Computer Ethics - Philosophical Enquiry (CEPE) Proceedings

Volume 2019 *CEPE 2019: Risk & Cybersecurity* Article 16

5-29-2019

A Ulysses Pact with Artificial Systems. How to Deliberately Change the Objective Spirit with Cultured AI

Bruno Gransche
University of Siegen, Germany

A Ulysses Pact with Artificial Systems: How to Deliberately Change the Objective Spirit with Cultured AI

Bruno Gransche
University of Siegen

Abstract

The article introduces a concept of *cultured technology*, i.e. intelligent systems capable of interacting with humans and showing (or simulating) manners, of following customs and of socio-sensitive considerations. Such technologies might, when deployed on a large scale, influence and change the realm of human customs, traditions, standards of acceptable behavior, etc. This realm is known as the "objective spirit" (Hegel), which usually is thought of as being historically changing but not subject to deliberate human design. The article investigates the question of whether the purposeful design of interactive technologies (as cultured technologies) could enable us to shape modes of human social behavior and thus to influence those customary standards that determine what is considered (in-)appropriate. Which moral rules could (possibly) guide such interventions into the normative fabric of our human social relations? Would it, for example, be appropriate to design technology to exhibit (or simulate) cultured, polite, or moral behavior in order to educate (or nudge) individuals to socially desired behavior? This could be called “deception as a way to virtue” – a thought which can already be found in Kant. Another way of utilizing technology to deliberately influence the social and individual dimension of human behavior would be what the article – inspired by an episode in Homer’s Ulysses – introduces as “Ulysses pacts”: Ulysses, the cunning hero, enters into a pact with his crew to tie him to the mast and disobey all orders while near the Sirens – a strategy of assisted self-constraint. Ulysses-pact technologies might be a means to force us to stick to our commitments by resisting giving in to our weakness of will at a future point in time so that we cannot revise earlier decisions and commitments. Do Ulysses-pact technologies support us, for example, in the autonomous achievement of desired behavior, thus helping us on a path to virtue, or are they but a ruse, devised to drill predictable consumers?

Keywords: philosophy of technology, ethics, objective spirit, culture, manners, social appropriateness, assistive systems, cultured technology, interactive systems, Ulysses, Hegel, Kant.

This article reflects on the consequences of living with technology, with and among artificial agents. Interactive robots and virtual agents are entering our everyday life and playing a new role in our social relations. The company we keep and the way we keep it influence what kind of company we are and how we relate to others. The appropriateness of our social behavior – the ways in which we act and relate – depends on our so-

cial relations – to whom or what we relate and who we are. While it might seem appropriate for our cats to jump on a guest's lap right away, and this kind of behavior[1] might seem odd yet cute and acceptable from our three-year-old child, on the part of our partner it certainly is not and would cause quite some confusion. But what about our interactive robot? Should it offer our guests a handshake or address them by their nicknames – having learned such manners by observation? Or, to put it more generally: Should artificial agents exhibit socio-sensitive behavior; should we build *cultured technology* and teach robots etiquette?

Here, *cultured* technology is on the one hand understood as systems that are *artificially* synthesized, man-made, and engineered, and yet partly *grown* (like cultured pearls): the emphasis is on self-dynamic aspects in their origin, as in the case of machine learning that is 'controlled' by defining thresholds and selecting training data (similar to adjusting the growing conditions of pearls) but not by detailed design of the final structure (Karafyllis, 2003). On the other hand – and particularly so in the context of this article – *cultured* is meant in the sense of showing (simulating) manners and socio-sensitive considerations such as tact, adaptive refinement in speech and behavior, etc. If the most profound technologies phenomenologically escape our attention; if they "weave themselves into the fabric of everyday life until they become indistinguishable from it" (Weiser, 1991, p. 94), then interactive technologies might also need to blend in in terms of manners. Otherwise, they will stand out and disturb social interaction as the artificial addendum they hitherto have been. The Google Duplex chatbot that successfully booked an appointment at a hairdresser's (Welch, 2018) simulated natural language with typical human 'flaws' such as "mhm…" as well as respecting typical rules of etiquette like greeting and thanking, allowing the other time to think and respond, giving options, etc. Its non-human nature thus went unnoticed. We, humans, are most accustomed to interhuman interaction and humanlike interaction behavior is therefore the most predictable and least confusing. Socially interactive artificial agents could be very successful if they managed to be almost indistinguishable from human participants,[2] although they risk falling into the *uncanny valley* (Mori, 2012) if they come close, but not close enough. This article focuses on possible long-term effects of high exposure to artificial agents simulating interhuman behavior. The standards by which the appropriateness of behavior is judged are themselves a result of social behavior. They belong – as part of the customs; the *ethos* – to the realm of the *objective spirit* (Hegel), which results from human action but is not open to individual alteration efforts. Designing socio-sensitive interactive technologies, therefore, not only results in easier and more comfortable human-machine interaction, but also offers a way – by exposure – to purposefully shape the objective spirit; the culture; the ensemble of shared customary practices. If the realm of the objective spirit is thus at the disposal of purposeful designing individuals, this

[1] In this article, I will not specifically distinguish between *action* (understood in terms of means and end), *behavior* (understood in terms of experience and expression; Dilthey (1965, pp. 191–226)), and *process* (understood in terms of cause and effect). *Behaving* – that is, doing something in an unaware habituated nonintentional way – has to be differentiated from *acting* – that is, doing something (choosing and using means) to intentionally actualize an end. When do humans *behave* and when do they *act*? Do artificial agents always perform cause/effect-related processes (answer: yes) or do they ever behave or act in the human sense (answer: no)? Those distinctions and questions will be put aside here for complexity reasons, for they are not the focus of this article.

[2] Recent research suggests that concealment of the non-human nature could make human-machine interaction more cooperative and efficient: Ishowo-Oloko et al. (2019)

raises the question of how it should be designed and how the transformation could be justified.

Alongside several ethical pitfalls (see below), we can also identify a hope: If our day-to-day interactive behavior deposits and sediments itself in the objective spirit in layers of routines, habits, customs, habitus, etc., if designing interactive technologies means significantly shaping our day-to-day behavior, and if the specifics of the objective spirit are the transcendental conditions of further behavior, then *we actually could shape our social capabilities and moral dispositions by shaping technology*. Alas, this broad-scale societal intervention could be dominated by just a very few individuals – for instance, the decision-makers at current tech giants or "Siren Servers" (Lanier). Ethical, cultural, political, aesthetical, etc. implications have therefore to be considered early on. One question that will be addressed here is whether *cultured behavior* can be deliberately induced by shaping *cultured technology*. If so, two areas of potential – or indeed problems – arise. Firstly, one empowering aspect is that design, engineering, and use practices could gain the potential to deliberately shape social behavior and moral judgments, actions, and eventually dispositions. Secondly, that power needs to be democratically controlled, yet tends to accumulate in the hands of just a few.

Cultured behavior via cultured technology

With Hans-Georg Gadamer, a notion of appropriateness can be grasped as *tact (Gadamer, 2011, pp. 13–15)*; as a kind of social sensitivity. David Kaplan linked the topics of tact/politeness, education, self-cultivation and appropriate technology in the following passage:

> What is this sense of appropriateness? For Gadamer it is 'tact.' It is a particular kind of social sensitivity to social situations and the judgment of how to behave in them. Tact is the tacit knowledge of appropriate action for a particular circumstance. It involves knowing what to say and do—and what not to say and do. Although not based on general principles or universal concepts, Gadamer maintains tact is a universal sense that requires of all that we remain both sensitive to particular situations, guided by the wisdom of the past, yet open to other points of view. Although it is difficult to prove any matter of tact conclusively, it is not an irrational concept; it is merely an acquired ability. How does one acquire it? Through education in culture, development, and self-cultivation in society. That is to say, *Bildung.* The only way to acquire interpretive tact is through practice. This connection between tact and practical wisdom has completely dropped out of the contemporary conversation of technology. But what is largely at issue in questions concerning the good life in a technological age is this notion of appropriateness in conduct. Technology is shot through with tact. It answers key questions, such as how things ought to be designed, how they should be used, how they should affect others, how they should be governed. Tact may not provide a precise answer to any of these questions, but if universalist and scientific concepts are ruled out (or not exclusively employed) then all that is left is practical wisdom, developed over time, through *Bildung.* After Gadamer, the notion of 'appropriate

> technology' takes on a whole new dimension. New answers might be found to vexing practical questions concerning technology. (Kaplan, 2011, p. 232)

In order to follow this idea, it is necessary to consider some basic aspects of social appropriateness in interhuman as well as in hybrid (human-machine) behavior. How is behavior judged as socially appropriate – or not – amongst humans? First of all, it has to be perceived via a set of observables such as voice and tone, gestures and facial expressions, posture and positioning, etc. that vary in *time,*[3] *space,*[4] and *mode*.[5] Which manifestation of those observables are judged as appropriate[6] then heavily depends on five interrelated dimensions:[7] 1. the *situational context* (work-meeting, night at the opera, with the family on the beach), 2. the pursued *action or task* (planting a tree, giving a talk, killing 'Il Commendatore' on stage), 3. *individual specifics* (*quod licet Jovi non licet bovi*, the Queen vs. Keanu Reeves vs. Kaspar Hauser, intelligence, attractiveness, reputation, etc.), 4. the *social relation* between the interacting participants (family, authority, status, power difference etc.). All of which finally relate back to 5. higher-level (shared, customary) *standards* like dignity and human rights, which explicitly prohibit distinctions by sex, race, origin, property, or status and trump situational relativity. And yet, given the same human rights, the same observable manifestations (a smile, proximity, even spitting in someone's face, etc.) are judged differently as appropriate for (un-) attractive men/women working as bartenders/priests/bodyguards in China/Mexico/Denmark in 1860/today/2100 and so on.

If insights could be gained about relations between such judgments and significant groups of observables that are typically connected to specific contexts, tasks, and social relations, then could those insights inform the socio-sensitive design of interactive technology? Or, could learning socio-sensitive technology even scan and compile those relations, and could this in turn enable us to learn about the dimensions of our appropriate behavior through an analysis of those machine compilations? Can those observables be made machine-readable – or which of them could be – and if so, could the interpretative steps that relate observable manifestations to appropriateness judgments be

---

[3] This dimension refers to giving or taking time in conversations, specific rhythms in interaction, respecting or ignoring entry, interrupting, or exit points of interaction scripts, etc.

[4] This dimension is, for example, adressed in proxemics, where distances between interacting agents are studied, or in the study of F-formations, where their specific positioning towards each other is studied. For recent work on F-formations, see: "A key skill for social robots in the wild will be to understand the structure and dynamics of conversational groups in order to fluidly participate in them." Hedayati, Szafir, and Andrist (2019 - 2019, abstract). See also for a recent automation attempt: Raman and Hung (2019).

[5] *Mode* could refer to ironic use of observables that indicate the opposite or different meanings than the ones shown, e.g. a pat on the back that might indicate familiarity could strategically be used to create and emphasize hierarchy, just as a compliment elevates the complimenting person into a position of being able to judge and grant praise.

6 The judgment on appropriateness is of course always a judgment on inappropriateness. Watts, for instance, proposed a conceptual framework that distinguishes non-politic/inappropriate (negatively marked) behavior from behavior politics/appropriate behavior that is subdivided into unmarked behavior and positively marked behavior. This constitutes a continuum that ranges from rude, impolite, non-polite, polite to over-polite behavior (Locher and Watts (2005, xliii)). For reasons of brevity, this will only be mentioned as judgments on appropriateness, if not otherwise specified.

[7] The credit for this overarching five-dimensional framework on the factors of appropriateness must go to the team of the research project "poliTE - social appropriateness for artificial assistance" (https://polite.fokos.de/en/home_en/#).

delegated in part to highly automated systems and integrated into socio-interactive technologies? And, if so, why should they be? There is a set of observables that are currently being researched and prototypically implemented in interactive systems, like machine judgments of the proper distance at which to keep participants (proxemics, e.g. Pepper). The entire strand of research on emotion-sensitive adaptive technology; humanlike interaction; assistive, companion technology; social robots, etc. shares an implicit premise: Humanlike or 'natural' – or in at least less artificial – interaction equals better interaction. Why would anyone build machines that simulate human behavior up to and beyond the uncanny valley? There are some obvious challenges, for instance potential deception that confuses users about the actual nature of the interacting entity – which would mean that those systems would pass the Turing test, at least temporarily. Such confusion could lead to a denial of fundamental respect for and the human rights (along with working rights and conditions) of fellow human beings in cases where AI systems instead of people are staged and assumed.[8] Another risk could be to invite parasocial bonding – i.e. falling in love with fictional characters like James Bond or nonhuman entities like God or robots. Parasocial relations with technology could have problematic consequences such as saving one's beloved robot (or car) instead of rescuing a fellow human in an emergency, or avenging property damage with bodily injury.

Nonetheless, cultured technology could offer considerable potential for pleasant human-technology interactions, and this could have positive effects on a person's mood, health, motivation, or performance. Respect researchers, for instance, consider a recipient-focused concept of respect in which a certain kind of respect is actualized if, and only if, a person feels himself or herself respected by someone regardless of whether his or her interlocutor does indeed respect him or her (van Quaquebeke & Eckloff, 2010). Even though technology could not respect a person in a fundamental sense – for instance, it would not be able to choose whom (not) to respect and would not have normative preferences that could shape such a choice – it could nonetheless simulate respect well enough for a person to feel respected by artificial agents. That would – according to a recipient-based concept of respect – be sufficient to actualize significant positive effects on the motivation or performance and health of the person who feels respected. Another potential advantage of socio-sensitive technology could be an overall reduction in inappropriate interruptions by interactive robots or AI systems that are involved in hybrid social settings such as a conference coffee break that might be catered by robots. This article will focus on potential indirect or long-term effects of exposure to socio-sensitive cultured technology on the part of humans such as potential upskilling, training, education and so on. The basis of the supposed effects lies in the mechanism

[8] This is strategically the case with many data economy services that are staged as AI services whereas legions of human precarious workers substitute or enable the service. "Digital information is really just people in disguise" (Lanier (2013, p. 15). "However, Amazon is also exploring how to get non-elite service jobs out of the way of the Siren Servers of the future. The company offers a Web-based tool called Mechanical Turk. The name is a reference to a deceptive 18th Century automaton that seemed to be a robotic Turk that could play chess, while in fact a real person was hidden inside. The Amazon version is a way to easily outsource – to real humans – those cloud-based tasks that algorithms still can't do, but in a framework that allows you to think of the people as software components. The interface doesn't hide the existence of the people, but it still does try to create a sense of magic, as if you can just pluck results out of the cloud at an incredibly low cost" (Lanier (2013, pp. 169–170). For recent work on the invisible human workforce that powers the Web, see: Gray and Suri (2019).

of habituation that will be explicated with – an Aristotelian – Kant and that allows the undeniable deception in interaction with cultured technology to be reframed as potential, allowing us to think of deception as beneficial rather than harmful.

Deception as a way to virtue

There has been some critique and concern about human-machine interaction such as Amazon's Alexa or Google's assistant Google Home. The response of the tech giants was to implement very basic manners into the speech-based interaction: "Google Assistant's Pretty Please helps your kids mind their manners" (Gebhart, 2018) or "Alexa and the Age of Casual Rudeness" (Gordon, 2018). Interestingly, the supposed effect that people, especially children, who predominantly interact in the form of spoken commands lose their manners and verbally degenerate to the grammatical imperative is not yet scientifically established. However, based on fundamental insights into the acquisition of skills – *use it or lose it* – such an effect seems not unlikely, which is why Google and Amazon reacted. Google's Pretty Please simulates manners if addressed in a respectful way. 'Hello Google, turn on the lights please.' will be answered, for instance, with 'Thanks for asking so nicely.'

Now politeness, as a special form of appropriate behavior and a symptom of being cultured, has a difficult relationship with sincerity and truthfulness; it implies some degree of mandatory deception and lying. Tact as a kind of "social sensitivity", for instance, can be shown by *dissimulation*, that is, pretending that something – such as impolite or embarrassing mishaps – did not happen even if it actually did. The fact that we often have to choose between an honest and a polite answer also shows that it is difficult to give an honest polite answer. Yet the politeness deception is not actually a case of genuinely deceiving someone and thus is not only not harmful, but indeed beneficial to the interlocutors and ultimately to the polite person him-/herself. In order to understand the beneficial dimension of polite or cultured behavior and the potential impact of cultured technologies, it is worth looking at the following passage from Kant:

> § 14. Collectively, the more civilized men are, the more they are actors. They assume the appearance of attachment, of esteem for others, of modesty, and of disinterestedness, without ever deceiving anyone, because everyone understands that nothing sincere is meant. (Kant, 1996, p. 37)

Being an actor means playing a role, staging an 'As if' layer that by definition differs from the layer of the actual.[9] Since ancient theater, there have been two possible aspects to such pretense: either showing something that is not – *simulating* – or not showing something that is – *dissimulating*. Ancient Greek theater had two paradigmatic roles

[9] The philosopher Hans Vaihinger – who drew on Kant and inspired Sigmund Freud, amongst others – described the omnipresence and importance of our "useful fictions" in his *Philosophy of 'As if'*: "'As if', i.e. appearance, the consciously-false, plays an enormous part in science, in world-philosophies and in life." Vaihinger (1935, xli). The notion of the consciously-false applies to the Kantian non-deceptive deception of politeness. Being polite is being false, but everybody knows it and thus no one is harmed by the falsehood.

hereof, *Alazon* the simulant – the poser, imposter – and *Eiron* – hence irony – the dissimulant (Buttkewitz, 2002, p. 23; Gransche, 2017). The reason why this untruthfulness or deception – both concepts Kant otherwise argues against in the interest of self-respect, which for him is to respect mankind represented in oneself[10] – is not harmful but beneficial lies in the fundamental familiarity of all participants; it is because "everyone understands." Kant sees an active habituation mechanism at work that links *hexis*[11] – clearly an Aristotelianism – to virtue:

> Persons are familiar with this, and it is even a good thing that this is so in this world, for when men play these roles, virtues are gradually established, whose appearance had up until now only been affected. These virtues ultimately will become part of the actor's disposition. To deceive the deceiver in ourselves, or the tendency to deceive, is a fresh return to obedience under the law of virtue. It is not a deception, but rather a blameless deluding of ourselves. (Kant, 1996, pp. 37–38)

In highly simplified terms, this means that the saying 'fake it until you make it' also applies to the cultivation of virtues. Kant then relates this fundamental anthropological diagnosis to cultured, "good and honorable formal" behavior and to politeness, and emphasizes its enabling function for virtue:

> Nature has wisely implanted in man the propensity to easy self-deception in order to save, or at least lead man to, virtue. Good and honorable formal behavior is an external appearance which instills respect in others (an appearance which does not demean). Womankind is not at all satisfied when the male sex does not appear to admire her charms. Modesty *(pudicitia),* however, is self-constraint which conceals passion; nevertheless, as an illusion it is beneficial, for it creates the necessary distance between the sexes so that we do not degrade the one as a

[10] Kant is rather famous for rigorously condemning untruthfulness. He argues in *The Metaphysics of Morals* that lying does not need any harmful effect (against others) to be reprehensible because even if no one (else) is harmed (or even if someone is saved by a lie), lying always harms a) the lier, b) mankind represented in the lying person, and c) the law of truthfulness – which is the basis for all duties based on contracts. The idea of a harmless lie (as in politeness) seems inconsistent in this perspective. Two aspects stand against that impression of inconsistency: firstly, the 'empty signs of well-wishing' or the 'non-deceiving deception' Kant sees in politeness do not qualify as a lie in the strict sense of *The Metaphysics of Morals.* Especially the criterium of importance is hardly met because polite gestures are usually not a matter of life and death – contrary to the famous example of lying to a murderer about the whereabouts of a person hidden in one's house (Kant (1797a, pp. 429–430)). Secondly, in the article *Über ein vermeintes Recht aus Menschenliebe zu lügen* (On a supposed right to lie from altruistic motives [my translation]) Kant repeatedly emphasizes that if a truthful answer would cause harm, avoiding the answer is best. Only where the answer cannot be avoided, truth should be spoken regardless of any harm. "Each human being has not only a right but even the strict duty to be truthful in statements he cannot avoid making, whether they harm himself or others." [my translation]. Original: "Jeder Mensch aber hat nicht allein ein Recht, sondern sogar die strengste Pflicht zur Wahrhaftigkeit in Aussagen, die er nicht umgehen kann: sie mag nun ihm selbst oder andern schaden." (Kant (1797b, p. 428). Politeness can be seen as an art of avoiding potentially harmful statements. Being polite – with Kant – could mean the art of walking the line between not making harmful statements and not lying in the strict sense at the same time.

[11] "Aristotelian *hexis* denotes not just ethical behaviour, but also knowledge and technical skill, otherwise known as *epistēmē* and *tekhnē.* Aristotelianism, indeed, accommodates practical, theoretical, and poetic or technical 'ways of being' *(hexeis):* in each case these are stable qualitative aspects of the subject *and* of the objective situation. From a doctrinal point of view, *hexis* thus has a very wide scope that includes the domains of *theôria,* of *poēsis* and of *praxis;* thus it is important not to confine it to ethical behaviour." Rodrigo (2011, p. 7)

> mere instrument of pleasure for the other. In general, everything that we call decency (*decorum*) is of the same sort; it is just a beautiful illusion.
>
> Politeness *(politesse)* is an appearance of affability which instills affection. Bowing and scraping (compliments) and all courtly gallantry, together with the warmest verbal assurances of friendship, are not always completely truthful. 'My dear friends,' says Aristotle, 'there is no friend.' But these demonstrations of politeness do not deceive because everyone knows how they should be taken, especially because signs of well-wishing and respect, though originally empty, gradually lead to genuine dispositions of this sort. (Kant, 1996, pp. 38–39)

This is not the place to discuss issues of potential sexism in a piece of 18th-century anthropological work. In principle, Kant is not wrong about the need for or satisfactory effects of compliments, admiration, or other forms of positive social and relational work and acknowledgment. It is safe to say that it holds true not only for womankind but for humankind in general, regardless of whether the "bowing" is a literal bending of the spine or performed through social media likes and followers. The important point here is that Kant gives a reason for why the illusion that is created by self-constraint is beneficial: "for it creates the necessary distance". It is the same benefit Gadamer ascribes to tact: "Thus tact helps one to preserve distance. It avoids the offensive, the intrusive, the violation of the intimate sphere of the person." (Gadamer, 2011, p. 15) Kant speaks about the distance between the sexes that is necessary to ensure the prohibition of instrumentalization as set out in his Categorical Imperative; to treat humanity *never merely as a means to an end, but always at the same time as an end*. To use a human being as a mere instrument, therefore, is to degrade that human in his or her human dignity. Cultured behavior, for instance, the beautiful illusion of gallantry, enables the necessary distance between "the sexes" that prevents degradation to a mere object of pleasure. This holds true for social interaction between humans in general, in other words men, women and any sexes in any direction. If we are about to introduce artificial interaction partners, one (non-capitalist) argument for making them socio-sensitive is that they could thus be made cultured or polite – that is, capable of self-constraint – in order to ensure the *necessary distance* between humans and artificial agents (and the disguised humans behind them) so that we do not degrade the one as a mere instrument of profit for the other. We need, therefore, to constrain ourselves in social interactions not to be overly direct or brutally honest and not to intrude on others. Modesty *(pudicitia)* is one facilitator of distance and should be used not for the sake of artificial systems – they cannot be offended – but for ourselves to establish virtue gradually. The boomerang effect of being rude to someone with whom we interact justifies the effort of politeness. It is considered morally wrong to mock someone even if that someone is not able to detect the mockery; it could therefore be seen as equally reprehensible to mock young children and to mock artificial agents. Ultimately, we are not polite to others –

only or mainly – out of consideration for them, but for ourselves.[12] In this sense,[13] it does not matter who or what the others are. This is one reason why we might consider manners in human-machine interaction and keep saying pretty please to our interactive systems and maybe ask for help rather than demand it from our artificial assistants. On the other hand, manners are not the only way to ensure a "necessary distance" between humans and artificial agents along with all other humans behind those systems (operators, data clients, etc.): carefully crafted levels of personalization and intimacy in those human-machine relations are another, if not the major other one. Any behavioral distancing effort will remain futile if the systems continuously gather data and those ultimately in charge of the systems – 'The Lords of the Siren Servers', to coin a Homeric-Tolkienian Lanierism – compile and sell the most personal and intimate information. As cultured behavior overcomes harassment or impertinence, we should design and enable cultured human-technology relations that help us overcome current *technological impertinence*.

Deceiving machines

By virtue of creating necessary distance, deception, which is a cunning self-deception, becomes beneficial. The mainstream and as such naïve moral demand 'Be honest!'[14] has to be put into perspective, and deceit and illusion are normatively reversed from malicious to beneficial. But why does deception having beneficial effects defuse its potentially harmful effects? Kant offers a reason for this as well: "these demonstrations of politeness do not deceive because everyone knows how they should be taken" – and – "everyone understands that nothing sincere is meant." Just as ancient Greek theater audiences shared a (dis-)simulation literacy and thus knew how to take an *Eiron* (search for the hidden) or an *Alazon* (distrust the shown) on stage, most people share a basic culture-specific politeness literacy. Thus, in cultured social interaction, there is deception but "without ever deceiving anyone" (Kant). It follows that potentially harmful behavior exists but without harmed victims because a shared understanding prevents the actualization of harm. In addition to that, through the mechanism of habituation, this kind of deception and self-deception leads to genuinely virtuous dispositions (*hexis* as the stable state of having those dispositions). The shared knowledge of how such deception is to be taken (*hexis* as cultural/ethical knowledge) and customary practices of how it is usually taken or meant are necessary preconditions for this deception to be harmless and beneficial. Within human-machine interaction, if we maintain good manners even towards artificial agents, we could benefit from the virtuous enabling effects while worrying even less about potentially harmful effects, because machines cannot be deceived

[12] "It [*Bildung* as keeping oneself open to what is other, BG] embraces a sense of proportion and distance in relation to itself, and hence consists in rising above itself to universality. To distance oneself from oneself and from one's private purposes means to look at these in the way that others see them." Gadamer (2011, p. 15)

[13] This holds only for the mentioned boomerang-character. Of course, the ban on instrumentalization as expressed in the Categorical Imperative refers to "humanity" and therefore cannot justify polite behavior towards nonhuman artificial agents.

[14] This is still a wide-spread go-to principal as seen for instance in one of the 12 rules for life by best-selling author and influential public mentor figure Jordan Peterson: "Tell the truth – or, at least, don't lie." Peterson (2018).

in a moral sense. ‘Deceiving’ machines (e.g. treating the robot politely) is not problematic; being deceived by machines or by other humans via machines actually can be. Living with socially intervening cultured – that is, deceptive! – technology is a problem because a shared knowledge of how to take that technology or those machines cannot easily be assumed and corresponding defusing practices have yet to be found, formed, and habituated. This means that as well as designing cultured technology we also need to develop cultural techniques for relating to new technologies – and for relating to new human-technology relations on a higher level[15]. Such a development takes time, probably longer than current technological innovation cycles.

To sum up, social interaction that respects human dignity requires the necessary distance between the interacting parties. Cultured, “good and honorable”, polite behavior is one way to create that distance. It is not the only one and not a sufficient one. The distancing effect is achieved by playing a role and via deception, yet this deception is harmless – because everyone understands it – and beneficial. The benefit lies in the aforementioned distance as well as in the potential to lead to genuine virtue. The mechanism that enables the latter benefit is habituation, or: *fake it until you make it*. This connection has to be transferred to artificial interacting entities. Artificial agents are never in danger of being harmfully deceived. In contrast to human agents, therefore, they do not even need specific knowledge in order to defuse the deception. To artificial agents, we can put on a show that can help us gradually to establish virtuous dispositions, while worrying even less about potentially harmful deception. The challenge – apart from implementing a complex phenomenon of social appropriateness and technology – lies in the potential harm that cultured technologies’ deception imposes on human participants. Firstly, people – being part of humanity – are at risk of becoming a mere instrument (most likely of profit) in the process; necessary distance has to be created and preserved in order to prevent technological impertinence and to protect human dignity. Secondly, people are absolutely deceivable in a moral sense; the harmful potential of machine deception must be met on the parts of humans with a corresponding knowledge of how to understand the technology in its ‘As if’ layers. Such an understanding and knowledge are a prerequisite for benefiting from deceptive technology and at the same time precluding harm from that technology. What exactly is artificial intelligence? What do algorithms do? How are all the systems with which I interact interrelated in a data economy and as a technosphere? If we further engage with interactive systems; if we want to enjoy the benefits of socio-sensitive technology, the deception competence inherent in a politeness literacy that facilitates and simultaneously protects interhuman social relations must be complemented by technology literacy, data literacy, AI literacy and so on.

Ulysses pact… with technology

[15] This higher-level relation to new human-technology relations is then a meta-relation of sorts, meaning that the new cultural techniques for dealing with new technologies that are segmented as new routines and perhaps future traditions must in turn be embedded in a cultural, normative context.

The notion of self-constraint is crucial in this context. Kant mentioned modesty as self-constraint that conceals passion. Concealment and self-constraint are at the very core of civilization and being cultured. Impulse control or the ability to defer rewards, for instance, are essential developmental steps every child has to take in order to become sociable. We use hygiene processes and products to conceal or not to impose our unaltered body (e.g. its odors or noises) on others. A groomed and controlled body is nature constrained by culture (whose standards, of course, vary greatly depending on the culture in question).[16] Self-constraint is a difficult task that requires certain dispositions, practice, skills, and strategies. Such a strategy is evident in Kant's detour to virtue via self-deception and the habituation mechanism. It is useful and necessary because it is difficult just to decide to be and then be genuinely virtuous in the intended way. Even if we want to have certain virtues, we struggle and, as with everything that we struggle with, we use technology to have it anyway. If we struggle to lift something, we use a lever. If we are unable to run or want to survive a marathon – unlike Pheidippides, who died upon arrival – we go by car. If we fail to actually exhibit the desired behavior, we use strategies, along with several technologies. One such strategy is called a *Ulysses pact*.

Ulysses, the most witty, resourceful, strategic, and cunning hero (all epithetically summarized as *polymechanos*) in Greek mythology, encounters the Sirens, bird-women who sing beautifully and lure sailors – completely absorbed by the Sirens' performance – to their death by shipwreck and drowning. Ulysses, however, has adequate knowledge of 'how they should be taken'. This heroic Siren literacy includes the fact, for instance, that the Sirens' power lies only in their song and not in their visual appearance, so that you lose self-control and die by hearing them, yet not by seeing them. The Sirens are – just like the Sphinx – mantic creatures, which means that they possess prophetic knowledge and great wisdom (mantic truth). Ulysses desires this truth, which is why he does not want to avoid them completely. Due to his Siren literacy, he knows that he would lose his self-control the moment he listened to that sung truth. He wants to have it but struggles to get it. The solution is a pact he makes with his crew. He orders them to tie him to the mast of his ship – which is a form of assisted self-constraint – and – which is the cunning part – to disobey any orders, especially to untie him, while under the influence of the Sirens. The crew then seal their ears, the gap in their defense against Sirens, with wax and navigate the ship into the sweet spot well within hearing range of the Sirens yet still at a safe distance from the cliffs. Fortunately, navigating a ship works with a bound and temporarily demoted captain and deaf but unbound crew members. Thus, Ulysses is able to learn the Sirens' wisdom and survives to live with it. This is more or less the essence of technology use: enjoy the benefits while not having to suffer from the harm.

Today, Ulysses contracts are a subject of discussion in clinical ethics, for example where patients have a psychological condition that leads them to distrust their future selves.[17]

---

[16] Interestingly, the German word for toiletry bag is *culture bag* (Kulturbeutel).

[17] "Like Ulysses afraid to be lured by the sirens, these patients feel they cannot trust themselves and are prepared to give up their freedom for a limited time in order to protect themselves. For example, BPS [borderline personality syn-

> Such contracts are a means to ensure against *akrasia*—weakness of the will, or not being willing to follow the course of treatment—or relapse of mental illness—not being able to follow the course of treatment. Ulysses contracts have many potential applications. In general medicine, patients might use such contracts to aid in quitting smoking or to ensure that a course of painful, yet beneficial physical therapy is completed. In mental health, patients could use the contracts to ensure against relapse of mental illness. (Spellecy, 2003, pp. 375–376)

Ulysses insured himself against a temporary Siren-induced *akrasia*, just as people might against alcohol-induced *akrasia* on a night out – forming a pact with their friends to ensure a safe journey home because their substance use literacy tells them that, once inebriated, they will lose the will to quit the party at an otherwise sensible point. Kant's detour to virtue is a strategy that instrumentalizes self-constraint (modesty, politeness) in order to achieve long-term self-deception (pretending to care about or respect someone or something) that ultimately results in the desired moral disposition (being a caring and respectful person). Here, what someone wants to 'have'[18] is not mantic truth but genuinely virtuous dispositions. But there is good reason not to trust our future selves to follow through on our past or present resolutions. Otherwise, the gyms would be overcrowded throughout the year and not just in January and before the beach season. We clearly need strategies to help or even force our future selves to exhibit behavior that we desire today or that we desire in principle, yet tend to put aside all too easily. For lack of a ship's crew to make commanded self-constraint contracts, we turn to the common strategy of using technology as a surrogate for human staff, a process that is often called automation. We can make a Ulysses pact with our smartphone and grant a given application actual punishing (constraining) power such as the power to transfer predefined amounts of our money. For instance, the application *StickK* (StickK, 2019) is a Ulysses pact platform that is designed to help users to "finally stick to their commitment" – be it a ten-minute walk after each meal, quitting smoking, working at least one hour each day on their dissertation, etc.[19] Even though "the market for self-punishing products may be small" (Ubel, 2014) – as yet! – the idea of automating sanctionable self-constraint on an everyday level in order to achieve a desired outcome, behavior, or disposition on a higher level is compelling. In a way – albeit simple and one-dimensional one – Google's and Amazon's 'ask nicely' functionalities position their devices (e.g. Echo) and the connected infrastructures and services (e.g. Alexa) as potential parties to

---

drome, BG] patients sometimes come to psychiatric emergency units and display a help-seeking behavior and pronounce suicidal intentions, but at the same time, they communicate, directly or indirectly, that voluntary care is not an option for them staying at the hospital since they cannot trust themselves." Lundahl, Helgesson, and Juth (2017, p. 83)

[18] NB: *hexis* in Aristotle is associated with *possession*, *having* something (knowledge, skills, virtues, etc.), yet in a way that is interrelated with *being*. *Having* virtues equals *being* virtuous, just as *having* shoes on is the same as *being* shod and *having* skills as *being* skilled, etc. See Rodrigo (2011).

[19] "Some people commit the money to a charity they detest, to further motivate themselves—knowing that the failure to take a ninety-minute bike ride will mean contributing twenty dollars to an organization on the wrong side of the right-to-life/right-to-choice debate offers a certain motivation to hit the road. According to a *Harvard Business Review* study, published in April, the company's success rate for contracts without stakes is 42.7 per cent; with stakes, it is 82.8 per cent. And if the money is going to a charity the user dislikes, the success rate is even higher, 87.1 per cent." Ubel (2014)

a Ulysses contract with whom users whose resolution is constantly tested by a plethora of everyday Mini-Sirens can team up against their own weakness of will (*akrasia*). Kant's way to virtue is facilitated by the detour via self-deception and habituation and it can be protected from vicious *akrasia* by concluding Ulysses pacts with artificial crew-members: with apps, devices, services, and infrastructures. Google's "Pretty Please" feature and even today's pre-commitment automation are but an attention-seeking diversion that aims to soothe concerns that might arise in the face of the pervasion of assistive technology in our everyday lives. Emerging and future socio-sensitive interactive technology that could potentially consider many manifestations of the above-mentioned dimensions of social appropriateness – *situational context, action or task, individual specifics, social relation, shared/customary standards* – could be a powerful crew with which to contract.

Artificial Ulysses crew – sassy assistants

Artificial Ulysses crewmembers could help to bridge two interrelated gaps in consistent action: one in time and one in awareness. Firstly, there is a time gap in every action due to the sequential nature of tasks. We have to perform one after the other in order to get something done. Each task or action consumes a certain amount of time or requires specific timing. Ethics – the discipline of 'What should I do?' – addresses this time gap at its core as it always asks 'What should I do *now* or next, in order for X to happen in the future?' In turn, previous actions determine the scopes of possible responses to this question and the options now available. The basic problem here is why a present individual should be entitled to make decisions for a future individual, a decision on which the latter must follow through and for which the latter can be held responsible. After all, we clearly differ in terms of knowledge, preferences, and will over time. On the other hand, why do we allow and often demand present individuals to revise and ignore their past choices? In order for society to function and groups to succeed in cooperation, there has to be some degree of predictability and accountability (Bratman, 2018) and strategies such as announcements, contracts, threats and promises serve to bridge this time gap. By one part of the objective spirit – namely law – you have to follow through on the consequences of a contract that your former self concluded even if your present self no longer shares its object or implications. However, this pragmatic, social, inter-individual reason does not apply to the intra-individual time gap: I can always change my mind; this is a sign of education and personal development and the opposite of stubbornness, dogmatism, or fundamentalism. A Ulysses pact is a strategy for using inter-individual cooperation – through the social construct of a contract – in order to ensure an intra-individual bridging of the time gap. This is a difficult phenomenon: why does the will of a contracting patient with borderline personality syndrome (BPS) at moment t_1 trump and constrain that person's will at moment t_2? In other words, why are Ulysses contracts allowed at all and specifically in a clinical context? To move on from potentially pathological special cases to our everyday condition of action: why do *I at* t_2 have to follow through on a decision that *I at* t_1 made before; why do we appreciate reliability, commitment, and consistency? Simultaneously, why do we invest so much as a society in changing an individual's knowledge and preferences deliberately, a process that is

called education? And if changing one's will, knowledge and preferences is paramount for any society and the notion of progress, how can we allow strategies that preempt actions for following through on the consequences of such a change? How we value consistency or spontaneity, predictability or flexibility, stability or change are fundamental questions that need to be scrutinized before socio-sensitive interactive technology or artificial Ulysses crewmembers are introduced into our everyday life. Interactive technology can preempt or facilitate significant potential for future change, progress, and development. The contracts we enter into can substantially transform how open our future is and the range and type of possibilities open to us.

The potential to use such contracts with artificial crewmembers to facilitate progress and further open up our future depends on the second gap, the awareness gap. Although our will changes over time, we often have a number of goals that are fairly invariable; this applies in particular to more abstract objectives such as to be more attractive, to lose weight, to work out, etc. or to be wiser, to study, to engage in discussion, etc. Yet we tend to shape our lives and daily decisions in relation to different levels of awareness of those objectives at any moment. A procrastinating philosopher, for instance, does not lose the will to write a book, but temporarily loses track of the goal or has doubts about the suitability of certain steps while, for instance, watching *Epic Rap Battles – Western vs. Eastern Philosophers* on YouTube – which might or might not ultimately somehow help with the book project. The diet-breaker does not lose the will to get healthy or attractive through dieting when eating foie gras with melted raclette cheese. S/he simply values the enjoyment and hedonistic effects more highly at that moment than the health-related long-term effects.[20] Is this *akrasia* or a sign that the higher-level goal was ill-chosen, maybe even merely adopted as a result of social mainstream pressure? It is impossible to think of options in binary terms of suitable or non-suitable for a specific goal or end. It is impossible to grasp all the means that can actualize a certain end, just as it is impossible to conceive of all the ends a single means could actualize. Using emerging technology in order to ensure an (initial) pursuit of goals and to render that immune to second thoughts and misappropriation has to deal with this changing normativity and this gap of awareness or heightening and lessening levels of intentionality. The dynamic of this means-end-goals-wish hierarchy and of the varying awareness is intricately interlinked with the development of our preferences and objective systems over time. It is not so that we reflect on a *wish* – e.g. the good life – come up with suitable *goals* whose achievement brings us closer to that wish – e.g. being loved, respected, healthy, and rich – come up with suitable *ends* that bring us to those goals – e.g. being kind and respectful, eating healthy food, being industrious – and then come up with means that are suitable for actualizing those ends – e.g. practicing, learning, studying, etc. Rather, we jump in at some point in this chain and when we encounter difficulties in finding a means, actualizing an end, or reaching a goal, or when we see other normative takes on what might be a part of a good life, how to weight wealth and affection, etc., we adjust the entire chain. Difficulties in getting comparatively rich? It is much more important to have friends and love anyway. Not so lucky with your

[20] For a compelling discussion on health and security, for example, as ends in themselves versus a hedonistic and aesthetic way of life using and occasionally suspending health as a means to more meaningful ends that are "worth living for", see: Pfaller (2018).

social relations? Success is paramount in life and money will buy you respect and eventually affection anyway. Your first novel flopped? Helping people is much more meaningful than becoming an eccentric writer anyway.

However, bearing in mind the challenge of this fluid normativity of ends, goals, and wishes, assistive systems could be deliberately designed to help compensate for undesired awareness loss and *akrasia*. Therefore, we could deploy an entire crew of contracting assistive systems that help us follow through on our higher-order resolutions and goals, by constraining – as commanded self-constraint – our options and the availability of our means on a lower-level. Competence-oriented assistive systems (Gransche, 2018), for instance, could strategically refuse assistance on a lower means-end level (not navigating onto the Sirens' cliffs even if commanded) in order to assist with higher end-goal-wish connections (listening to the Sirens' song and surviving it). In the context of socio-sensitive systems and their potential role in leading to genuine virtue (or whatever behavioral ideal is set by an individual in a social and historical context), Kant's detour could quite possibly be assisted by cultured technology that (in order to work) exhibits and demands a certain behavior that is enforced by constraining all other options. Ultimately, this means assisting the pursuit of a higher goal by assistance-denial on a lower level. This kind of crew ensures the survival of its captain by refusing to obey orders while near the Sirens. It is the conditional time-gap equivalent of the paradoxical command: 'Do not follow this order!' Contracting with artificial assistants could get us into the position of a technosphere Ulysses with a crew of sassy assistants who assist – *polymechanos*-style – by denying assistance. Possible goals range from acquiring and surviving mantic truth to mastering passions through modesty with polite self-constraint. On an education level, this assistance would not mean delegating the performance of a task, but rather supporting the supervision and regulation of de- and upskilling processes for desired abilities such as navigating by map, organizing a wedding, or behaving in a socially appropriate, personally recognized way.

What ability to constrain? Level-selective assistance.

Socially intervening technology can play a role in technologically mediated de- and upskilling processes. Design, development, and implementation of and relating to such technology can deliberately train people in desired and "cure" them of undesired behavior – which poses the question of which behavior is which. To help designers and engineers create technology that allows for stable relations with the desired behavior, desired behavior must be distinguished from undesired behavior. If pleasant, somehow cultivated social interactions are commonly appreciated, then technology that successfully simulates and demands cultured (polite) behavior could help us to enact corresponding (at first non-genuine) behavior until it leads to genuinely virtuous dispositions – regardless of the self-deception and 'As if' difference – that in turn give rise to behavior that is authentically the way it was previously staged yet enforced by an artificial Ulysses crew.

To navigate this question away from the cliffs of relative subjective desires, we can steer it towards collective conceptions of social appropriateness. Bearing in mind the above-mentioned connection of ends, goals, and wishes, it is safe to say that what

an individual or society deems desirable depends in one way or another – ex negativo or ex positivo – on what is customarily judged as socially appropriate or not. Even if, for example, there is not one specific number of centimeters that separate an appropriate talking distance from an inappropriate one, it cannot be denied that physical proximity is, all over the world, one factor in whether an interacting entity is judged more or less appropriate. Which specific distance will be judged as appropriate in a specific situation between specific agents with a specific relation performing a specific task etc. cannot be generalized: as the five aforementioned dimensions of social appropriateness show, it depends on the culture, customary standards or ethos, or as Hegel terms it, the objective spirit (*objektiver Geist*). The objective spirit refers to all those phenomena that are man-made but cannot be altered by the individual, i.e. that the individual encounters as given. Other prominent phenomena that fit this definition of the objective spirit are history, custom, state, law, art, religion, science, economy, and language – the latter being "the finite, historical heir to the objective spirit" (Ricœur, 1996, p. 65) – in another word: *culture*. Hegel defines the objective spirit as "a form of reality as a world that was created and has to be created by man, in which freedom is a present necessity." [my translation] (Hegel, 1986, p. 32) In modal terms, the phenomena of the objective spirit fall into both the modal spheres of the possible and the necessary depending on the collective or individual level and depending on what are usually large timescales. They are part of the possible (accidence) sphere because they are man-made; they could be otherwise; they could be changed. Yet they are not at the disposal of an individual changing will.

The set of customary practices, the *ethos*, is by definition part of the objective spirit. Those customary standards are but the sum of individual behavior and actions of certain groups, collectives, or societies.

> [T]he objectifications of life tend to deposit and sediment themselves in a durable acquisition which assumes all the appearances of the Hegelian objective spirit. If I can understand vanished worlds, it is because each society has created its own medium of understanding by creating the social and cultural worlds in which it understands itself. (Ricœur, 2016, p. 12)

The objective spirit could be seen as something like a hill that was not created geologically but by the remants of and material left by a past settlement. Over time, the material accumulates in layers and future generations can either walk on the top layer or dig into the sediments for an archaeological investigation. They can even come to the erroneous belief that the hill's origin was geological (that is mistaking the objective spirit as nature), but they can neither ignore nor undo a deeper layer. Even though each layer consists of single individual contributions, none of them are made by a sole individual. Behavior creates deposits in the form of traditions if it is sufficiently collectively manifested and repeated. The sediments are then the transcendental condition affecting any further behavior. Which behavior becomes a customary standard and which collective standards eventually shape individual behavior depends on exposure.

Socially intervening systems will – and technology generally, yet less specifically, always already does – create enormous exposure. The more we act and interact with

technology and within technologically transformed conditions, the more technology becomes a major factor in shaping our behavior. Apart from powerful radical attempts – like propaganda, dictatorship, or religion – it is almost impossible to explicitly design the *ethos*, our customs, and thus to shape what counts as appropriate behavior, because the sedimentation of the objective spirit is a matter of multitudes, if not of humanity. Deliberate attempts to change it either tend to average out or can lead to customs or sedimentary layers that are partly caused by the intervention but deviate significantly from the intention. Even education, one aforementioned attempt to deliberately change every individual's knowledge in a society, despite its relatively consolidated curricula, creates its exposure by way of thousands of individuals (teachers), thus including their individual differences, and so on. Education, a prerequisite for being cultured, makes a huge difference in the objective spirit, yet even the entire educational system cannot purposefully shape a desired form; it rather transforms it somehow into some vague direction. You cannot start a new custom, you can just set a new rule. Whether the latter leads to the former is not up to your will.

Deliberately shaping the objective spirit?

The world economy, capitalism, and data and surveillance capitalism (Zuboff, 2019) tend to favor winner-takes-all “star-system” (Lanier, 2013, pp. 41–47) dynamics. One of the essential features that enable the technosphere and its services and business models is the interconnectedness of technology. This interconnectedness needs rigid norms and communication standards. Relating to the technosphere, relating to interconnected services implies relating to those norms and standards that are and have to be unified all over the world. If the other side of a relation (here: technology) is rigidly standardized, then the entire relationship becomes more standardized, regardless of the differences of ‘one side’. Billions of Facebook users use standardized profile page options to present themselves; millions of Twitter users submit to the standard of 140 or 280 characters; the standard interfaces of the handheld smartphones are palm-sized screens that standardize the human-device relations and bow more than three billions necks every day (GSMA Intelligence, 2019).

Today's billions of smartphones – and with them the potential artificial assistants – and millions of assistance devices such as the Echo are designed and developed by a few thousand people and even fewer ultimately decide their final form. Yet those decisions pre-structure a rather narrow set of potentially stable relations the users could possibly establish. With more than 5 billion mobile phone users worldwide in 2019 (GSMA Intelligence, 2019), customary practices of socializing, dating, making an appointment, etc. drastically change. Globally converging effects due to exposure to standardized technology will probably increase with learning technology or artificial intelligence because one service (e.g. Alexa) with a plethora of devices (e.g. Echo) can learn and integrate the informarion from all interaction entities from all over the world (Brown University, 2016). Interactive technology, communicative devices, and in particular cultured artificial agents create exposure to deliberately shape ‘the entities not at the disposal of just a few’; the comparatively few decision-makers in today's techno-

sphere create standards and new rules of social interaction that, as a result of the tremendous exposure of the products, accelerate the habituation mechanism and thus the sedimentation processes of the objective spirit. Along with this acceleration, well-established mechanisms of selection, variation, and reflection are undermined. By way of technology and habituation, some ‘Lords of the Siren Servers’ gain an as yet indirect but nonetheless much shorter route to deliberately influencing entities of the objective spirit and thus to shape customary standards almost directly to their liking (for now, primarily standards of consumption). This position could be compared – yet in an almost romantic, subcomplex way – to medieval, Renaissance or absolutist monarchs: whatever the King did and how he did it – although himself subject to highly rigid customs and protocol – was adopted by customary multipliers like courtiers and set as the new customary standard. However, courtiers and eventually the rest of society (who could afford to) adopted the ruler's standards knowingly as ‘the ruler's standards’, whereas today we might very much object to granting the most influential ‘Lords of the Siren Servers’ such as Steve Jobs, Jack Ma, Larry Page, Marc Zuckerberg, or Jeff Bezoz the status of a ruler, especially not in terms of culture or aesthetics. Yet in a way they are much more powerful than a medieval King, because they are less bound by customary standards, and their influence is global.

Ethical considerations

The idea presented has significant normative implications. The sequence of steps can be grasped as follows: design technology – interact with that technology – interaction shapes behavior – exposure habituates behavior – change of the structure of the objective spirit. Of course, designing technology, in turn, is influenced by the transcendental condition of the objective spirit as well – there is a protocol for the King. There are two fundamental perspectives here. One focuses on the empowerment of society to influence the transcendental condition of their lifeforms, the objective spirit, in a more deliberate way. This would presuppose that the power of the Lord of the Siren Servers could be controlled democratically and an open normative debate about tolerated, accepted, respected, or desired lifeforms precedes and informs the design and development of cultured artificial agents. A second perspective focuses on the potentially threatening implications and vulnerability of society to the will, biases, and naiveties of a few tech giants that cast customary superstrata around the world.

Both perspectives reveal a series of philosophical considerations that must accompany, complement and precede the development of socio-sensitive cultured technology and their pervasion of everyday life. If we are to preempt possible customary tech-giant despotism, the normative assumptions that influence the selection of one behavior over another, assumptions that are then implemented in the interactive systems and standardized, must be explicated, scrutinized and democratically sanctioned. Actual choices must be available on which system to use and there must be an awareness of the fact that this means choosing which transformation of the objective spirit to support and to promote. This still leaves an ethical dilemma surrounding who could be asked to judge the appropriateness of an action – not least given the importance of intercultural differences in such a judgment – that is then globally rolled out. Who could declare a

sufficient consensus on the tolerability, acceptability, respectability, or desirability of certain behavior? If the objective spirit becomes in part deliberately designable, these choices must then be justified to the collective that is impacted by this design. How could minorities be protected or deviant behavior kept possible if they are not or it is not part of the current mainstream of desirability but nonetheless admissable and the hallmark of an open pluralistic society? Who could possibly train or design the interactive systems so that they exemplify desired behavior – and in turn expose us to it and educate us with it – without also teaching the systems all too human flaws, prejudices etc.? AI, algorithms and learning technology in a way mirror the capacities, limits, and mindsets of those who create learning mechanisms and those who provide the training data. But it is more of a distorted mirroring like the one we get by looking into Anish Kapoor's *Cloud Gate* sculpture. Could there possibly be any standard, any single action or behavior that we could sufficiently agree on as welcome in the objective spirit and as a transcendental condition for any further behavior, or is the mere attempt and therefore any global technology always inherently paternalistic? How could knowledge of "everyone knows how they should be taken" or "everyone understands"[21] as part of data literacy or AI literacy possibly be achieved? Such knowledge is a precondition for the concept of beneficial deception as opposed to harmful deception. Therefore, if the answer to this last question were that such knowledge was impossible, then the beneficial potential of artificially assisted self-deception would be void. Subsequent questions would be how to limit the use of potentially harmfully deceiving technology or whether to ban it entirely.

Education on how systemic interventions, deceptions or simulations from socially interactive cultured technology 'should be taken' would be a good start to mitigate the harmful potential of systemic deception. How deliberately to design the objective spirit – if possible via the detour of cultured technology – how to justify those design decisions vis-à-vis all affected persons, how to balance global technology against culture-specific customary standards, and – when considering all these implications – whether it is better not to shake our robots' hands and say pretty please to Alexa at all: all these are major questions facing a society that is researching, developing, and releasing social robots, companion technology, and artificial assistive agents, and that is enabling them to simulate cultured behavior and to dissimulate the Lords they really serve.

[21] Still referring to Kant: "But these demonstrations of politeness do not deceive because everyone knows how they should be taken." Kant (1996, 38–39 [152])

References

Bratman, M. (2018). *Planning, time, and self governance: Essays in practical rationality*. New York: Oxford University Press.

Brown University (2016). Million Object Challenge. Retrieved from http://h2r.cs.brown.edu/million-object-challenge/

Buttkewitz, U. (2002). *Das Problem der Simulation am Beispiel der Bekenntnisse des Hochstaplers Felix Krull und der Tagebücher Thomas Manns* (Dissertation). Retrieved from http://www.thomasmann.de/sixcms/media.php/471/Diss.%20Das%20Problem%20der%20Simulation.pdf

Dilthey, W. (1965). Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften. In *Gesammelte Schriften* (VII). Stuttgart: Teubner.

Gadamer, H.-G. (2011). *Truth and method* (2., rev. ed., reprint). *Continuum impacts*. London: Continuum.

Gebhart, A. (2018, May 8). Google Assistant's Pretty Please helps your kids mind their manners. Retrieved from https://www.cnet.com/news/googles-pretty-please-feature-wants-to-help-you-enforce-manners/

Gordon, K. (2018). Alexa and the Age of Casual Rudeness. Retrieved from https://www.theatlantic.com/family/archive/2018/04/alexa-manners-smart-speakers-command/558653/

Gransche, B. (2017). The Art of Staging Simulations: Mise-en-scène, Social Impact, and Simulation Literacy. In M. M. Resch, A. Kaminski, & P. Gehring (Eds.), *The Science and Art of Simulation I: Exploring - Understanding - Knowing* (pp. 33–50). Cham: Springer International Publishing.

Gransche, B. (2018). Assisting Ourselves to Death: A Philosophical Reflection on Lifting a Finger with Advanced Assistive Systems. In A. Fritzsche & S. Oks (Eds.), *Philosophy of engineering and technology. The Future of Engineering: Philosophical Foundations, Ethical Problems and Application Cases* (1st ed., pp. 271–289). Cham: Springer International Publishing.

Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Boston: Houghton Mifflin Harcourt.

GSMA Intelligence (2019). The Mobile Economy 2019. Retrieved from https://www.gsmaintelligence.com/research/?file=b9a6e6202ee1d5f787cfebb95d3639c5&download

Hedayati, H., Szafir, D., & Andrist, S. (2019, March - 2019, March). Recognizing F-Formations in the Open World. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 558–559). IEEE. https://doi.org/10.1109/HRI.2019.8673233

Hegel, G. W. F. (1986). *Enzyklopädie der philosophischen Wissenschaften im Grundrisse: 1830. Dritter Teil. Die Philosophie des Geistes. Mit den mündlichen Zusätzen* (9. Auflage). *Werke in 20 Bänden: Vol. 10*. Frankfurt am Main: Suhrkamp.

Ishowo-Oloko, F., Bonnefon, J.-F., Soroye, Z., Crandall, J., Rahwan, I., & Rahwan, T. (2019). Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation. *Nature Machine Intelligence*, *1*(11), 517–521. https://doi.org/10.1038/s42256-019-0113-5

Kant, I. (1797a). *Die Metaphysik der Sitten*. *AA: VI*. Retrieved from https://korpora.zim.uni-duisburg-essen.de/Kant/aa06/429.html

Kant, I. (1797b). *Über ein vermeintes Recht aus Menschenliebe zu lügen. AA: VIII*. Retrieved from https://korpora.zim.uni-duisburg-essen.de/kant/aa08/425.html

Kant, I. (1996). *Anthropology from a pragmatic point of view* (3 opl). London: Southern Illinois University Press.

Kaplan, D. M. (2011). Thing Hermeneutics. In F. J. Mootz & G. H. Taylor (Eds.), *Continuum studies in continental philosophy. Gadamer and Ricoeur: Critical horizons for contemporary hermeneutics* (226-240). London, New York: Continuum.

Karafyllis, N. C. (Ed.). (2003). *Biofakte: Versuch über den Menschen zwischen Artefakt und Lebewesen*. Paderborn: Mentis.

Lanier, J. (2013). *Who owns the future?* London: Allen Lane.

Locher, M. A., & Watts, R. J. (2005). Politeness Theory and Relational Work. *Journal of Politeness Research. Language, Behaviour, Culture*, *1*(1), 9–33. https://doi.org/10.1515/jplr.2005.1.1.9

Lundahl, A., Helgesson, G., & Juth, N. (2017). Ulysses contracts regarding compulsory care for patients with borderline personality syndrome. *Clinical Ethics*, *12*(2), 82–85. https://doi.org/10.1177/1477750916682623

Mori, M. (2012). The Uncanny Valley: First English translation of the original essay from 1970 authorized by Mori. *IEEE Spectrum*. Retrieved from https://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley

Peterson, J. B. (2018). *12 rules for life: An antidote to chaos.*

Pfaller, R. (2018). *Wofür es sich zu leben lohnt: Elemente materialistischer Philosophie* (7. Auflage). *Fischer: Vol. 18903*. Frankfurt am Main: Fischer-Taschenbuch-Verl.

Raman, C., & Hung, H. (2019, July 19). *Towards automatic estimation of conversation floors within F-formations.*

Ricœur, P. (1996). *The hermeneutics of action*. *Philosophy & social criticism*. London, Thousand Oaks, Calif: SAGE Publications. Retrieved from http://search.ebsco-host.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=53288

Ricœur, P. (2016). *Hermeneutics and the human sciences: Essays on language, action and interpretation* (Cambridge philosophy classics edition). *Cambridge philosophy classics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316534984

Rodrigo, P. (2011). The Dynamic of Hexis in Aristotle's Philosophy. *Journal of the British Society for Phenomenology*, *42*(1), 6–17. https://doi.org/10.1080/00071773.2011.11006728

Spellecy, R. (2003). Reviving Ulysses Contracts. *Kennedy Institute of Ethics Journal*, *13*(4), 373–392. https://doi.org/10.1353/ken.2004.0010

StickK (2019). Self-improvement. Powered by behavioral science. Retrieved from https://www.stickk.com/

Ubel, P. (2014, December 11). The Ulysses Strategy. Retrieved from https://www.newyorker.com/business/currency/ulysses-strategy-self-control

Vaihinger, H. (1935). *The philosophy of 'as if": A system of the theoretical, practical and religious fictions of mankind* (Nachdr. der Ausg. New York : Harcourt, 1925). *International library of psychology, philosophy, and scientific method*. London: Kegan Paul, Trench, Trubner & Co., LTD.

Van Quaquebeke, N., & Eckloff, T. (2010). Defining Respectful Leadership: What It Is, How It Can Be Measured, and Another Glimpse at What It Is Related to. *JOURNAL OF BUSINESS ETHICS*, *91*(3), 343–358. https://doi.org/10.1007/s10551-009-0087-z

Weiser, M. (1991). The computer for the 21st century. *Scientific American*, 94–104.

Welch, C. (2018). Google just gave a stunning demo of Assistant making an actual phone call. Retrieved from https://www.theverge.com/2018/5/8/17332070/google-assistant-makes-phone-call-demo-duplex-io-2018

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for the human future at the new frontier of power*. London: Profile Books.