The Social Robot as 'Charismatic Leader': A Phenomenology of Human Submission to Nonhuman Power [1]

Matthew E. Gladden [2]

CCPE, Georgetown University, Washington, DC, USA

Abstract. Much has been written about the possibility of human trust in robots. In this article we consider a more specific relationship: that of a human follower's obedience to a social robot who leads through the exercise of referent power and what Weber described as 'charismatic authority.' By studying robotic design efforts and literary depictions of robots, we suggest that human beings are striving to create charismatic robot leaders that will either (1) inspire us through their display of superior morality; (2) enthrall us through their possession of superhuman knowledge; or (3) seduce us with their romantic allure. Rejecting a contractarian-individualist approach which presumes that human beings will be able to consciously 'choose' particular robot leaders, we build on the phenomenological-social approach to trust in robots to argue that charismatic robot leaders will emerge naturally from our world's social fabric, without any rational decision on our part. Finally, we argue that the stability of these leader-follower relations will hinge on a fundamental, unresolved question of robotic intelligence: is it possible for synthetic intelligences to exist that are morally, intellectually, and emotionally sophisticated enough to exercise charismatic authority over human beings—but not so sophisticated that they lose the desire to do so?

Keywords. Social robotics, trust, obedience, philosophy of technology, innovation management, posthumanism

Introduction

There exists a rich body of thought that explores the nature and possibility of human trust in social robots. A related topic that has not yet been so thoroughly explored is the question of human *obedience to* robots—i.e., the possibility that robots might participate in leader-follower relationships with human beings. While obedience often occurs in conjunction with trust, the two phenomena also appear independently of one another, which makes obedience worthy of deeper investigation in its own right. Moreover, the consideration of human obedience to robots takes on particular importance, given the fact that we are already taking our first tentative steps toward accepting robots as leaders, and the development of such relationships will likely only expand and accelerate in the future.

One human being who obeys another is not simply trusting that the leader will behave in a certain way; the human being is temporarily or permanently giving up a certain part of his or her autonomous decision-making power and right of self-determination by transferring it to the leader. Although such obedience is manifested most explicitly in contexts such as the military chain of command or vows of obedience within a religious order, it exists in countless subtle, everyday forms, such as when our views on some policy issue are influenced by the arguments of an admired political figure, or when we find ourselves buying a product endorsed by a favorite celebrity. What are the ways in which social robots

might inspire—or even demand—obedience from human beings? And what are the implications of human submission to robot leadership? We explore these questions by focusing on the particular case of human submission to the referent power and charismatic authority of social robots.

1. Referent power and the exercise of charismatic authority by robots

French and Raven proposed a model comprising five bases of social power, to which Raven later added a sixth: coercive, reward, legitimate, referent, expert, and informational power. An effective leader must possess and employ one or more of these bases of power in order to secure, maintain, and exercise influence over followers [1]. It is relatively easy to imagine a future world in which a social robot may be in a position to threaten human beings, dole out rewards to them, fill some official role assigned to it by a human government, cultivate the perception that it possesses expert skills, or control human beings' access to information; these cases would represent the use of coercive, reward, legitimate, expert, and informational authority, respectively.

The case that is perhaps most difficult to imagine is that of a robot who influences human beings by exercising *referent power* over them. A leader possesses referent power when others are drawn to the leader because they feel a sense of fondness or admiration for him, feel a sense of kinship or affinity with him, identify with the leader's moral values, or yearn to win his personal approval. After all, how likely is it that human beings would voluntarily take a robot as their leader because they feel a greater sense of affinity with the robot than with their fellow human beings, or identify more closely with the robot's 'moral values' than with those of human beings, or long to feel the robot's personal approval and affirmation?

One particular manifestation of referent power is seen in a human leader who possesses what Weber calls 'charismatic authority,' which is manifested when followers' obedience arises from their "devotion to the exceptional sanctity, heroism or exemplary character" of their leader and "the normative patterns or order revealed or ordained by him" [2]. Creating such a leader-follower relationship between a robot and human being would involve two participants who—at first glance—seem unsuited to their roles. In order for a robot leader to exercise charismatic authority, it must possess (or at least be perceived to possess) traits such as holiness, divine ordination, moral righteousness, personal charm, or a hypnotic strength of personality—traits which we might expect to find in particular human beings, but which we do not normally see as possessable by robots. And one might imagine that, for their part, human beings are not inclined to spontaneously shower sentiments of admiration, love, and loyalty on robots.

However, a deeper investigation shows that it is not only possible for robots to hold sway over human followers through the use of referent power and charismatic authority, but that the advent of such relationships is fast approaching. The fact that social robots can lead *other robots* is now taken for granted; computer scientists have already designed communities of robots that spontaneously organize their own social structures, with members taking on roles as 'leaders' or 'followers,' and researchers are developing frameworks to allow leaders of robot communities to emerge through robots' use of all the bases of power, including referent power [3]. For our part, human beings are not only able to—but are even naturally *inclined to*—form social relations with computerized systems as though they were human [4], [5]. In many ways, human beings already demonstrate obedience to robots and their machine kin. We listen to the automated voice telling us to step away from the subway car's doors as they are about to close; we follow the instructions of a voice on the phone telling us to 'press 1' for this or 'press 2' for that; we obey traffic signals telling us to stop and go.

However, in these elementary cases, we obey a robotic system not because we consider it to be an intelligent agent possessing social power of its own, but because it represents and extends the power of the human beings who have constructed it and situated it in a particular role, and who *do* possess some social power over us. Human beings do not yet interact widely with robots in a way that reveals the robots to be autonomous agents who possess their own social power, let alone the more specific forms of referent power and charismatic authority. However, the first tentative steps in this direction are already being seen, and I would maintain that humanity stands on the threshold of a wider embrace of such charismatic robot leaders.

2. Three ways in which we envision charismatic robots as leaders

Some scholars argue that human beings' ongoing, almost instinctual embrace of the latest gadgets and technologies shows that "giving robots positions of responsibility is not only unavoidable but is rather something desired and that we are trying to achieve" [6]. In other words, we yearn for robot leaders and are striving—whether consciously or unconsciously on our part—to create them; we want to be led by our mechanical creation. The popular depiction of robots as charismatic leaders that dates back more than a century lends credence to this position. I would suggest that a review of current robotic design efforts and our relationships with robots in fiction allows us to identify three key avenues by which robots might exercise charismatic authority over human beings: 1) through the possession of superior morality; 2) through the manifestation of superhuman knowledge that may take on aspects of religious revelation; or 3) through interpersonal allure, seduction, and sexual dynamism.

2.1. The Type 1 charismatic robot leader: The Saint-Martyr

A charismatic robot leader that attracts and influences human followers through the possession and display of a superior morality we can refer to as the 'Type 1' or 'Saint-Martyr' robot. In fiction, the demonstration of the Type 1's moral superiority is often underscored through the robot's decision to sacrifice its own existence in order to save the lives of its human colleagues. While the robot's destruction means that—in a practical sense—it is no longer capable of serving as a leader of human beings, it also serves to emphasize the fact that the robot was precisely the sort of noble, selfless leader whom human beings *should have* taken as their leader, and whose instructions they could have obeyed with a sense of moral confidence.

One of the most thoughtfully developed examples of a Type 1 Saint-Martyr robot is the Tachikomas (or 'think tanks') from the anime series *Ghost in the Shell: Stand Alone Complex*, based on the manga by Masamune Shirow. In the series' climax, the last surviving Tachikomas become aware of the fact that one of their former human colleagues was in danger, and they rush to their old companion's aid and mount a frenzied battle to rescue him, thereby sacrificing themselves in the process [7]. We realize that the 'glitches' and 'unreliability' that the Tachikomas had previously begun demonstrating were not glitches at all, but rather flashes of their dawning sapience and humanity—a humanity underscored when, in the moment before its destruction, one of the Tachikomas spontaneously utters a plaintive prayer to God, lamenting the fact that it had not been able to do more to save its human companion. Examples of similar sacrifices by Type 1 Saint-Martyr robots can be found in the novel *2010: Odyssey Two* and its film adaptation, as well as in the film *Star Trek: Nemesis*.

It is not difficult to imagine that in the real world, some human beings will prefer to place their trust in robot leaders of this sort, rather than in human leaders who time and again have demonstrated a propensity for injustices such as nepotism, wastefulness, corruption, war, and even genocide.[1] The groundwork for the development of such leaders is being laid by frameworks like Pereira's schema for the use of referent power by artificial agents [3], along with works of science fiction that are helping human society to appreciate and explore the fact that—regardless of what system of moral principles a human being might adhere to—social robots' ability to weigh conflicting interests, utilize their best judgment to make imaginative, just, and wise decisions in accordance with those principles, and take decisive action in moments of great difficulty can, in principle, match or even surpass human beings' abilities.[2]

2.2. The Type 2 charismatic robot leader: The Superintelligence

A charismatic robot leader that attracts and influences human followers through the possession and manifestation of superhuman knowledge is the 'Type 2' or 'Superintelligence' robot. It does not influence human beings by controlling their access to information; in that case it would be relying on informational rather than referent power. Neither does it influence human beings' behavior simply by virtue of the fact that it is perceived to possess unique skills or experience, in which case it would be utilizing expert power. Rather, the Type 2 robot accumulates human followers because it is perceived

to be so much wiser and more knowledgeable than human beings that humans feel drawn to it on an emotional level—not because the knowledge that it possesses is *useful,* but because it is *intoxicating*.[3]

Human beings frustrated and anguished by our intellectual limitations may see a Type 2 robot as a radiant beacon of almost limitless knowledge amidst the depressing world of human ignorance.[4] The Type 2 robot has plumbed the mysteries of the cosmos; its far-reaching sensory input allows it to experience empirical reality in a way that human senses cannot; its cognitive storage and processing capacities allow it to assimilate, correlate, and extract meaning from functionally infinite bodies of knowledge; it grasps the relationship of time and space, energy and matter, life and death, in a way that the human mind is too limited to comprehend.[5] If a human mind is incapable of directly fathoming the knowledge that a Type 2 robot possesses, it might well decide that it should at least pledge its fealty to the one who *has* fathomed it.

In fiction, there is significant overlap between Type 1 and Type 2 robots; the possession of superior factual knowledge and superior moral courage often go hand in hand. In his ethical analysis of entrusting responsibility to robots, Arthur Kuflik also addresses the two kinds of robots as a single species: he considers the possibility of robots "who unfalteringly do what is just and kind" and are "our moral, not merely technical, superiors," posing the question: "[W]ould we—inverting the relationship between creator and creation—faithfully strive to serve them (as many now think of themselves in relation to what they take to be an all-wise and loving deity?)" [8]. Such a benevolent Type 2 robot guardian might use its intimate knowledge of our personal strengths, weaknesses, and longings to guide us gently toward happiness and self-fulfillment. On the other hand, Jerold Abrams argues that a robotic entity (e.g., in the form of a nanoswarm or utility fog) that possesses superhuman sensory and information-gathering capacities could just as easily become a nightmarish overseer of a surveillance society of the type foreseen by Foucault [9].

Fictional examples of Type 2 Superintelligence leaders include the intelligent starship 'Minds' from the Culture series of novels by Iain Banks. Each Mind is capable of carrying on conversations with millions of human passengers simultaneously, and human beings accept Minds as tacit leaders within the novel's utopian anarchist society due in part to their desire to join themselves to the boundless knowledge that the Minds represent [10]. A more modest (and less serious) example of a Type 2 robot leader may be found in the novel *The Stainless Steel Rat Gets Drafted*, in the character of the charismatic robot Mark Forer, whose superhuman knowledge contributes to a planet's decision to adopt the philosophical and economic system that he expounds [11]. In our real world, human beings already regularly and voluntarily delegate a portion of their decision-making process to data-distilling technologies like Google's search engine, Wikipedia, Microsoft's Cortana, or Apple's Siri. It is not difficult to imagine that the sort of loyalty demonstrated toward such brands by billions of information-consumers around the world could someday be transferred into a sense of loyalty and affinity for social robots that serve as the custodians and repositories of even greater knowledge.

2.3. The Type 3 charismatic robot leader: The Seducer

A charismatic robot leader that attracts and influences human followers through interpersonal allure, physical attractiveness, and sexual dynamism represents the 'Type 3' or 'Seducer' robot. The notion of a robot that controls human beings through romantic or erotic appeal predates Čapek's introduction of the word 'robot' in 1920. Setting aside the ancient Greek myth of Pygmalion's love for an animated ivory statue (which was not a 'robot' in the modern sense), we find such early examples as the female android Hadaly in Villiers de l'Isle-Adam's 1886 novel *L'Ève future*. Hadaly is designed by a (fictionalized) Thomas Edison to be as physically alluring as a real human woman, but with an even more engaging intellect and personality. Edison's electromechanical creation is so effective that a British nobleman abandons his human fiancée to take Hadaly as his consort, despite realizing that she is not human [12].

A more widely known depiction of a Type 3 Seducer robot who quite literally and spectacularly takes on a role as a charismatic leader of human beings is that of the robot Maria from Thea von Harbou's novel *Metropolis* and its film adaptation by Fritz Lang.[6] The amoral robot Maria, designed to replace a heroic young woman whose appearance she mimics, is sent by her makers to a nightclub

where she demonstrates her seductive powers by performing an erotic dance that provokes the city's male elite into a frenzy. Having passed this test, the robot Maria then establishes herself as the leader of the city's oppressed working masses and uses her hypnotic rhetoric to incite them into a self-destructive revolution that nearly results in the deaths of their own children [13].

While efforts to produce commercially viable sex robots are already underway, the first-generation devices appear to be envisioned as little more than mechanically animated dolls which—lacking independent judgment, volition, and moral agency—are incapable of 'leading' human beings in any meaningful sense. However, if such physically attractive robots were to be developed in a direction that envisions them as autonomous, intelligent agents that can serve as emotional and intellectual *companions* for their owners (e.g., as contemplated in Project Aiko), it would be possible for such robots to influence the thoughts and behavior of their human loved ones as Type 3 charismatic robot leaders. [14].

3. The decisions that we will (not) make when adopting our robot leaders

If in the future there exists a diverse array of charismatic robots—each with its unique strengths, weaknesses, values, and allure—this raises the question of exactly how we human beings will consciously and carefully sort through the universe of charismatic robots to choose those whom we will anoint as our leaders. I would suggest that the answer to how we will make this conscious decision… is that we will *not*.

Some philosophers of technology adopt a contractarian-individualist approach to human trust in robots, arguing that we can (and should) collectively and individually make conscious, rational decisions about the extent to which we entrust robots with control over our lives. In contrast, Mark Coeckelbergh proposes a phenomenological-social approach which posits that "the social or the community is prior to the individual, which means that when we talk about trust in the context of a given relation between humans, it is presupposed rather than created" [15]. He elaborates, stating that:

> Adaptation to environments … does not necessarily require the exercise of agency. Often we cannot help trusting technology and trusting others, and luckily we often do so without having a reason and without calculation (not even afterwards). In so far as robots are already part of the social and part of us, we trust them as we are already related to them.

Building on this framework, I would argue that in practice, human obedience to a technological entity is something that evolves naturally through innumerable actions, most of which are, in themselves, insignificant. It may occur that at no point during the growth of this relationship does a human being make a conscious 'decision' to submit to the guidance and mastery of a particular piece of technology,[7] and even at the point when the obedience relationship has grown to be fervent and unshakeable, the human being might not even realize that it exists.[8] Neither individual humans nor society as a whole will enjoy the opportunity to make conscious, rational decisions about whether to submit to charismatic robot leadership. Rather, such leadership will evolve gradually, without fanfare, within the network of social and environmental relationships within which we all exist. We will not 'decide' to adopt charismatic robots as leaders; we will only 'discover' after the fact that we are already obedient to them.[9]

4. How long will robots have the desire and ability to lead us?

It is an open ontological and empirical question as to whether the sort of synthetic mind that is capable of demonstrating the moral courage, intelligence, or romantic allure needed to draw human beings to it would necessarily possess the cognitive traits that would cause it to see human obedience as a thing to be cultivated and permitted rather than shunned.[10] Moreover, it is a tautology, but a meaningful one, to point out that robots can only have human followers so long as human beings are capable of being led by them. Coeckelbergh argues that human trust in a robot requires that both parties be able to

communicate with one another using language [15]. From the perspective of current AI research, the development of such communication seems an achievable goal. However, I would suggest that the more relevant question is not whether human beings will develop a world in which robots are able to relate to us effectively using language, but whether our robot progeny will be content to *remain* in that world.

Abrams suggests that today we view Rorty's notion of a 'conversation' among human beings as a sort of worst-case fallback position: if we are unable to achieve moral consensus within humanity on the thorny questions of the day, at least we can maintain a minimal level of social connection by 'keeping the conversation going.' But with the advent of Nietzschean 'Overmen' in the form of nonhuman and transhuman intelligences, the possibility for even the most basic social communication begins to break down [9]. Those persons who have self-fashioned an existence of sufficiently enhanced hyperintelligence will no longer have the desire (or even ability) to communicate with the 'natural' human beings they have left behind. Similarly, our robot creations may become so boundless in their capacities and ambitious in their yearning for self-fulfillment that they lose the desire and ability to communicate with the 'primitive' beings who created them. As Abrams argues, any hope of 'maintaining a conversation' among the sapient inhabitants of our world will become increasingly futile as human beings and our synthetic creations of evolvable robots and AIs and swarm intelligences and living software (not to mention cybernetically or genetically altered humans) fragment into numerous, mutually incomprehensible societies.

This raises a question that remains theoretically and empirically unresolved: is it possible for social robots to exist that are intellectually, emotionally, morally, and aesthetically sophisticated enough to exercise charismatic authority over human beings, but not *so* advanced that their manner of thought and being become incomprehensible to us—that in their pursuit of intellectual (or even 'spiritual') self-fulfillment, they lose their desire and ability to communicate with the disappointingly limited beings who created them? To put it bluntly: will we be able to create charismatic robots that are smart enough to lead us but not smart enough to leave us behind?

The first generation of social robots to possess robust artificial general intelligence will be capable of interacting socially with human beings, because practical realities dictate that this will likely be the purpose for which they have been designed. However, future generations of social robots (which may themselves be built by robots) might 'outgrow' the sort of hardware and software—or mind and body—needed for them to serve as charismatic leaders of human beings. It seems likely that proliferations of different strata of robot cultures (and even civilizations) that simultaneously display varying levels of technological advancement may eventually come to coexist: while the most advanced robots may only be capable of being 'social' among themselves and not with human beings, at any given point there may remain a society of robot 'laggards' that have been designed—or have chosen—to forgo more advanced technological evolution in order to continue serving as our colleagues, companions, and leaders.

Conclusion

Some scholars argue that the restless human urge to develop new technologies demonstrates that we yearn for robot leaders and are striving to create them, whether we realize it or not. In this paper, I have suggested three ways in which—through various artistic, scholarly, and commercial pursuits—human beings are seeking to create charismatic robotic masters to whom we can relinquish portions of our moral decision-making and the responsibility for guiding and controlling parts of our lives.

In some aspects, the creation of such robots has already reached the preliminary design and engineering stage; in other respects, it still stands in the 'pre-preliminary design' stage that is represented by science fiction. As a form of R&D, such fictional works allow us to develop and even conceptually 'test' the technologies that we will, in the future, summon into being. Through such works of fiction, humanity is carrying out strategic planning that explores our reaction to, formulates our goals for, and plots out our uses of technologies that our engineering capacities do not yet allow us to create. However, this design and testing of future social robots through fiction takes place at a fairly general level; it does not involve making conscious, deliberate decisions about whether we, as a human society, will agree to enter into relationships of obedience with particular social robots. The phenomenological-social approach to understanding our human relationship to technology would seem to indicate that

such moments of rational, purposeful decision-making will rarely, if ever, arrive—even as we develop and deepen our individual and collective human submission to charismatic robot authority.

The exact nature of our future submission to robot leadership is not yet known. While it is possible to imagine frightening (and frighteningly plausible) scenarios, we might hope that our obedience to robots will be at least as morally and intellectually beneficial as our obedience to other human beings. This may be possible, insofar as successful social robots can in principle lack some of the flaws of unscrupulous and incompetent charismatic human leaders and can possess a greater abundance of the strengths that allow virtuous and talented charismatic human leaders to organize and guide their fellow humans in pursuit of the common good. For many human beings alive today, the notion that they could ever come to love or admire or be shaped by a robot in the same way they have been shaped by a favorite boss or professor or their own parents might seem farfetched. But for future generations, the notion of a world in which charismatic social robots play no role in influencing and guiding and leading us might seem just as difficult to accept.

Endnotes

[1] A Type 1 robot serving, e.g., as a workplace supervisor could inspire opposing reactions in its human subordinates: some might resent the robot for lacking 'humanity' or for having 'stolen' a human being's job; others may feel loyalty to their leader for its perceived fairness, selflessness, and lack of corruptibility. That some humans are resentful of a robot does not imply that it cannot exercise charismatic authority over others.

[2] The Type 1 charismatic robot's moral superiority does not mean that it will not make mistakes; indeed, the ability to 'err' may be an important leadership trait for a robot. If viewed as an organism, an organization such as a business is constantly experiencing processes of 'reproduction' on both a macro-scale (e.g., the launching of a subsidiary) and micro-scale (e.g., the training of a new employee) that can happen according to the stated plan or deviate from it. The 'mutations' caused by creativity, error, and insubordination produce both harm and breakthrough advances: human laziness might cause an employee to seek out a new shortcut for completing a task, just as jealousy might drive an employee to develop a new solution to garner the praise that was being directed to other employees. While I reject the notion that robots must possess human moral flaws like greed or envy in order to be effective leaders, a robot may need some source of 'non-rational' thought that allows it to think outside of its environment's dominant intellectual and moral paradigms.

[3] Some Type 2 robots may possess 'limited total access' to the sum of human knowledge (i.e., the theoretical ability to access any known fact, constrained in reality by issues like speed and cost); others may someday possess what is—from the human perspective—'functional omniscience,' insofar as they are capable of offering a meaningful answer to any factual question the human mind is capable of formulating.

[4] Type 2 charismatic robot leaders that possess a sufficiently advanced body of knowledge may not be able to impart it fully and directly to human followers in a form that the human mind can perceive or understand; the attempt to impart it directly might result in a sensory, cognitive, or spiritual 'overload' similar to staring into the sun. A robot might only be able to share such knowledge partially and obliquely, by distilling it into simplified metaphors, parables, or symbols that the human mind can understand. The fact that human followers cannot fully experience the robots' knowledge may only serve to increase the robots' mystique as beings who *can* grasp such knowledge. The dynamics at work in human beings experiencing such fleeting, metaphorical glimpses of an ultimately unfathomable reality may not be altogether dissimilar from what has historically been conceptualized as the process of divine revelation or an experience of the 'beatific vision' of God. It would not be surprising if the sentiments and behaviors displayed by human followers of Type 2 Superintelligence robots take on some of the characteristics seen in the human adherents of religions that involve devotion to a supernatural Being who is seen as the font, repository, and revelator of infinite wisdom.

[5] We may find it difficult to imagine a robot possessing such knowledge or sensory capacities if our definition of 'robot' only includes, e.g., current humanoid or zoomorphic devices such as the Telenoid R1 or PARO therapeutic robot. While such devices are incredibly sophisticated and valuable, they represent only one branch of the robotic family tree. A future robot could just as easily lack any permanent and clearly discernible 'body.' For example, its cognitive processes might be distributed among a shifting network of nodes; it might tap into existing civil, scientific, or commercial sensor networks to gather data, rather than possessing dedicated sensors; and it might project itself temporarily into particular environments to manipulate physical reality through access to the Internet of Things, rather than being tied permanently to a single set of robotic limbs. Such a robotic entity could be transnational or—if part of its cognitive processing takes place within satellites—even partially extraplanetary. A more comprehensive ontology of robots is needed in order to identify the full universe of potential robotic entities and their unique characteristics.

⁶ The bifurcated depiction of robots in fiction parallels that of another para-human being, the vampire. On one side are archetypes of 'non-human' robots such as Tik-Tok from L. Frank Baum's *Ozma of Oz* (1907) or The Automaton from *The Master Mystery* (1919). These walking calculators are mechanistic and lack any emotion; they correspond to the vampire as Nosferatu—a pestilential beast that lacks any erotic appeal and shares nothing in common with humanity beyond a roughly humanoid form. On the other side are archetypes like Maria from *Metropolis*, who is suffused with an occult sexual energy and confidence in her ability to bend human beings to her will. Such a robot is so emotionally potent that to describe its traits as 'human' would be an understatement; they enter the realm of the 'hyper-human.' Such robots correspond to the vampire as Dracula—the suave, seductive, aristocrat irresistible in his erotic dominion over human prey.

⁷ The closest that some of us come to making such a conscious decision of submission is in our role as consumers debating, for example, which brand of smartphone to buy. Such a decision brings with it the commitment to a particular brand image, an ethos, a limited universe of apps and accessories, an ecosystem of online purchases and backups and media-sharing that will shape all of the ways in which we interact with our digital world, with the *noosphere*. However, even if we study product reviews and consciously weigh the pros and cons before deciding between the next Galaxy Note or iPhone, how often do we stop to consider whether we truly wish to submit ourselves to the supervision and tutelage of a smartphone *at all?*

⁸ This pattern of creeping, unconscious entanglement with new technology is reflected in the fact that in 1983, only 1.4% of American adults used the internet and only 23% of American computer owners thought that email would be a technology that they would find very useful [16]. Now, responding to emails occupies up to 25% of a typical manager's workday [17], and the inescapability of a technology that can reach employees through their mobile devices wherever they are, 24 hours a day, means that the old wall between 'work time' and 'personal time' has crumbled [18]. However, few of us can recall moments when—after careful investigation and consideration—we *consciously decided* to make email communication a part of our daily routine, or consciously decided for the first time that we would no longer just check email while 'at the office' but also at night or on weekends—or consciously determined that it was acceptable and even desirable to type out an email reply to one person while in the middle of a face-to-face conversation with another.

⁹ Even if some human beings should, out of principle, consciously decide that they will never allow themselves to love or admire or be seduced by a robot, it may as a practical matter become impossible for them to live out this conviction. Our social relations are increasingly mediated by technology; we trust people we have met online because of our virtual experience of them, without knowing their gender or race or place of residence. Grodzinsky et al. note that this increasing 'virtualization' of our interpersonal relationships will combine with advances in the sophistication of robots and AI to mean that there will come a day when we trust and befriend and let ourselves be influenced by online entities because of our direct experience them, without even knowing—or perhaps even being able to discover—whether the entities are human or artificial [19]. No longer will it be simply an online entity's sex or race or age or location that are hidden from us; we will increasingly find ourselves being inspired and led by online personalities whom we experience only as virtual avatars, without knowing (or even caring?) whether they are human or artificial.

¹⁰ A robot must possess many characteristics to serve effectively as a leader of human beings; more conceptual and empirical research is needed to confirm that these characteristics are mutually compatible. I would argue that the full array of necessary traits includes a primary physical substrate for computation (which might include a single serial CPU, multiple CPUs, a CPU-based cluster, CPU-based grid, combinations thereof, or more exotic substrates used in techniques such as photonic or chemical computing); a primary form of processing inputs and making decisions (such as a computer program or neural computing process); a suite of ontological properties such as identity, temporality and change, and physicality; and a suite of cognitive properties such as intelligence, knowledge, and sociality. Desirable but optional traits include the use of supplemental forms and physical substrates for computation that could include human minds or even books. In order to better understand a robot's leadership potential, one can organize these traits into a 'profile' that categorizes them into core attributes (the most essential aspects of the robot's being); secondary attributes (each of which extends one or more core attributes, without which it could not exist); skills and forms (which are the processes, techniques, or arts whereby the robot makes use of core and secondary attributes in order to carry out activities or accomplish certain tasks); and possessions and relations.

References

[1]   D.R. Forsyth, *Group Dynamics, 5e*, Belmont, California, Cengage Learning, 2010.

[2]   M. Weber, Economy and Society: An Outline of Interpretive Sociology, New York, Bedminster Press, 1968.

[3]   G. Pereira, R. Prada, and P.A. Santos, Conceptualizing Social Power for Agents, *Intelligent Virtual Agents*, Lecture Notes in Computer Science **8108** (2013), 313-24.

[4]   M. Rehm, E. André, and Y. Nakano, Some Pitfalls for Developing Enculturated Conversational Agents, *Human-Computer Interaction: Ambient, Ubiquitous and Intelligent Interaction*, Lecture Notes in Computer Science **5612** (2009), 340-48.

[5]   J. Friedenberg, Artificial Psychology: The Quest for What It Means to Be Human, Philadelphia, Psychology Press, 2011.

[6]   H.A. Samani, J.T.K. Valino Koh, E. Saadatian, and D. Polydorou, Towards Robotics Leadership: An Analysis of Leadership Characteristics and the Roles Robots Will Inherit in Future Human Society, *Intelligent Information and Database Systems*, Lecture Notes in Computer Science **7197** (2012), 158-65.

[7]   K. Kamiyama, C: Smoke of Gunpowder, Hail of Bullets – BARRAGE, *Ghost in the Shell: Stand Alone Complex*, episode 25, Production I.G, first broadcast October 1, 2003.

[8]   A. Kuflik, Computers in Control: Rational Transfer of Authority or Irresponsible Abdication of Autonomy?, *Ethics and Information Technology* **1:3** (1999), 173-84.

[9]   J.J. Abrams, Pragmatism, Artificial Intelligence, and Posthuman Bioethics: Shusterman, Rorty, Foucault, *Human Studies* **27:3** (September 1, 2004), 241-58.

[10]   I.M. Banks, *Consider Phlebas*, London, Macmillan, 1987, and the nine subsequent novels in the Culture series.

[11]   H. Harrison, *The Stainless Steel Rat Gets Drafted*, New York, Bantam, 1987.

[12]   A. Villiers de L'Isle-Adam, *Tomorrow's Eve*, trans. by Robert Martin Adams, U. of Illinois Press, 2001.

[13]   F. Lang, *The Complete Metropolis*, DVD, Kino International, 2010.

[14]   "It's not about sex, says Aiko robot inventor Le Trung," *News.com.au*, December 16, 2008, http://www.news.com.au/news/its-not-about-sex-says-fembot-inventor/story-fna7dq6e-1111118332678.

[15]   M. Coeckelbergh, Can We Trust Robots?, *Ethics and Information Technology* **14:1** (2012), 53-60.

[16]   S. Fox and L. Rainie, The Web at 25 in the US: Part 1: How the Internet Has Woven Itself into American Life, *Pew Research Center's Internet & American Life Project*, February 27, 2014, http://www.pewinternet.org/2014/02/27/part-1-how-the-internet-has-woven-itself-into-american-life/.

[17]   A. Gupta, R. Sharda, and R.A. Greve, You've got email! Does it really matter to process emails now or later?, *Information Systems Frontiers*, **13:5** (November 1, 2011), 637-53.

[18]   J. Shih, Project time in Silicon Valley, *Qualitative Sociology*, **27:2** (June 1, 2004), 223-45.

[19]   F.S. Grodzinsky, K.W. Miller, and M.J. Wolf, Developing Artificial Agents Worthy of Trust: 'Would You Buy a Used Car from This Artificial Agent?', *Ethics and Information Technology* **13:1** (2011), 17-27.