fairly certain that these devices – whether short term assistive-care mechanical devices or genuine friendly robots in the longer term – will be sufficiently expensive that they will only spread in this sector if government pays a sizable part of the cost. Perhaps this assumption is wrong, but the available literature suggests that this is part of the expectation in the technology community. But, as pointed out in the first section, this is likely to favor the development of less complex, less ambitious robots in favor of more easily created simpler robots or mechanical devices. This is not necessarily incorrect as policy – in fact it is probably right – but it does leave a gap for the demand to develop more ambitious, friendly robots.

On the supply side of capital investment for research and development of the sector, I find my research results so far very inconclusive and even contradictory. I will only say that there appears to be a belief in some quarters that, with respect to this particular sector of the robotics market, research and development funds are lacking, while others believe that the entrance of the tech giants into the field is, and will continue to, change that picture. It is hard to research this area as far as actual numbers; financial statements are hard to come by for an industry that is often privately held or whose numbers are not broken out for this subsector, and although the very welcome and gradually growing robotics industry press is getting better at researching the business and finance of the industry, I still do not have very much confidence in my grasp of the funding for the industry.

Finally, demand side and supply each have a role to play in the regulation of safety and usefulness in the emergence of the assistive care robotics industry. But at this stage, they are empirical unknowns in nearly all ways. The unknowns include such questions as to whether these devices will be regulated – given their role in nursing functions or elder or disabled care – in significant part by the FDA. And there might be a variety of other regulators, quite apart from issues of litigation and products liability, at both the federal and state level. Some of them will be not entirely obvious outside of the industry – the need to train staff or health aide workers or the ultimate users of these machines in their use, not just for safety, but to be able to use them. It is impossible to estimate the impacts that these kinds of regulatory issues will have on price, cost, or anything else related to the market for these devices.

Even if they are only lightly regulated in any or all of these ways, however, such regulatory concerns will also have a supply side impact – as well they should – in favor of the early introduction of normative engineering in order to address what can be known in advance about those issues.

*   *   *

The conclusion from this is that, at least with regard to a sector that has to be regarded as a core area of friendly robotics business, it is premature to be able to discuss the market or the business model in most aspects in ways that would be seriously quantitative. There is a large amount of research that needs to be done to understand that market as an economic matter – but there is also no getting around the fact that a reason for the many question marks is that fundamental parts of the technology are still under development at a stage that is about basics of movement, motion, manipulation, and so on. The promise is there, but a genuinely useable business plan is not yet doable – unless one's answer is simply, "We're Google and we'll spend what it takes - and until we get there, that's our business plan."

But I now want to switch gears quite radically and return to the fundamental question of friendly robots, on the following basis. If we accept, as I do here, that how human beings respond to these machines is a crucial issue for being able to interact with robots at close quarters, then the question is the right approach to that.

One view of this is the famous concept of the Uncanny Valley, as found in Masahiro Mori's justly renowned 1970 article. It deals with a specific aspect of friendly and unfriendly robots, the problem of the uncanny, eerie, or disturbing under specific psychological and design circumstances, but I would like to talk about it as a stand-in, so to speak, for the broader issue of friendly robots. I will talk about it, and then turn and talk about the emerging view, it appears at this early stage, that the problem of human relations to robots is as much one of attachment and affection as the Uncanny Valley. This has, I want to suggest in a very preliminary way, important consequences for the normative engineering of friendly robots – but also a possible difficulty for the ethics of their engineering. Viz., robots will be much easier to commericialize and diffuse into human social life – provided we have decided to go for something beyond "smart appliances" only – if they are "friendly."

Certainly we don't want the robots to be scary or threatening; but there are many business reasons why we would want our friendly robots to be "friendly." The problem is, our ability to make robots friendly – in conjunction with human tendencies to want to attach to robots and project positive affect onto them – means that we risk creating tendencies to trust and finally, reliance on a robot that might not be able to perform as the human consciously or unconsciously believes it can. So the bottom for this whole paper might be: given all of this in the commercialization of friendly robots, as a general design principle, do not create a robot that invites more positive human affect, attachment, affection, trust or reliance than the actual performance of the robot can deliver. Over-trust and over-reliance is likely to lead to tears.

[To the reader at this conference: the draft is going to be quite a bit stiched-together from several papers at this stage; things got more complicated than they should have.]

III. The Uncanny Valley: Human Affective Response to Robots

"The Uncanny Valley," Masahiro Mori's path-breaking 1970 article on human affective response to robot design and human-robot interaction, pursues two simultaneous, yet distinct, intellectual paths. One is the path of empirical psychology, proposing a hypothesis that human affective response to robots is not a "monotonically increasing function" in which gradually increasing positive human feelings about a particular robot rise as more and more "human" features—physical appearance, but also behaviors in many forms—are added to the robot. On the contrary, at some point, the accumulation of human features but also the accumulation of small flaws and errors results in a human perception of the robot as "uncanny," "eerie," "creepy"—hence the Uncanny Valley.

This first is an empirical hypothesis, a proposal for scientific testing through the tools of psychology. It might be right or wrong, or else point toward a much more complicated interplay of human affective dimensions far beyond the simple, exemplary model Mori provides. The other path, by contrast, is not about science or empirical psychological studies, but is instead a theory of aesthetics. By this is not meant superficial appearances, the valences of fashion and style, but instead its formal disciplinary meaning: the interpretive connection between appearance and what lies beneath. Readers familiar with Mori's other writings and life-long activities know that he has expressed this in explicitly religious, Buddhist terms. Yet even those who do not share his religious convictions might still understand sympathetically that, as aesthetic, it is the 'evocation' of the 'sublime'.

Not precisely his words, but they reflect his twin concerns. Mori's concern is to connect form and substance, so to satisfy both the sense and sensibility of the connection—which is not so very far from the Western intellectual tradition's long preoccupation with the "sublime." This means the surface appearances of robots and their underlying actuality, as our present subject, and the ways in which human beings tend to respond to them in feelings, emotions, moods, and affect.

Design of surface appearances matters. This is true, obviously and importantly, for the commercialization, marketing, and general diffusion across society of these new, and in some respects affectively and cognitively confusing, technologies—technologies which, as Mori points out across his writings, are not just about "robots," but about a broader and much older category of "automata."

The cognitive and affective mixed signals and confusion that automata raise are sufficiently provocative that even as hard-headed and empirical a scientist as Peter H. Kahn and his co-researchers in psychology and engineering have proposed these emergent complex robotic systems as a new ontological category. For reasons discussed toward the end of this Essay, I don't think this is the case—and not just because of *entia non sunt multiplicanda praeter necessitate* and all that—but it does point to a need to understand these emergent robots in categories with greater depth than might have been thought necessary in order to market them across many applications in ordinary social settings. Business models by tech companies seeking to place social robots in functional niches in the kitchen, office, school, hospital, retirement home, battlefield, farm, mine, street, retail store, restaurant, and many places more, might discover that what began as a marketing effort based around pure machine functionality, safety, attractiveness of design and appearance, etc., will turn out to depend upon much deeper propensities of users of these machines.

Better to get them on the table now. But half of Mori's point in "The Uncanny Valley" is that the human affective response is not merely a perception by human beings of the robot, studied by empirical psychology, but a perception by human beings of what the robot's appearance evokes. Understanding this depends in no small part, however, upon careful attention to the nuances of language that point to the "direction" of perception and the "relational" subtleties that characterize how beings—us—that are intentional and self-aware, but whose self-consciousness is partly defined by its sociality. This is the intellectual domain of the humanities—moral psychology, criticism as a genre of aesthetics, and intellectual history. In this instance, it offers a help-meet and companion to empirical psychology, because it can help reveal possible confounding interpretations of what people mean that depend upon linguistic and conceptual orientations not captured by an experiment as such.

In economics it is sometimes said that apparently pragmatic, purely practical businesspeople propound views, policies, strategies and plans, unaware how much they are in intellectual thrall to the views of some long-dead economist or philosopher. Something similar is not unknown technology including, today, the commercialization of robotics. Given a new technology possessed of socially and economically transformative possibilities, but which is also possessed of emergent, unexpected behaviors, as well as a tendency to elicit surprising attitudes and unexpected behaviors among people who interact with it—well, trust me, speaking as a lawyer if nothing else, the underlying conceptual categories matter to the business plan.

But technologists in Silicon Valley, far from denying this, have always embraced it—though less as intellectual history than futurism intended to build a scaffolding of ideas and intellectual vision into which technological innovations can be socially framed, as a necessary step in framing the practicalities of commercializing and selling the innovation. The underlying conceptual categories framing robotics matter and ought to matter to robot designers. They matter as well to the "normative engineering" of robots—the features of ethics, law, and regulation that, in tandem with the practical directions of innovations in robotics, will play large roles in structuring the diffusion of robots into everyday social settings.

IV. CONCLUSION: OUT OF THE UNCANNY VALLEY?

Why should we care about this, given that our goal is to climb out of the Uncanny Valley and create friendly robots that will avoid these problems? One thing to note is that Mori himself says in his article that the solution to the problem of the Uncanny Valley is not to try and make perfect simulacra – but instead to back off, to go backwards on the line of human affect, and rather than aim for verisimilitude, instead seek to "evoke": that aesthetic again, and it is not surprising that he makes reference to puppets in the Japanese tradition of the theatre. But more directly, the problem is, perhaps surprisingly, a very different one than the Uncanny Valley.

My impression of parts of the robotics design community is that its belief, using the device of the Uncanny Valley, that the fundamental problem of robot design is to make robots friendly, sweetly humanoid, unthreatening, inviting, is possessed of a personality, and is finally not just a mobile-automated-programmed device, not a thing but a "being." The problem with robot design is to get out of the Uncanny Valley, in other words, and to the sunny uplands beyond. I could be wrong about this, and I suspect than many of us are revising our views to take account of new psychological research into human response to robots.

That research – which is being carried out by people such as Kate Darling, Julia Carpenter, and other members of this interdisciplianary community – seems to indicate that people are inclined (or eager or even desperate) to form attachments and affective bonds with a remarkable variety of robotic machines, and even without affective bonds, surprisingly willing to impute "intentionality" to machine actions and movements.

But, we should be clear, if that's what one sees as the fundamental attitude of human beings toward robots, from a normative standpoint, believing that the problem of designers is to make robots to which it is easier to impute intentionality rather than simply programmed behaviors, and easier to develop affective attachments, is, quite simply, a very bad idea. Or at least, it's a bad idea unless the robot designers are very sure – i.e., deep-pockets sure – that this robot will perform as human beings in ordinary social life, in their homes or on the street or taking care of Grandma, believe it will. There is an important role for normative engineering here, because there are reasons to believe that over time, business models and commercialization based around friendly robots might come to rely on human affective responses to their products that are, in fact, a product of overreliance.

END

[Thanks to readers, comments welcome, and apologies for the the sketchiness at the end; still an early draft. KA]