Science and Engineering Ethics (2020) 26:2497–2510
https://doi.org/10.1007/s11948-020-00245-x

ORIGINAL RESEARCH/SCHOLARSHIP

Dignity and Dissent in Humans and Non-humans

Andreas Matthias[1]

Published online: 29 June 2020

Abstract

Is there a difference between human beings and those based on artificial intelligence (AI) that would affect their ability to be subjects of (human-like) dignity? This paper first examines the philosophical notion of (human) dignity as Immanuel Kant derives it from the moral autonomy of the individual. It then asks whether animals and AI systems can claim Kantian dignity or whether there is a sharp divide between human beings, animals and AI systems regarding their ability to be subjects of dignity. How this question is answered depends crucially on one's understanding of what constitutes human dignity and autonomy, and what requirements one places upon systems in order for them to be seen as morally autonomous.

Keywords Human dignity · Kant · Autonomy · Robots · Cyborgs · Extended mind

Introduction

The primary question here is whether autonomously acting, decision-making artefacts, like robots and computer programs, can possess some property analogous in function to human dignity. Is the dignity of human beings something that is inextricably linked to their biological humanity, or is it a property of potentially any sufficiently complex, intelligent and autonomous system, of which human beings just happen to be the historically first instance to appear on the planet? The following sections first describe the most prominent concept of human dignity: Kantian dignity, and contrast it with Philipp Balzer and colleagues' (Balzer, Rippe and Schaber 2000) proposal for a concept of dignity that could apply to animals. Then Kant's understanding of moral autonomy is applied to robots and other AI systems, considering, particularly, the difficulties that arise from the distributed character of modern AI systems.

✉ Andreas Matthias
matthias@ln.edu.hk

[1] Department of Philosophy, Lingnan University, 8 Castle Peak Road, Tuen Mun, Hong Kong

What is Human Dignity?

The Concept of Dignity

The concept of human dignity dates back to antiquity in the form of the Roman *dignitas,* a kind of respect awarded to particular social roles, and can be traced through Christian variations right to present day (Rosen 2012). This paper is not concerned with these historical conceptions of dignity. Although they still exert influence on our everyday understanding of the term, the philosophically most important and influential conception of human dignity goes back to Immanuel Kant, and it is this conception that is examined here.

There are a multitude of alternative conceptions of human dignity (Ashcroft 2005; Mattson and Clark 2011), which are not discussed in detail here. One could distinguish dignity as capabilities or functional characteristics (based on Sen and Nussbaum, see Nussbaum 2001), virtuous behaviour, or a particular kind of rank (Waldron 2007, 2014). One particular approach that might be fruitful for the present question is Philipp Balzer and colleagues' (Balzer, Rippe and Schaber 2000) analysis of the dignity of non-humans, which is examined in a later section. One could see human dignity as the basis of human rights (as, for example, the German Constitution does), or one could dispute that human rights need to be justified through human dignity (Schroeder 2012).

Finally, one might mention that not all philosophers agree that human dignity as a concept in philosophy makes sense (Pinker 2008). Ruth Macklin (2003) famously published an editorial titled "Dignity is a useless concept." She asserts: "It means no more than respect for persons or their autonomy. … Appeals to dignity are either vague restatements of other, more precise notions or mere slogans that add nothing to the understanding of the topic" (Macklin 2003). Others consider dignity to be at least a redundant concept that can be replaced by more precise notions (for example, autonomy, or the capacity to assert claims) (Feinberg 1966; Griffin 2008).

This paper subscribes to a Kantian conception of human dignity and attempts to see how such a conception can be used to argue for or against the dignity of artefacts.

As a side note about the language used, the terms "machines," "automata," "artefacts," and "robots," will all be used interchangeably, although there are important differences between the terms. An "artefact" is something made artificially by human beings, that is, utilising a process that is not found in nature. An "automaton" is an artefact that is able to act autonomously, that is, can act without human supervision. The term "machine" is historically more complex and can either mean a particular kind of an abstract, deterministic transformation in a state machine (going back to Ashby's Cybernetics and constructs like the Turing "machine"). A "robot," finally, is generally understood to be a contemporary, physical automaton, a computerised, electronic device that is able to autonomously move in space and interact in an unsupervised way with the physical world, typically within the everyday living environments of humans. What is

often labelled "robot ethics" should therefore, correctly, be called "autonomous machine ethics," since for the ethical dimension of an action, it does not matter whether the action is executed in a physical environment or not (a desktop computer, for example, might be able to act autonomously by granting or refusing a loan in a banking scenario, or by recommending sentences for offenders to a judge, although it is not mobile, and thus not a robot). The ethical problems involved do not, generally, require locomotion or physical action, and so "robot ethics" is a somewhat limiting and misleading term. But since its use is customary, the term is used here.

Kantian Dignity

Arguably, the most influential source for our modern conception of human dignity is Immanuel Kant's *Groundwork of the Metaphysics of Morals* (Kant 1900). Kant speaks of 'Würde' (related to Wert=worth), which is the same word used in the German Basic Law for the concept of human dignity ('Menschenwürde,' worth or dignity of man).

The crucial passage is:

> In the kingdom of ends everything has either a *price* or a *dignity.* What has a price can be replaced by something else as its *equivalent*; what on the other hand is raised above all price and therefore admits of no equivalent has a dignity … That which constitutes the condition under which alone something can be an end in itself has not merely a relative worth, that is, a price, but an inner worth, that is, dignity. Now, morality is the condition under which alone a rational being can be an end in itself, since only through this is it possible to be a lawgiving member in the kingdom of ends. Hence morality, and humanity insofar as it is capable of morality, is that which alone has dignity (Kant 1900, pp. 4:434–435; cited after Gregor 1996).

The last sentence is the core of this thought. Only humanity has dignity, *insofar as it is capable of morality.* Kant adds: "Autonomy is therefore the ground of the dignity of human nature" (4:436). As discussed later, this is still, after more than 230 years, the most precise attempt to ground human dignity in a particular ability of the human mind.

Kant's influence reaches all the way to the present. The interpretation of human dignity provided by the German Constitutional Court (Bundesverfassungsgericht) is, in its main lines, based on Kant (Rosen 2012). Similarly, although Roger Brownsword frames his concept of dignity in terms of 'empowerment' of the individual to make a choice (or the constraints imposed on the choices of others that affect the individual), this empowerment itself is described in Kantian terms:

> Human dignity as empowerment is committed to a framework of action in which humans may choose to do the right thing as they may choose to do the wrong thing. To take away from humans their capacity to make wrong choices is an insult to their capacity for choice, the worst kind of affront to their dignity. (Brownsword 2004, p. 213).

It is also important to note that Kant contrasts dignity with "price" rather than, say, worthlessness. Things that lack dignity still may have a value, but their value can be expressed as a price. A price signifies that something is fungible, that it can be replaced by another thing of the same type. Individuals, even artefacts insofar as they are unique, are often said to be "priceless": their value is enhanced because of their uniqueness and irreplaceability. Irreplaceability alone would not be sufficient for human dignity, though, since dignity, for Kant, is clearly rooted in human *autonomy,* not in mere uniqueness.

Autonomy and Determinism

Now, one could dispute whether human beings actually have this kind of moral autonomy themselves. Are we not, like pre-programmed machines, driven by the requirements of our genes and the structure of our neural networks to make particular choices? Do we really have that kind of free moral autonomy that Kant seems to advocate? This paper does not dive deeper into the question of free will and determinism, but since the focus of the present paper is to ask whether machines can be subjects of human-like dignity, the discussion about whether *men* are essentially free or determined in their actions is beside the point. If one subscribes to a Kantian notion of dignity, and assumes that human autonomy (such as it is) sufficiently grounds human dignity, then it can be argued that *if* machines could act in a similarly autonomous way as humans, they would also qualify to be subjects of dignity, just as humans are.

Kant himself sidesteps the problem by, on the one side, admitting that human beings, as physical and biological entities, can be described as following causal laws in their behavior; but, on the other hand, as moral beings, they are capable of self-determination ("transcendental" vs. "practical" freedom: Reath 2006, p. 277).

Dissent

For Roger Brownsword, as noted above, there is a value in the ability of human beings to freely make choices, even if these choices are bad (either disadvantageous to the individual, or morally wrong) (Düwell et al. 2014; Brownsword 2004, p. 213). The same does not seem to be true of Kant. According to what has been called the "enactment" view (Dean 2006), human beings must actually follow the Kantian 'moral law' in order to possess dignity, and may lose their dignity when they behave immorally (Bayefsky 2013, p. 821). It seems that for Kant, some actions lead to the loss of the human dignity of the agent. Voluntarily giving up one's autonomy, complaining and whining, lying, drinking and committing "unnatural sins" annihilate the dignity of a human being and degrade the person to a level below that of animals (Bayefsky 2013, p. 817). This might be surprising, given that Kant is often seen as a defender of the unconditional ascription of inviolable dignity to human beings. Kant summarises his demands on moral human action in a series of formulations of the so-called Categorical Imperative. The two most prominent forms of the Categorical Imperative stipulate that

a moral action must be *generalisable* (being able to be made into a universal law that all agents can follow) and that a moral action must not treat human beings merely as means to some end, but always also as ends (Kant 1900), thus recognising the infinite worth (dignity) of human beings.

For human dignity, therefore, it is not sufficient that the agent is rational and autonomous, but it is also required that the agent's moral choices are *of a particular kind,* namely such that they are in agreement with these stipulations of the Categorical Imperative, i.e. that they are rational and generalisable, and that they treat human beings as ends and not merely as means. If even a simple lie "annihilates" the dignity of the human being, as Kant says in the passage cited by Bayefsky above, then probably every moral transgression will do the same. Thus, only agents that consistently act morally right could claim human dignity for themselves.

On the other hand, mere, passive or enforced adherence to the moral law is not sufficient to provide grounds for human dignity. Kant clearly states that *moral autonomy* is required. Kantian autonomy can be understood in various ways, either as "the ability to give oneself the moral law through reason," or "the property of actually acting on [self-given] principles" (Bayefsky 2013, p. 821). One would be what is usually called "moral autonomy," while the other would be a metaphysical kind of autonomy or self-determination. In both cases, the aspect of the individual being morally self-legislating is crucial and is the source of the "respect" that should be shown to human beings.

For autonomy (and dignity), it is therefore necessary that the agent is able to effectively act on self-given principles and to give themselves the Kantian moral law through their own reason. Just following the commands of another does not constitute autonomous action. The ability of the individual to voice *dissent* is therefore a marker of the individual's ability to be self-governing, autonomous, and the subject of human dignity.

This is, among other issues, also an epistemic point, and one that is crucial for the classification of machines as moral agents. When an agent agrees with and follows the moral law (understood as the rules that are compatible with the Categorical Imperative) in the form of morally prescribed and accepted behavior, we are unable, from the outside, to judge whether the agent is a morally autonomous one, since we cannot know whether the agent's obedience is a direct consequence of their own, morally autonomous decision, or whether it has been enforced by external means. A remote-controlled robot, for instance, *might appear* to be autonomously following Kant's moral law (understood as action that conforms to the stipulations of the Categorical Imperative), but the reality of the matter is that it has no choice but to act as the remote controller prescribes. The same is true if the moral deliberation happens inside the machine but following an algorithm that prescribes particular moral choices and renders others unavailable to the agent.

It is only in cases of *dissent* that we can unambiguously witness moral autonomy in action. "Dissent" here shall just be understood as a demonstrated ability to refuse to act in a particular way that is prescribed by external forces, based on one's own rational deliberation and values. In this way, dissent as the consequence of a rational deliberative process differs from a mere inability to follow orders due to, for example, a human being unconscious, or a machine damaged.

One may think of a car rolling down a hill. If one wants to be assured that the steering wheel works and that the driver indeed has control of the vehicle, it wouldn't be sufficient to just let the car roll straight down the hill. This might be what the car's driver would like to do anyway, but letting the car roll on its default course does not demonstrate the driver's control over it. Only when the driver attempts to turn the wheel (even if this might be suboptimal in itself, perhaps causing the car to veer dangerously from side to side), she can be assured that the car indeed follows the instructions given through the steering wheel and that the driver herself is in effective control of the rolling car. Similarly, if an agent *refuses* to act according to some prescribed standard of action (the refusal being not just casual inaction but a reasoned response), then this refusal demonstrates the agent's capacity for genuine moral reasoning and freedom. It seems, therefore, that we can use the possibility of reasoned dissent as a diagnostic tool to evaluate the ability of an agent to be a Kantian, that is, autonomous *moral* agent.

The Dignity of Non-humans

The Dignity of Animals

A brief look at the possibilities of dignity for animals can perhaps help clarify the requirements for dignity in non-humans more generally. First, a competing concept, a functional, quasi-Aristotelian account of animal dignity proposed by Balzer et al., and why it does not seem to be convincing as an attempt to justify the dignity claims of nonhumans is discussed. Then what a Kantian approach to animal dignity would look like is considered.

Balzer's 'Teleological' Account

The dignity of animals has been the subject of extensive research (e.g. Meyer 2001; Bekoff 2004; Bilchitz 2009; Schindler 2013; Zuolo 2016) as well as political action. In a referendum on May 17, 1992, the citizens of Switzerland ordered their government to issue regulations on the use of the genetic material of animals, plants and other organisms by taking into account the dignity of non-human organisms (Balzer et al. 2000, p. 7).

Balzer assumes that for human beings the only plausible (in his view) interpretation of human dignity would be "the moral right not to be degraded" (p. 12), where "degradation" is described in terms of self-respect. If this is the case, then, obviously, human dignity cannot be applied to animals or plants, since these life-forms lack the ability to have self-respect because they lack the necessary neural infrastructure and mental states. Of course, not all humans have the full range of higher mental states either. Young children, for example, would not have a fully developed sense of self-respect. Still, Balzer maintains, we ascribe human dignity to them not as individuals but as members of the species out of social and psychological considerations (as opposed to these groups of humans who have a genuine claim to

dignity). Balzer et al. maintain that there might be good (though indirect) reasons to "grant the moral right not to be degraded even to those who cannot be degraded" (Balzer et al. 2000, p. 14).

Moving away from humans (and, possibly, higher apes and dolphins) Balzer asks what could possibly be the grounds for ascribing dignity to other animals and plants? He then proposes a broadly Aristotelian concept, in which animals "have their own good" (p. 15), pursue individual goals, and can be described as organic units. The challenge here is, of course, to avoid over-extending the notion of dignity, for example to insects or plants, which would water it down so much as to become meaningless as a basis for claiming rights. Balzer argues that illness, wilting and rotting are ways in which animals and plants can fail to achieve their own good (to 'flourish' in Aristotelian terms), while the same does not apply to stones. A broken stone is no less a good stone than an unbroken stone, while a dead animal is less of an animal than a healthy animal.

The point of the creature having individual goals is meant to exclude machines: "Of course, one could also ascribe such an own good to a machine, as it may rust, and fall into such a bad condition that it is unable to perform its function properly" (Balzer et al. 2000, p. 16). But machines, according to Balzer et al., don't have their own ends: "they exist only to fulfill certain purposes for which human beings have designed them" (p. 16), and so they don't qualify for dignity. Although Balzer here is trying hard to make sense of the concept of non-human dignity, multiple problems seem to present themselves.

First, about human dignity itself, it does not seem like 'degradation' is a reliable detector for violations of human dignity. Degradation of one's humanity is a concept that is highly variable over time. As living conditions in a society change, what is considered as degrading one's dignity also changes and the standard of comparison can slide up or (more commonly) down. In today's overcrowded cities, it is not uncommon to see human beings sharing tiny living spaces (7 sqm per person in a recent IKEA ad)—a situation that, in other circumstances, might be considered to degrade one's humanity.

Second, the distinction between humans, higher (possibly self-conscious) animals, lower animals, plants and rocks in terms of dignity seems hard to justify. Especially the mention of larger primates (p. 14) seems to suggest that consciousness and self-awareness come in degrees along a continuum of mental abilities. Despite that, Balzer does not follow up by conceding that dignity can also come in degrees. For him, the right to not be degraded is absolute and even applies where no degradation is possible (for example, because the individual does not sufficiently understand the conditions of his own degradation, or because an animal is not even sufficiently developed to have self-reflective mental states.) Along the idea of degradation, Balzer then introduces the concept of things having "their own good" (p. 15) as another ground for dignity, and the relationship between the two concepts, degradation and having an Aristotelian 'telos' is never clearly stated. Is one of the two already sufficient for something to have dignity? Do both conditions need to be fulfilled? Is one more important than the other? And if one applies to humans and higher animals, and the other to plants and rocks, then what is the underlying common principle that would allow us to call both demands for the fulfillment of these

two different conditions "dignity"? Don't we obscure rather than clarify things when we use the same term for two entirely different and unrelated concepts?

Third, further down the chain, it is also unclear why inanimate objects cannot participate in the same notion of non-human dignity as animals or plants. A lake, for instance, can clearly be a good, well-functioning, or a degraded, poisoned lake. The same can be said of ecosystems, societies, and other entities that are not animals, but that seems to allow for a sensible notion of "well-being as the thing they are." The Aristotelian *telos* extends to all things, and Balzer does not clarify why his notion of dignity should not be extended to a lake when it can apply to a tree.

Fourth, it is not entirely clear what the idea behind machines "not having their own ends" is. Yes, machines are designed to further particular human ends, but one could say the same, of, for instance, transgenic animals. Goats that produce spider silk in their milk exist only to further the ends of human beings—are they therefore to be devoid of dignity? And what of autonomous robots that can operate without supervision in a human social environment? Do they not also pursue their own goals? Not all machines are pre-programmed to fulfil a particular task. Particularly deep and reinforcement learning, coupled with autonomous, unsupervised operation, can enable an artefact (for example a chatbot or an AI personal assistant) to take up many different roles and to autonomously pursue its own goals in a shared environment with human beings. Sure, the transgenic goat is alive, while the robot is not, but if that is the relevant difference, then Balzer et al. fail to explain why being alive would be a necessary condition for having dignity.

And last, but most importantly, we can ask whether it is a good idea at all to reuse the term "dignity" for something that is explicitly said to be a different kind of thing, resting on different conditions than what we call "human dignity." The concept of human dignity derives whatever power it has from particular connotations (mostly Kantian or Christian) and it would be very confusing to create a sound-alike, second version of the same term that denotes an entirely different concept and has no relation to either the Kantian or the Christian roots of the original concept. Suitable terms for Balzer's concept already exist. The capabilities approach (Nussbaum 2001; 2008), natural law and deep ecology all are based on similar ideas about the flourishing of a species being something valuable in itself. There is no additional gain to be obtained from attaching the label "dignity" to such a concept.

Individuality and Dignity

The Kantian conception of human dignity suggests another interesting point: that individuality might be a necessary condition for human dignity. If some things have a price and some have dignity, and the latter have dignity precisely because they are not fungible and, therefore, their value is unique and beyond any possible exchange-value (price); then we might conclude that what is fungible must lack that specific value precisely *because* it is interchangeable with other, similar things, rather than being an irreplaceable individual.

As parents, we can understand the argument that our child cannot be replaced by another child of the same age, gender and general features. But a car could well

be replaced by another without a loss of value (say, in the case of a manufacturer recall). This is, following Kant, because a person has dignity where a thing has only a price. And the reason that a person is different is, for Kant, that a person can be autonomous and a creator of their own individual moral law, while a car cannot.

This thought has a problem, though, that has been explored more closely in the philosophy of love, in the form of substitution arguments (Soble 1990; Driver 2014). What happens if the thing that has autonomy is a mass-manufactured object that lacks individuality? Then the properties of autonomy and non-fungibility do not necessarily appear together in the same object anymore. We can imagine a possible, sentient, morally autonomous robot that is mass-produced and exists, as the functionally same individual, in thousands of materially different copies. It seems that the availability of these copies, and the resulting possibility of replacing the particular robot with another, indistinguishable copy, would somehow harm the robot's standing as an individual with a specific and unique moral worth and make that robot into a thing that is fungible, that can be bought and sold, and that does not qualify to be a subject of dignity (since it is already a thing with a price, and for Kant the two categories are not meant to overlap: something has *either* a dignity *or* a price).

A Kantian Account

On a more Kantian account, we would have to ask the questions:

1. Are animals individuals? Do they have sufficient individual differences that give them "worth" rather than a "price"? At least for some animals, this clearly seems to be the case. Pets typically are individuals, carry a name and have an individual history of personal interactions with humans that render them unique and irreplaceable. Wild animals would not qualify. One nameless sheep in a herd of a thousand would not have the same kind of worth. Instead, it would be considered a mere thing, something that can be bought, sold and replaced without loss by another sheep of the same type. Notice that, just as with robots, the point here is the fungibility of the particular sheep, not the species. One could say the same of a wild dog, which is as good as another wild dog, if I don't personally relate to either of them. On the other hand, a pet sheep that carries a name and to which I personally relate through a shared history of interactions, would qualify as a non-fungible thing, and therefore something that is unique and, potentially, beyond "price." Now this condition is alone not sufficient for Kantian dignity, because I also have to consider:
2. Does the animal have a sense of the Kantian 'moral law' and can it direct its decisions toward establishing a moral law for itself and then following it? The moral law, for Kant, has to be rational in the sense that it must fulfil the requirements of the Categorical Imperative. Instinctual action that compels the animal to act, for example by prioritising its own needs over the requirements of reason, would therefore not qualify as proper moral deliberation. In this sense, it is questionable whether any animal can fulfill Kant's requirements.

Autonomy, Individuality, and Dignity

Autonomy, Error, and Dissent

Moving from animals to machines, now one must ask whether computers, robots and other "autonomous" artefacts could possibly qualify as subjects of (human-like) dignity.

First, it is important to notice that Kant would not consider algorithmic morality (that is the subject of much contemporary research (Wallach and Allen 2008; Arkin 2009; Winfield and Jirotka 2018) to be genuine moral autonomy. The self-driving car that executes an algorithm that prioritises one kind of outcome over another does not exercise any moral autonomy. The result of the algorithm's execution is predetermined by the programmer (or, in the case of deep learning, by the patterns that have been learned during training and the current state of the environment). At no point is the program *free* to make a genuine moral choice (Matthias 2011). We can also phrase this in terms of dissent. The self-driving car does not have the ability to *disagree* with the moral judgement that is prescribed by its software on purely moral reasons.

True, the car's software can disagree with the human driver (or the programmer) about how to evaluate a particular situation on the road in terms of the optimal reaction that the car should exhibit. But this is not a form of dissent that is morally relevant in a Kantian sense. It is just a different conclusion reached on the basis of the data that the car, through its sensors, has available and that are different from the information that a human driver can access in the same situation. In this case, the values and moral imperatives that are implemented in the car's software are still identical to those of the human designer or operator. What differs is the car's evaluation of the facts at a particular moment, its perception of how the situation that is perceived through its sensors maps into the conceptual framework provided by its software. Either the car's algorithm or the human operator/designer/programmer must be in error when such a disagreement occurs. One of the two sides either has a limited access to data or is using that data wrongly, thus reaching different results. If we could make sure that both the car's algorithms and the human operator/designer/programmer see the situation in exactly the same way, having access to the same data, and that neither is mistaken as to the facts, then we could expect their suggestions for action to agree with each other.

Genuine moral dissent therefore is more than just disagreement about facts. Genuine dissent persists even in the absence of factual error. For example, two soldiers might have access to exactly the same data about a person on the other side of the battlefield. They might both classify the person as an enemy and a combatant, judge the person's intentions and abilities in the same way, obeying the same laws of war and rules of engagement; but they might still disagree about how to act, depending on their different moral values, their opinions on the rightness of that war's cause, their conceptions of fairness, their religious beliefs and many other evaluative factors. We would only speak of a morally *autonomous*

machine if the war robot, despite all factual agreements, were able to assess the morality of a possible response *differently* from its designer, operator or commander, and, in the consequence of such a differing assessment, refuse to execute a particular command. Of course, not the actual refusal is what counts, but the *ability* to refuse a command on moral grounds.

Individuality and Dignity

The problem is complicated by the realisation that technological systems often are not as "closed" regarding the location of their agency as biological systems are; and they are usually not "individuals," that is, they do have a "price" in the Kantian sense, rather than "worth" (dignity).

We are used to perceiving biological agents as individuals, separated from each other by the borders of their respective bodies. Although this can be argued to be fiction, even in the case of biological organisms (see, for example, the role of mycelia in the communication of trees (Gorzelak et al. 2015) or the influence of gut microbiota on brain and behavior (Cryan and Dinan 2012)), it becomes even more apparent in the case of technological artefacts like computers and robots. A robot's "cognitive" processing abilities depend entirely on sensors and processing equipment that does not need to be located inside the robot and that can both depend on and respond in complex ways to environmental conditions (Bluetooth and Wi-Fi signals, remote sensors, "cloud" computing resources, centralised memory on remote servers, access to remote knowledge bases and expert systems, online image recognition and natural-language processing via, for example, remote IBM Watson or Amazon services). The "individuality" of a modern robot is therefore a difficult notion to make sense of. Each Tesla car profits from the learning and experiences of any other Tesla car, via the sharing capabilities of the built-in car software (Viereckl et al. 2015), essentially rendering any connected car into a "clone" of any other car that utilises the same software. Also, machines in general, by the nature of the industrial production process, lack individuality in their hardware make-up, being constructed from identical, mass-manufactured components; unlike living organisms that contain their highly individual blueprint as DNA molecules within each single cell. For Kant, this already would place the inherent worth of a machine into question, since it is the essential individuality of a thing (or human being) that makes it able to be a carrier of "worth" rather than of a price tag. From a Kantian perspective, the commercial availability of identical machines might itself be a reason to deny them the ability to be subjects of dignity.

Conditions for the Autonomy of an Artefact

In the case of an autonomous robot:

1. One would first have to identify the locus of its (moral) autonomy. Where is the actual moral deliberation taking place? Where is the decision to act taken,

and which hardware module is responsible for that decision? Is that module an individual (for example, a neural network that has been trained in a way that now renders it substantially different from all other modules of that type) or is it just one of many identical modules? In this case, it would have a "price" but not a moral "worth."

2. Then one would have to analyse the external factors that were causal to the particular decision being taken. Did these factors *compel* the robot to take that particular decision, or did the machine have an actual choice? Was a step of free deliberation part of the decision-making process? With Kant, we can disregard the fact that ultimately, in a physical system operating in a causally closed world, the "freedom" of the decision-making process can never be absolute. Of course, one will be able to identify causal factors that led to the particular decision being taken, in the same way as one might be able to name a series of good reasons that compel a human being to take a particular decision. Being able to do that does not mean that the agent was not free, though. It might be helpful, as a heuristic approach, to ask whether the agent did have the freedom to act differently had they been able to dissent. In the case of a robot, there must be at least one point in the decision-making process where the *robot itself,* and not its designer or programmer in advance, autonomously makes a particular decision based on an evaluation of all available facts and utilising a set of values that the robot has itself formed and adopted.
3. One would have to exclude that the moral decision is taken remotely, by a centralised processing unit, and merely transmitted to the particular physical "body" of the robot. In this case, the robot would not be an individual but one of many identical manifestations of the remote brain's body. Not being an individual, the particular robot body could not claim moral "autonomy" in the Kantian sense. Instead, perhaps the remote, controlling unit could, if it itself is a singular, separate individual, be able to take its own decisions.
4. If, on the other hand, the robot is controlled by multiple processing units, then again one could not consider it *one* individual. Each processing unit alone would not be solely responsible for the robot's actions and could therefore not claim autonomy.

Conclusion: Dignity of Artefacts

Assuming for a moment, accepting the Kantian version of human dignity, what would be our judgement regarding the dignity of intelligent, autonomous artefacts?

First, the justification for human dignity for Kant lies in the ability of *individual* human beings to be creators of the 'moral law' as well as its subjects. It is based on the essential moral autonomy and freedom of the human being. Given this, artefacts could only claim dignity if they can exhibit a comparable degree of moral autonomy to human beings. As is explained briefly above, this requirement is *not* fulfilled by machines that can implement and follow moral rules in an algorithmic way, that is, by following a pre-programmed sequence of steps for conducting moral deliberations. Such algorithmic morality would miss the point of true moral autonomy,

which includes the ability to disobey the algorithm or the values of the machine's creators and to opt for a creative and even a morally 'bad' course of action (Matthias 2011).

For Kant, it seems that moral autonomy is a necessary condition for the ascription of dignity, but not a sufficient one. The agent still has to *behave in a way that is consistent with the requirements of dignity,* e.g. to fulfil the duties that arise out of the Kantian 'moral law' (Bayefsky 2013, p. 821) and it has to be an *individual,* that is, something that cannot be replaced by another unit of the same type. This is what makes it possible for it to be the carrier of "worth" rather than a "price."

A further question would be whether ascribing dignity to morally autonomous artefacts means that they would qualify for (human) rights. Certainly, it seems that human rights can be justified in a number of ways (see, for example, Schroeder 2012; Nussbaum 2001, 2008) and that human rights do not require human dignity as a prerequisite. On the other hand, if one assumes that there is something like human dignity, and that it works like Kant thinks, then it seems natural to derive at least some (human) rights from the presence of human-like dignity in a subject. Kantian human dignity (if accepted at all as a concept that applies to machines) would be sufficient but not necessary for the ascription of some human-like rights to autonomous robots.

References

Arkin, R. (2009). Governing lethal behavior. In *Autonomous robots*. CRC Press.

Ashcroft, R. E. (2005). Making sense of dignity. *Journal of Medical Ethics, 31*(11), 679–682.

Balzer, P., Rippe, K. P., & Schaber, P. (2000). Two concepts of dignity for humans and non-human organisms in the context of genetic engineering. *Journal of Agricultural and Environmental Ethics, 13*(1), 7–27.

Bayefsky, R. (2013). Dignity, honour, and human rights: Kant's perspective. *Political Theory, 41*(6), 809–837.

Bekoff, M. (2004). Wild justice and fair play: Cooperation, forgiveness, and morality in animals. *Biology and Philosophy, 19*(4), 489–520.

Bilchitz, D. (2009). Moving beyond arbitrariness: The legal personhood and dignity of non-human animals. *South African Journal on Human Rights, 25*(1), 38–72.

Brownsword, R. (Ed.) (2004). What the world needs now: Techno-regulation, human rights and human dignity. In *Human rights* (pp. 203). (Global governance and the quest for justice; v. 4). Oxford: Hart.

Cryan, J. F., & Dinan, T. G. (2012). Mind-altering microorganisms: The impact of the gut microbiota on brain and behaviour. *Nature Reviews Neuroscience, 13*(10), 701.

Dean, R. (2006). *The value of humanity in Kant's moral theory*. Oxford: Oxford University Press on Demand.

Driver, J. (2014). Love and duty. *Philosophic Exchange, 44,* 1.

Düwell, M., Braarvig, J., Brownsword, R., & Mieth, D. (2014). *The Cambridge handbook of human dignity: Interdisciplinary perspectives*. Cambridge: Cambridge University Press.

Feinberg, J. (1966). Duties, rights, and claims. *American Philosophical Quarterly, 3,* 8.

Gorzelak, M. A., Asay, A. K., Pickles, B. J., & Simard, S. W. (2015). Inter-plant communication through mycorrhizal networks mediates complex adaptive behaviour in plant communities. *AoB Plants*. https://doi.org/10.1093/aobpla/plv050.

Gregor, Mary J. (Ed.). (1996). *Kant: The metaphysics of morals*. Cambridge: Cambridge University Press.

Griffin, J. (2008). *On human rights*. Oxford: Oxford University Press.

Kant, I. (1900). Groundwork of the Metaphysics of Morals. In: Königliche Preußische Akademie der Wissenschaften (ed.). *Kants gesammelte Schriften*. Berlin: Georg Reimer.

Macklin, R. (2003). Dignity is a useless concept. *BMJ, 327*(7429), 1419–1420. https://doi.org/10.1136/bmj.327.7429.1419.
Matthias, A. (2011). Algorithmic moral control of war robots: Philosophical questions. *Law, Innovation and Technology, 3*(2), 279–301.
Mattson, D. J., & Clark, S. G. (2011). Human dignity in concept and practice. *Policy Sciences, 44*(4), 303–319.
Meyer, M. (2001). The simple dignity of sentient life: speciesism and human dignity. *Journal of Social Philosophy, 32*(2), 115–126.
Nussbaum, M. (2001). *Women and human development: The capabilities approach* (Vol. 3). Cambridge: Cambridge University Press.
Nussbaum, M. (2008). Human dignity and political entitlements. In *Human dignity and bioethics. Essays commissioned by the president's council on bioethics* (pp. 351–380). Washington, DC.
Pinker, S. (2008). *The stupidity of dignity. New Republic, 238,* 28–31.
Reath, A. (2006). Kant's critical account of freedom. *Blackwell Companion to Kant*. https://doi.org/10.1002/9780470996287.ch19.
Rosen, M. (2012). *Dignity: Its history and meaning*. Cambridge: Harvard University Press.
Schindler, S. (2013). The animal's dignity in Swiss animal welfare legislation—Challenges and opportunities. *European Journal of Pharmaceutics and Biopharmaceutics, 84*(2), 251–254.
Schroeder, D. (2012). Human rights and human dignity. *Ethical Theory and Moral Practice, 15*(3), 323–335.
Soble, A. (1990). *The structure of love*. New Haven: Yale University Press.
Viereckl, R., Ahlemann, D., Koster, A., & Jursch, S. (2015). Connected car study 2015: Racing ahead with autonomous cars and digital innovation. Retrieved from March 7, 2016, from Strategy & http://www.Strategyand.Pwc.Com/Reports/Connected-Car-2015-Study.
Waldron, J. (2007). Dignity and rank: Memory of Gregory Vlastos (1907–1991). *European Journal of Sociology/Archives Européennes de Sociologie, 48*(2), 201–237.
Waldron, J. (2014). What do the philosophers have against dignity? NYU School of Law, Public Law Research Paper No. 14-59. Retrieved from http://dx.doi.org/10.2139/ssrn.2497742.
Wallach, W., & Allen, C. (2008). *Moral machines: Teaching robots right from wrong*. Oxford: Oxford University Press.
Winfield, A. F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133), 20180085.
Zuolo, F. (2016). Dignity and animals. Does it make sense to apply the concept of dignity to all sentient beings? *Ethical Theory and Moral Practice, 19*(5), 1117–1130.