philosophies

Article

Responsibility and Robot Ethics: A Critical Overview

Janina Loh

Philosophy of Media and Technology, University of Vienna, 1010 Vienna, Austria; Janina.loh@univie.ac.at

Received: 13 July 2019; Accepted: 23 October 2019; Published: 8 November 2019

Abstract: This paper has three concerns: first, it represents an etymological and genealogical study of the phenomenon of responsibility. Secondly, it gives an overview of the three fields of robot ethics as a philosophical discipline and discusses the fundamental questions that arise within these three fields. Thirdly, it will be explained how in these three fields of robot ethics is spoken about responsibility and how responsibility is attributed in general. As a philosophical paper, it presents a theoretical approach and no practical suggestions are made as to which robots should bear responsibility under which circumstances or how guidelines should be formulated in which a responsible use of robots is outlined.

Keywords: robots; responsibility; moral agency; moral patiency; inclusive approaches

1. Introduction

It is currently assumed that technological developments are radically changing our understanding of the concept of and the possibilities of ascribing responsibility. The assumption of a transformation of responsibility is fed on the one hand by the fundamental upheavals in the nature of 'the' human being, which are attributed to the development of autonomous, self-learning robots. On the other hand, one speaks of radical paradigm shifts and a corresponding transformation of our understanding of responsibility in the organizational forms of our social, political, and economic systems due to the challenges posed by robotization, automation, digitization, and industry 4.0. It is also expressed widely that, thanks to these circumstances, our modern mechanized mass society sets ultimate limits to responsibility, even opening up dangerous gaps in the possibilities of attributing responsibility [1]. Nevertheless, the call for responsibility seems to continue unabatedly. The question is whether, despite all these changes, we can continue to build on the traditional concept of responsibility (Sections 5.1 and 5.2) or whether our traditional understanding of responsibility is actually changing or should change (Section 5.3).

In the following, I will, first, define the traditional understanding of the term "responsibility" via an analysis of its etymology and genealogy and outline what I call a "minimal definition" of the concept of responsibility [2,3]. Second, I will give a short overview over the philosophical discipline of robot ethics (Section 3) and its three fields of research (Section 4), in order to be able to give a theoretical approach of how to ascribe responsibility in man–robot interaction in the fifth part of this paper.

The history of responsibility is comparably short: the adjective "responsible" first appeared in the thirteenth century in France and in the seventeenth century in Germany. A reflective and systematic usage of the term is not found until the beginning of the nineteenth century [4–6]. Responsibility is a tool for organizing and regulating contexts that are not transparent because they involve vast numbers of people and inscrutable hierarchies. Because one can be called responsible for past, present and future actions, the concept of responsibility complements less complex phenomena such as ascriptions of duty or guilt that fall short in situations when courses of action are mediated by a large number of instances and authorities (such as in the industrial revolution [7]). During the twentieth century new potential responsible agents entered the scene: robots.

Robot ethics is a comparatively young philosophical discipline, and frequently faces two accusations. The first is that robot ethics has no specific object of analysis, because ethics does not concern stocks and stones, i.e., objects. The second is that even if it is justifiable to include artificial systems in ethical reflection, they do not raise any questions that have not been asked long before in more traditional ethical arenas. However, in response to the first accusation, robots can occupy a place in the moral universe even if we are not willing to identify them as moral agents, comparable to countless (partly) inanimate entities that to human eyes have a value—landscapes, ecosystems, the planet earth, even houses, cars, and smartphones. Where, if not in ethics, should we discuss the potential value of artificial systems? As for the second accusation, there is not a lot to answer; however, this criticism applies not only to robot ethics, but to any ethics restricted to a specific context (such as animal ethics, climate ethics, and health care ethics), as long as we agree on the human being as origin and pivot of ethical reflection per se. Robot ethics indeed poses traditional philosophical questions and also rebrands particular challenges that confront other ethical systems (such as animal ethics [8–12]). For instance, which competences define agency? What are the prerequisites for moral agency? With what moral values should artificial systems be equipped? What moral self-understanding underlies 'bad' behavior towards robots? In what areas of human specialisms—be it industry, military, medicine, elderly care, service, or others—do we still want to rely (partly or significantly) on human rather than artificial expertise? It is intuitively evident that questions of ascribing, delegating, sharing and dividing responsibility are raised in these spheres.

2. What is Responsibility[1]

Etymologically speaking, the traditional understanding and the usage of the term "responsibility" includes three components [2] (pp. 3–41). Being responsible means, firstly, the state or fact of being answerable for something. It is the ability to provide an accounting of one's actions [14,15]. Secondly, responsibility is a normative concept, that is, it is not only descriptive and causal. In calling the sun responsible for melting the candle wax we use the ascription of being responsible in a metaphorical sense, since the sun is not able to explain itself or be answerable. In contrast, in calling someone responsible for killing another person we usually do not want to state a simple fact or see the person in question as a cause in a purely descriptive way. We want the murderer to explain herself or himself and to accept her or his guilt. Thirdly, responsibility includes a specific psycho-motivational constitution of the responsible subject: we think she or he is answerable in the sense of being an autonomous person, able to take on her or his responsibility, and equipped with capabilities such as judgement and the faculty of reflection [2] (pp. 39–41) [16].

This etymologically minimal definition of responsibility leads to five relational elements that I discuss in more detail in the following paragraphs. An individual or collective subject or bearer of responsibility is the responsible agent or person (who is responsible?). The subject is responsible for a past or future object or matter (what is x responsible for?). The subject is responsible to a private or official authority (to whom is x responsible?) and toward a private or official addressee or receiver; the addressee defines the reason for speaking of responsibility in the context in question. Finally, private or official normative criteria define the conditions under which x is responsible; they restrict the area of responsible acting and by this differentiate moral, political, legal, economic and other responsibilities—or better: domains of responsibility. For instance, a thief (the individual subject) is responsible for a stolen book (the retrospective object), or, better, the theft (a sequence of actions that have already occurred) to the judge (the official authority) and towards the owner of the book (the official addressee) under the conditions of the criminal code (the normative criteria that define a legal or criminal responsibility).

[1] Some of the Thoughts Expressed in the Introduction to Section 2 Have already Been Presented in a similar Way in [13].

The first to explicitly address the relationality of the concept of responsibility was Alfred Schütz, who claimed there were two relational elements—object and authority [17,18]. In later accounts, the number of relational elements varies, from three to four [5] (p. 16) [19–22] to six [4] (p. 570), and even seven [23].

If one is willing to agree to this minimal definition of responsibility it becomes clear that a complex cluster of capacities is needed to call someone responsible: (i) the ability to communicate, (ii) autonomy, which includes being aware of the consequences, i.e., knowledge, being aware of the context, i.e., historicity, personhood, and a scope of influence, and (iii) judgement, which includes several cognitive capacities such as reflection and rationality, as well as interpersonal institutions such as promise, trust, and reliability. It is important to take into consideration that these three sets of capacities can be ascribed in a gradual manner and that a potential subject of responsibility must learn these competences over a period of several years. As it is possible to speak of better or worse communication skills, to say that someone is more or less able to act in a specific situation, is more or less autonomous, reasonable, and so on, it follows that, according to the present prerequisites, responsibility itself must be attributed proportionately. Assigning responsibility is not a question of 'all or nothing', but one of degree [24].

2.1. The Subject of Responsibility

The subject or bearer of responsibility is the one who is answerable. I have already elaborated above on the competences that a potential subject of responsibility has to be equipped with in order to bear responsibility. Given these prerequisites, one might ask whether only 'healthy and grown-up' people are potential responsible subjects or whether children, for instance, are too. Maybe (some) animals or plants, or even inanimate things (such as artificial systems), could also be called responsible? After all, we claim the sun is responsible for melting the candle wax. But within responsibility research, the phenomenon of responsibility is traditionally interpreted as an individualistic concept that is closely linked to personhood [2] (pp. 66–69). This core aspect of the conventional understanding of responsibility has not been questioned until recently. Against the backdrop of these thoughts it becomes clear why every purely descriptive or causal usage of the term "responsibility"—one that does not allow for a normative interpretation, such as in the example of the responsible sun—is only metaphorical [2] (pp. 37–39) [7] (p. 242) [25]. The sun cannot be answerable. In this paper I will not discuss whether (some) animals and children might justifiably be called subjects of responsibility,[2] but rather concentrate on the question of whether (some) robots could be interpreted as bearers of responsibility (see Section 5.1).

An accurate definition of responsibility in a particular context includes asking whether the responsible subject in question is an individual or a collective [2] (pp. 71–94). In the case of an individual bearer of responsibility, one has to differentiate between sole, personal and self-responsibility. In a responsible collective there are mechanisms of delegating, sharing and dividing responsibility to be identified, which ascribe partial, full, or no responsibility to the members of the collective. In Section 5.2 I will introduce the term "responsibility network" [30] to focus on the different functions and roles that the involved parties within man-robot interaction have.

2.2. The Object of Responsibility

No one is simply answerable: one is always answerable for someone or something—the object or matter of responsibility. At first sight, one might differentiate four categories of objects of responsibility: people are responsible for creatures (e.g., children or animals), for things (e.g., my father's glasses that I am told to get from the optician), for events (e.g., a robbery), or for actions and their consequences. However, on closer consideration, one can recognize that every responsibility can be translated into a responsibility for actions and the consequences of actions. The parents' responsibility for their

[2] On the issue of learning to bear responsibility, see [2] (pp. 143–159) [26–29].

children is a responsibility for their children's well-being, which the parents can guarantee via specific actions. As well as being an event, the robbery is an arrangement of actions that collectively define the responsibility for the robbery. The subject and the object of a specific ascription of responsibility are linked via roles that organize and structure our everyday life. Roles define objects that we are responsible for—notwithstanding that some roles are much more clearly defined than others [14] (p. 292) [25] (p. 543) [31].

Objects of responsibility are per se part of a retrospective (ex-ante, backward-looking) or prospective (ex-post, forward-looking) responsibility [2] (pp. 103–104) [32,33]. To determine the specific moment of a responsibility, the object in question has to be known, at least approximately. In the majority of situations this is possible without any difficulties: the responsibility of a defendant for a theft, for instance, is obviously a retrospective responsibility, because in order to bring someone before a court the matter has to have already happened. However, in some contexts a temporal localization of the object of responsibility in the past or future is less clear without further explanation—for instance, in sentences such as "humans are responsible for climate change". Here, the term "climate change" might be used to denote the result of changes to the global climate that have already occurred, in which case the responsibility in question would be a retrospective one. On this interpretation, the said people are responsible by their actions for the climate change that has led to the current state. On the other hand, the people's responsibility for climate change could be a prospective one, if understood as having a responsibility to improve the global climate.

In any case of ascribing responsibility, specifications regarding its prospective or retrospective nature are necessary in order to allow the responsible subject to know what her or his being answerable for the said object requires from her or him.

2.3. The Authority of Responsibility

Along with the subject and the object, the authority of responsibility is the least questioned relational element of the concept of responsibility, perhaps owing to the fact that the term first appeared in the sphere of criminal law, where a defendant is answerable to a court and a judge. As long as the ability to respond is closely linked to personhood, inanimate things, plants, animals, and infants are not able to fulfil the role of the authority [34].

The court and the judge are examples of external authorities, while personal conscience is the most well-known internal authority in contexts of ascriptions of responsibility. External authorities have often limited and disputable scope and power. Absolute, indisputable, and 'final' authorities [35]—such as God [36] or history [37]—are highly controversial in the responsibility research community [25] (pp. 546–547) [35,38]. Furthermore, authorities are public or private. A public authority is intersubjectively accepted and is in a position to impose sanctions on the responsible subject. Publicity does not refer to presence in the media, but rather to the recognition and approval that results in a psycho-social pressure on the responsible person to follow the authority's claims. The public status of an authority and its approval is not to be confused with the endorsement of, for instance, a law or a judicial order. Approval of the authority refers to acceptance of the legitimate procedure of a judicial order which leads to the defendant's meeting her or his obligations. In light of these reflections, an author's bestseller might enjoy great presence in the media without necessarily having public approval. Privacy of an authority, on the other hand, does not mean vagueness or ambiguity: social conventions are an example of public but not necessarily clearly defined norms. Private authorities and private norms are not intersubjectively accepted, and enjoy only little potential for applying sanctions. Whether an authority is private or public depends on the specific context and on the normative criteria that frame the ascription of the responsibility in question (see Section 2.5 below).

2.4. The Addressee of Responsibility

Within the responsibility research community, the addressee or receiver is the relational element that in my view provokes the most disagreement. Often one does not sufficiently differentiate between the authority and the addressee due to the mistaken belief that the latter is irrelevant [5] (p. 16) [35] (p. 127) or because both relata are conflated in theory [17] (p. 256) [34,39]. Very rarely, the addressee is defined as a genuine relational element of responsibility [5] (p. 570) [40].[3]

The addressee is affected by the subject's responsibility and defines the source for the existence of the said responsibility. Birgit Albs summarizes the fundamental role of the addressee in stating "no plaintiff, no judge" [40] (p. 27). In the case of theft, the person from whom the thief has stolen (furthermore, the citizens as a whole or the infringed legal norm itself), the addressee, is the ground for the thief's responsibility. This also shows that the addressee (unlike the authority) need not meet the prerequisites for bearing responsibility (autonomy, judgement, and the ability to communicate). Inanimate things such as norms and values as well as beings are potential sources for ascribing a certain responsibility. The addressee, as the why of a responsibility, realizes the normativity of the minimal definition of responsibility. Because purely descriptive acts of imputation and ascription exist, normative criteria require an explanation and justification that the addressee guarantees.

Like the authority, the addressee is private or public.

2.5. The Normative Criteria of Responsibility

The final relational element, that of normative criteria, defines the conditions under which a subject is responsible. However, the concept of responsibility does not include a specific set of norms; it is "parasitic" on a particular normative foundation [5] (p. 65). There are different normative criteria, for instance, values, principles, imperatives, maxims, laws, rules, orders, tasks, and instructions. Among these, duties traditionally play the most crucial role [38] (p. 54) [41], maybe because they are one of the most familiar to the responsibility community along with ascriptions of guilt [18] (pp. 197–219). Defined as a duty or set of duties, responsibility is a deontological term [42,43].[4]

A responsibility's normative criteria need justification via norms, which themselves require justification, and so on ad infinitum [35] (p. 106). They define spheres of responsible acting and types of responsibility such as criminal, legal, political, moral, or economic responsibility, bound by criminal, legal, political, moral, and economic norms. A principle such as "one is not allowed to kill somebody", for instance, can serve as a moral rule, a religious imperative, or a law—depending on the context [2] (pp. 125–128).

Like the authority and the addressee, the normative criteria are private or public. Moral norms tend to be private, whereas political and legal norms tend to be public. However, not every moral principle is equally private, and not every legal norm can impose sanctions against someone. Some norms of international law, for instance, are not binding in the same way that norms of the criminal code are. Also, the example given above, "one is not allowed to kill somebody", might have a more private status when formulated by parents towards their children as an imperative of education than during the Easter mass as a theological guiding principle.

3. What Is Robot Ethics?

Robot ethics is a sub-category of machine ethics and represents a relatively new form of applied ethics within the so-called Western cultural area [12,50–56][5]. All robots are machines, but not all

[3] The reasons for this are discussed in [2].

[4] "Responsibility" has also been understood as a consequentialist term [44–46] and as part of virtue ethical approaches [47–49] or as a virtue itself [25] (p. 543) [43].

[5] Another sub-category of machine ethics is computer ethics. Machine ethics, again, is a sub-category of the sub-discipline ethics of technology which, in turn, is a field in philosophy of technology.

machines are robots [57]. This classification of robot ethics as an applied ethics sub-category of machine ethics is based on an understanding of ethics as a category of action that is initially specific to humans. Thus, the philosophical discipline of ethics, since the classical definition of Aristotle, has been concerned with human customs, traditions and habits, with the good life, has carried out the scientific reflection of human practice and judged the criteria of good and bad action [58]. A lot of people assume that only human beings are acting beings, whose doings are not subject to blind instinct and drive, but are formed by intentions, norms and reasons.

I differentiate two types of applied ethics: One concerns those ethical systems that deal with a nonhuman counterpart. In addition to machine and robot ethics, these ethical systems also include animal, plant, environmental, computer ethics, and, in general, ethics of technology. Similar to robot ethics, animal ethics, for instance, focuses on the normative criteria that play a role in the breeding, domestication and keeping of animals, in general in dealing with animals and in the relationship between humans and animals ([59–67]; regarding plant ethics cf. [68]). The robot ethicist David Gunkel has, in *The Machine Question* (2012), already stated that robot ethics is related to animal ethics in particular, in so far as "[t]he machine question [...] is the other side of the question of the animal" [69]. René Descartes, as Gunkel understands him, initially attributed the same ontological status to animals and machines [69] (p. 3). It was not until the 20th century that this ontological equality between animals and machines was abolished in favor of animals.

The other group of applied ethics assembles ethical systems for special domains of human life, where values are represented, norms are enforced, and rules are formulated that are usually attributed a different status in people's everyday lives. Medical ethics, ethics of humanitarian interventions, ethics of war and of economics, and ethics of international relations can be cited as examples of this form of applied ethics. This second type of applied ethics is only marginally addressed, if at all, in this paper.

In German-speaking countries, robot ethics is not yet a generally recognized discipline within academic philosophy, even though interest in interdisciplinary collaborations, including philosophers, is growing. In comparison with the English-speaking world, where the ethical examination of artificial systems has produced a canon of classical literature since the middle of the last century [10,11,70–79], the German-language discourse is, although increasing, still relatively small [8,9,57,80–83].

Robot ethics was above referred to as sub-form of machine ethics, since all robots are machines, but not all machines are robots. A machine is an artificial structure that consists of individual parts moved by a drive system (motor, wind, water, etc.) and converts energy [84,85]. Robots are special machines. Historically, the term "robot" goes back to the Czech word "robota" for work, compulsory service, and forced labor, and was coined in 1920 by the artist Josef Čapek. His brother Karel Čapek used it in the play *R.U.R. Rossum's Universal Robots* (1921) for humanoid apparatuses that render services to humans. In this original understanding is anchored the idea of robots as artificial slaves (from the Slavic root "rab" for "slave"; [86]) and assistance systems that relieve humans of tiresome, boring (e.g., repetitive) and dangerous work. The first established human domain where robots were deployed to undertake these dull, dangerous and dirty jobs, namely in industry, reflects this vision, which Čapek creates in the above-mentioned play. It is also the core of the so-called industry 4.0, the technological transformation of the human working environment through digitization and automation. Čapek is, however, no blind technology enthusiast and in *R.U.R.* deals with numerous challenges which accompany the creation of robots. The plot of his play ultimately amounts to a revolt of robots who are no longer willing to submit to their human creators, to turn against humans in order to take over world domination. But Čapek also raises many other philosophical questions, such as the nature of 'the' human being, the responsibility of scientists for their artificial creatures, and what it means to form an emotional bond with another being. His play ends with the prospect of a love relationship that begins to develop between two robots. Thus, in the historical understanding of the robot founded by Čapek, a broad basis is laid for the discussions that should arise in the following decades.

The following reflections are based on an extended version of the definition proposed by Catrin Misselhorn, according to which a robot is an electro-mechanical machine that (a) has some form of independent body, (b) consists of at least one processor, (c) with sensors that collect information about the world, (d) and effectors or actuators that translate signals into mechanical processes. The behavior of a robot (e) is, or at least appears to be autonomous and it can (f) influence its environment [71] (p. 48) [82] (pp. 28–31) [86] (pp. 41–59) [87,88]. This understanding is not unproblematic, since some of the conditions cited (such as embodiment, autonomy and influence) are ambiguous and therefore at least in need of explanation. It also excludes those artificial systems, such as computers, chatbots, medical assistance systems, and drones, which have so far often been referred to as robots. These machines open up the large grey area which deserves to be illuminated by philosophers of technology and robot ethicists. Isaac Asimov has already suspected this robot-related sphere in which we also meet these other relatives and acquaintances of robots [89]. In fact, according to the definition given here, something is not a robot in the strict sense if one of the mentioned conditions (a)–(f) is not fulfilled: Computers do not meet condition (f) and whether they possess an independent body—condition (a)—remains to be discussed. The specific kind and manner of embodiment, on the other hand, plays no role, robots exist in every conceivable shape, the humanoids among them are called androids [90]. In a metaphorical sense, computers represent the 'brain' of a robot, just as algorithms can metaphorically be regarded as their 'nervous system', or even better as their 'mental behavior patterns' and 'learned processes', but not as the actual robot itself. Artificial systems such as the surgical assistance system DaVinci and drones lack condition (e) [86]. The conditions (a)–(f) are necessary in detail, and together are sufficient to provide a complete definition of "robot" in the strict sense anticipated in this paper.

4. The Three Fields of Robot Ethics

When I talk about fields of research in robot ethics in the following, I don't mean that these are different ethics or ethical systems, but different areas of the discipline of robot ethics, in each of which different ethical approaches are developed. Traditionally, research in robot ethics distinguishes two fields. In one area, it is discussed to what extent robots are to be understood as moral patients, i.e., as having value and perhaps even rights, but in general above all as objects of moral action. The other field deals with the question of the extent to which robots themselves have to be interpreted as moral agents and thus as subjects of moral action. The two fields of work are not necessarily exclusive; they complement each other. Both face—thanks to their first ethical (Aristotelian) premises—some challenges that the third (and younger) field intends to meet.

Within the field in which robots are regarded as moral agents, the question is asked to what extent robots are capable of acting morally, what competences they have to be equipped with, and to what extent, if they are to be able to act morally. Depending on the underlying understanding of agency, morality and the competences that are necessary for its realization, this field of robot ethics focuses, for instance, on the attribution of freedom and autonomy as a condition for moral agency, on cognitive competences (e.g., thinking, mind, reason, judgement, intelligence, consciousness, perception, and communication), but also on empathy and emotions [55,76,79,91] [87] (pp. 30–42). It should be noted, however, that the robots that exist so far are candidates for the attribution of moral patiency and not of moral agency.

Within the field in which robots are discussed as moral patients, as objects of moral action, the issue is how to deal with artificial systems, what value they have, even if one holds that they are not capable of morally acting themselves [92–97]. Here robots are consistently understood as tools or complements of human beings. Topics within this robot-ethical field include the formulation of codes of ethics in companies, the desirability and possibility of relationships with robots, the 'enslavement' of robots or the assessment of the use of robots for therapeutic purposes. Some thinkers also advocate or discuss the possibility of attributing (some) robots rudimentary or even fundamental rights. Similar to Immanuel Kant in § 17 of the second part of his *Metaphysics of Morals*, which argues against cruelty to animals because this leads to a brutalization of humans, Kate Darling [98], for instance, argues for robot rights, since, according to her, humans are then more likely to succeed in remaining 'human'. The European Parliament, too, is currently working on a concept that would allow (some) robots to be given the status of "electronic persons". Within this area of robot ethics, the moral competence (i.e., the decision on the behavior of a robot, which happens through the personalization of a respective artificial system by its users) remains consistently with the human owners. The competence–competence (i.e., the decision on any framework values and principles which are specified in the programming of the respective robot and cannot be changed by the users) remains with the manufacturers and distributors or with the law. Within this field of robot ethics, it is therefore only humans who decide on the morality of their creatures and who are responsible in the event of an accident.

The group of moral agents is generally much smaller than that of moral patients, because usually we only distinguish people with moral abilities in the genuine sense. However, a whole series of beings and things is attributed a moral value or even (rudimentary) rights, so they are moral patients—at least to the extent that these entities are morally worthy of consideration, even if they may not have any intrinsic value, but only a high instrumental value. As a moral agent, a being is at the same time a carrier of values; conversely, not all carriers of values are moral agents equally. The attribution of moral values to living beings and objects depends, for example, on the perspective taken in each case. An anthropocentric position argues that only humans have an intrinsic value. Anthropocentrism means on the one hand that people have a moral advantage and thus special status over all other beings (moral anthropocentrism), and on the other that only people have the capacity for knowledge and can develop a capacity for judgement (epistemic anthropocentrism) [99]. An alternative to anthropocentrism is, for instance, pathocentrism, which morally ascribes an intrinsic value to beings capable of suffering and epistemically advocates the position that values come into the world through beings capable of suffering. As a result, they must also be objectively accepted outside the human cognitive horizon. Biocentrism, which considers all living morally and epistemically, and physiocentrism, which considers all nature (holistic physiocentrism) or everything in nature (individualistic physiocentrism) morally and epistemically, can also be classified in the circle of centrist approaches, as they are called below. The inclusion of robots in the horizon of things endowed with an intrinsic value could perhaps open up a further perspective, a mathenocentrism, for instance (from the Greek "matheno", "learning"), which measures all this with an intrinsic value that is controlled or programmed or capable of learning in a specific way.

The above-mentioned understanding of ethics in general, which goes back to Aristotle, and the resulting status of applied ethics, as well as these and all other conceivable centrist approaches, are now accompanied by some serious philosophical challenges, which the inclusive robot-ethical approaches (the third field of robot ethics) intend to meet. For the starting point within the classical two fields of robot ethics (robots as potential moral agents and moral patients) is always first of all 'the' human being who defines the ethical commonsense, depicts the 'normal case' or the genuine and 'ideal type' of ethical agency. Thus every (Aristotelian) understanding of ethics that defines 'man' as the pivot rests, at least implicitly, on an anthropological foundation.

The anthropological disciplines in biology, cultural studies, ethnology and not least philosophical anthropology each take their own paths in order to define the human being and to differentiate humans with the greatest possible clarity from all other beings. In doing so, they usually proceed in an essentialistic way, i.e., they seek to describe 'the' human being by means of an individual attribute or a series of characteristics. The *Historical Dictionary of Philosophy* presents in the entry "Mensch" a total of four common anthropological approaches, all of which are ultimately essentialist [100]: The Alkmaion topos defines 'the' human being by a single attribute. Alkmaion, a pre-Socratic from the late 6th or early 5th century BC, distinguishes 'the' human from other living beings for the first time via a single competence. He defines 'man' "by the fact that he alone understands, while the remaining creatures perceive, but do not understand" (my translation; [101]). On the other hand, the compositional topic ("Kompositionsformel"; my translation) captures 'the' human being not only by one property, but by two or more attributes. The microcosm–macrocosm approach understands 'the' human being as what the cosmos is on a small scale. The horizon topic ("Horizontformel"; my translation) finally determines 'the' human being "through the spheres into which he projects upwards and downwards" (my translation; [100] (p. 1073)).

This essentialism, which is taken for granted in philosophical anthropology and other (anthropological) disciplines, poses certain challenges, such as moral status attributions, which suggest a specific treatment of animals, machines and other alterities. Even radically exclusive positions such as speciesism, racism and sexism often argue essentialistically by denying the excluded beings certain characteristics. (Anthropological) essentialism continues to be accompanied by an epistemic uncertainty as to whether the being in question actually possesses the attributed qualities. Our attribution of certain competences, such as freedom of will, is often based on a metaphysical foundation. Actually, we do not only not know what it is like to be a machine or an animal, e.g., a bat—to quote the title of a famous paper by Thomas Nagel [102]—but we also do not really know what it is like to be another human. Because it cannot be determined with unambiguity whether humans are actually equipped with freedom of will and similar abilities. We cannot clearly prove them empirically. The only difference lies in the fact that we are willing to make an additional assumption with humans, namely that at least the beings we call humans have the competences in question [103,104].

Those applied ethics approaches that are (implicitly or explicitly) based on Aristotelian, anthropological, and centrist principles address a respective and also essentialist defined counterpart to 'man', who has a moral status due to certain characteristics. Within a pathocentric approach, for example, all beings to whom sentience can be attributed are given a place in the moral universe. Furthermore, these positions have to react not only to (moral) centrism and (anthropological) essentialism, but also to the consequences of an at least implicitly assumed subject-object dichotomy. This subject-object dichotomy underlies an Aristotelian understanding of ethics as well as centrist approaches in general and emerges, on closer examination, as a consequence of epistemic centrism. For instance, epistemic anthropocentrism recognizes a specific subject, namely 'the' human being, who, endowed with cognitive capacity, judgement and reason (or comparable competences), attributes values to all other beings as the objects of 'his' cognitive practice and is the only being that brings values into the world in the first place. The same applies to other morally centrist approaches, such as moral pathocentrism, in combination with epistemic anthropocentrism. But other epistemic centrisms also remain committed to the paternalism implicit in the subject-object dichotomy by defining a respective cognitive subject that assigns or denies values, abilities and competences to an imaginary counterpart as the object.

The inclusive approaches of robot ethics present in quite different ways alternatives to the Aristotelian understanding of ethics, the anthropological essentialism that accompanies it, the moral and epistemic anthropocentrism that is often embedded in it, and face the philosophical challenges of centrist approaches in general. Here, too, the debate revolves around the attribution of moral agency and moral patiency. With regard to the ascription of moral agency, however, the traditional understanding of the moral agent is questioned in order to extend it to nonhuman beings. On the other hand, competences that have been ascribed essentialistically to individual subjects of actions within the framework of the usual approaches should now be understood relationally and processually as realizing themselves in the interaction of different agents and non-agents [69,105–109]. With a view to robots as objects of moral actions, the debate in this third area of robot ethics also revolves around, for instance, the anthropomorphization of nonhuman beings and the possibility of entering into relationships with them.

In their concern, the inclusive approaches overlap in many respects with post-feminist, post-structuralist, post-colonial and critical-posthumanist positions. The unifying intention of all these perspectives is an inclusive program (morally and epistemically) that neither emphasizes the position of 'the' human being (whatever may be hidden behind this label in detail) over other beings or negatively devalues 'the' human to the 'level' of all other beings. Exclusive theories (with regard to the topic of this chapter approaches on robots as potential moral agents or moral patients) either value 'the' human being as special and better than all other beings (this position could be called an exclusive authoritarianism) or exclude 'the' human being from the genuine realm of morality (call this exclusive relativism[6] a position less frequently defended than the first). In contrast to these exclusive theories, the inclusive approaches seek to place all other beings on the same 'level' as human beings and thus include them as morally equal companions with accompanying fundamental moral rights in the moral universe in which human beings are located[7]. It must, however, be pointed out that it is not the aim of the inclusive approaches of robotic ethics to put an end to essentialism or to the distinction between subjects and objects in general. A critical awareness of the morally highly questionable consequences of some concrete essentialisms as well as of some specific differentiations between subjects and objects is needed.

The here introduced three fields of research within robot ethics pose the question of what is moral and of what is a moral judgment. In *Moral Machines: Teaching Robots Right from Wrong* (2009), Wendell Wallach and Colin Allen state that those beings are moral agents that are in situations that require moral judgement. They introduce Philippa Foot's famous thought experiment of so-called trolley cases [79] (pp. 83–100) as an opener to discuss whether driverless train systems in London, Paris and Copenhagen (since the 1960s) morally 'judge' when they are programmed to stop whenever there are people on the tracks, even though passengers might be injured due to the abrupt halt [79] (p. 14)[8]. Although the autonomous train, programmed with a specific algorithmic structure, is not genuinely able to act morally, this situation is phenomenologically comparable to those that humans might experience. That—according to Wallach and Allen—suffices to interpret artificial systems as quasi-agents without claiming them to be genuine moral agents in the same way as humans. I will elaborate on their approach as a version of the weak AI thesis in Section 5.1.

5. Ascribing Responsibility in Man-Robot Interaction[9]

It is often denied that artificial systems can bear responsibility, due to their supposed lack of the necessary competences that one normally claims only human beings to be equipped with: robots, following this traditional approach, don't have the ability to act or autonomy, judgement, the ability to communicate, or any other morally relevant capacity. Wallach and Allen outline an approach of functional equivalence to overcome this problem of lacking competences in artificial systems. In the next three sections, I elaborate on the role and function of responsibility within the three fields of research in robot ethics, robots as moral agents, robots as moral patients, and inclusive approaches of robot ethics (Section 4).

5.1. Robots as Moral Agents and the Prerequisites for Ascribing Responsibility: Wallach and Allen's Approach of Functional Equivalence

In asking whether robots are to be interpreted as "artificial moral agents (AMAs)", Wallach and Allen [79] (p. 4) define moral agency as a gradual concept [24] with two conditions: "autonomy and sensitivity to values" [79] (p. 25). Human beings are the genuine moral agents, but some artificial systems—such as an autopilot, or the artificial system Kismet—might be considered as "operational" moral agents. They are more autonomous and sensitive to morally relevant facts than non-mechanical

[6] A position that can perhaps be assumed with Joseph Emile Nadeau [110], who only attributes genuine moral ability to androids. John Danaher discusses in his text "The rise of the robots and the crisis of moral patiency" the thesis that the development of artificial systems "could compromise both the *ability* and *willingness* of humans to act in the world as responsible moral agents".

[7] To illustrate the distinction between in- and exclusive approaches I will use two current examples: The German AfD represents a morally exclusive authoritarianism, since it discriminates against some people morally, i.e., excludes them from the circle of the morally best qualified persons. US President Donald Trump and his advisor Kellyanne Conway can be interpreted with the phrases "fake news" and "alternative facts" as the implicit precursors of an epistemically exclusive relativism, which in the final analysis threatens to abolish any distinction between "right" and "wrong", "fact" and "fiction" and, until then, first tries to discredit a number of journalistic media.

[8] For the debate on autonomous driving systems, see [13,111–114].

[9] Some of the Thoughts Expressed in Sections 5.1 and 5.2 Have already Been Presented in a similar Way in [13].

tools such as hammers. However, they are still “totally within the control of [the] tool’s designers and users” [79] (p. 26) and in this sense are “direct extensions of their designers’ values” [79] (p. 30). Thus far, only very few robots have the status of “functional” moral agency, such as the medical ethics expert system MedEthEx [115]. In the sense defined by Wallach and Allen functionally moral machines “themselves have the capacity for assessing and responding to moral challenges” [79] (p. 9). The authors claim that “[j]ust as a computer system can represent emotions without having emotions, computer systems may be capable of functioning as if they understand the meaning of symbols without actually having what one would consider to be human understanding” [79] (p. 69).

With this notion of functional equivalence, Wallach and Allen subscribe to a version of the weak AI thesis, which holds that only the simulation of certain competences and abilities in artificial systems is possible, whereas the strong AI thesis claims that the construction of robots that genuinely are intelligent, conscious, and autonomous in the way humans are is theoretically possible [116]. According to Wallach and Allen, a strong AI understanding of autonomy is not a necessary condition for constructing AMAs. Instead they focus on the attribution of functionally equivalent conditions and behavior. Functional equivalence means that specific phenomena are treated ‘as if’ they correspond to cognitive, emotional, or other attributed competences and abilities[10]. The question of whether artificial systems can become intelligent, conscious or autonomous in the strong AI sense is replaced by the question to what extent the displayed competences correspond to the function they have within moral evaluation—in the context of this paper, the concept of responsibility. However, although Wallach and Allen claim the boundary between functional morality and full moral agency to be gradual [24] with respect to certain types of autonomy, for the foreseeable future it is hard to fathom how an artificial system might achieve functional equivalence with the genuinely human ability to set “second-order volitions” for oneself [117], to act as “self-authenticating sources of valid claims” [118], or to reflect on one’s own moral premises.

In tackling these questions, it may be helpful to supplement Wallach and Allen’s approach with Stephen Darwall’s distinction between four different usages of “autonomy”, namely, “personal”, “moral”, “rational”, and “agential” autonomy [119][11]. While personal autonomy means the ability to form values, goals and ultimate ends, Darwall defines moral autonomy as the competence to reflect on one’s own moral principles. One might conceive of these two forms of autonomy as being reserved exclusively for human agents. Rational autonomy, Darwall’s third type of autonomy, might be achievable for artificial agents as well as humans, since it is grounded in an action solely on the basis of the “weightiest reasons”, which may be represented in a functionally equivalent way by algorithms [119]. More important, however, is the ascription of agential autonomy to artificial systems, since this means identifying a certain behavior as a “genuine action”—that is, not completely determined by external factors. This may be functionally represented by the robot’s ability to change internal states without external stimuli.

On a computational level, these forms of autonomy and the ability to autonomously change internal states might be described by distinguishing three different types of algorithmic schemes [83] in referring to two types of algorithms. Determined algorithms always give the same output, whenever a particular input is given. Deterministic algorithms always give the same output, whenever a particular input is given, in passing through the same sequences of states. Hence, deterministic algorithms are a subcategory of determined algorithms, because every deterministic algorithm is a determined algorithm, but not every determined algorithm is a deterministic algorithm. Potentially, machines that predominantly function based on deterministic algorithms might be neither functional nor operational. They are almost closer to non-mechanical tools such as hammers than to operational robots. Artificial

[10] Technically this argument applies to humans as well, and all the more so to animals [102]. In general, humans are willing to prima facie grant capacities like reason, consciousness, and free will to other individuals. However, there is no guarantee that this assumption really holds [107] (p. 63).

[11] This idea is due to Wulf Loh [13].

systems that predominantly function on the basis of determined (but non-deterministic) algorithms might then be understood as operational robots. Finally, those few robots that are predominantly structured by non-determined (and thereby non-deterministic) algorithms are to be understood as functional artificial systems.

Let us consider a few examples. Wallach and Allen define the artificial system Kismet as an operational AMA. Supplementing their approach with my understanding of responsibility and the necessary prerequisites for ascribing the ability to act responsibly (autonomy, judgement, and the ability to communicate) Kismet possesses a rudimentary ability to communicate since it can babble in simple noises. Judgement (if one is willing to call Kismet's behavior reasonable at all) is barely recognizable in Kismet's reaction to very simple questions. The biggest challenge in regarding Kismet as an operationally responsible robot is clearly its autonomy, since the relevant sub-capacities (knowledge, historicity, personhood, and scope of influence) are very limited. In its rudimentary mobility, Kismet can autonomously move its ears, eyes, lips, and head, and respond to external stimuli such as voices. Hence, Kismet is, as Wallach and Allen suggest, still completely under the operators' and users' control; it does not artificially learn, and its algorithms only allow for deterministic results. To call Kismet responsible might seem comparable to calling an infant or animal responsible. However, in contrast to the hammer hitting one's thumb or the sun melting the candle wax, the case of Kismet might (as with infants and animals) open room for debate on ascribing responsibility to artificial systems like it, although understandably, the room for debate appears to be small.

Cog is a robot that can interact with its surroundings because of its embodiment. It might pass as an example of a weak functionally responsible agent, since its ability to communicate, as well as its judgement, has been greatly improved compared to Kismet. Furthermore, Cog's overall autonomy has evolved, since it includes an "unsupervised learning algorithm" [120]. For instance, after running through numerous trial-and-error attempts to propel a toy car forward by gently pushing it, Cog will eventually push the car only from the front or from behind, not from the side, since only then will it move. Cog has not been programmed to solve the task in this manner, but learns from experience. Because of its limited capacity to learn, one might understand it as a weak functional agent, or at least as a very strong case for an operational ascription of responsibility. Calling Cog responsible might be comparable to ascribing responsibility to a very young child.

Autonomous driving systems (the last example given here) might be identified as operational rather than functional artificial agents. While their communicative and judgement skills are as well-developed as Cog's capabilities, or even greater, their overall autonomy is still strictly limited, owing to their lack of learning and non-determined (non-deterministic) algorithms. I will show in the next section that responsibility with regard to autonomous driving systems can be distributed through a responsibility network, because one cannot ascribe responsibility to the autonomous cars themselves.

To sum up, in combining Darwall's four types of autonomy with Wallach and Allen's approach of functional equivalence, one can draw a clear line between full-blown (human) agency and artificial (operational and functional) agency. While human agents are equipped with all four types of autonomy, artificial systems may only possess rational and agential autonomy in a functionally equivalent way for the foreseeable future. As far as certain domains of responsibility are concerned, an artificial system can be called autonomous as soon as it meets the criterion of functional morality.

Ascribing responsibility to artificial systems is possible, yet to this day only in very restricted terms. In the field of evolutionary learning systems, the designers of machines that will be able to learn orient themselves by developmental-psychological theories for children learning. The roboticists' approaches are based on a metaethical premise on the context-sensitivity of morality: moral and responsible agency needs experience and situational judgement. Both are artificially possible only in embodied systems. The notion that a system's interaction with its environment was a prerequisite for the artificial development of abilities and competences was first considered in the 1990s by Rodney Brooks, who subsequently established the field of "behavior-based robotics" [120] (pp. 52–87) [121]. The idea to construct robots that develop competences in the way children develop competences has been used in the design of a great many artificial systems, such as iCub, Myon, Cb^2, Curi, and Roboy. I assume that these projects at least partly focus on non-determined (and thereby non-deterministic) algorithms that are able to learn (evolutionary algorithms). However, so far, machine learning is not possible in moral contexts, or only in weak moral contexts: to this day, one cannot ascribe responsibility to artificial systems–at least as long as one understands responsibility as competence, which presupposes the ability to communicate, the ability to act or autonomy, and judgement. Because these are skills that are not simply given, but have to be learned (Section 2).

5.2. *Robots as Moral Patients and the Relational Elements of Responsibility: Responsibility Networks*

From the reflections in Section 4, together with the functional equivalence approach of Wallach and Allen (Section 5.1), it can be concluded that artificial systems cannot yet be regarded as responsible agents, since they lack the necessary competences, possess them functionally only weakly, or possess them only operationally. Recalling the etymologically minimal definition of responsibility as the normative ability to answer for one's actions, based on the psycho-motivational constitution of the subject, our traditional understanding of responsibility is clearly individualistic in that we always need to define a subject or bearer of responsibility (Section 2.1). If the prerequisites are not present, it is not, or is only metaphorically, possible to ascribe responsibility. If we actually feel that we need to ascribe responsibility but do not know to whom, some researchers have outlined alternative approaches that do not define the subject position [122]. However, I am skeptical whether without explicitly defining a responsible subject the term "responsibility" is able to serve its primary tasks, that is, to structure, organize, and regulate contexts that are opaque because they involve a large number of people and inscrutable hierarchies.

On the other hand, there exist situations in which the involved parties are not (fully) equipped with the necessary competences for bearing responsibility. However, we are still certain that we need to ascribe responsibility to someone. Consider again the case of autonomous driving systems (see Section 5.1) as operationally responsible agents (equivalent to the responsibility of an infant or animal). The autonomous car might be a moral patient in so far as it is part of our moral universe and (instrumentally) morally worth considering. However, it is not a moral agent in a significant (that is, at least functional) way. For contexts such as these I'd like to adopt Christian Neuhäuser's concept of responsibility networks [30] and elaborate on it [13]. The first assumption of this approach is that the involved parties are to be held responsible to the extent that they possess the necessary prerequisites for responsibility.

Responsibility networks have the following characteristics: (1) They have an enormous scale; (2) it is very hard to define responsible subjects or other relative elements such as the normative criteria or the authority; (3) they are found in contexts when it is unclear whether we might be able to define concrete responsibilities at all; (4) they combine several different responsibilities; (5) the involved parties usually have different positions (e.g., they are subject of a responsibility, object of another responsibility, and authority of yet another responsibility); (6) relational elements often overlap in responsibility networks. Consider, for instance, the parents' responsibility for their children (although this is not an example of a responsibility network, but rather shows how relata in some cases overlap): here, the children and their well-being are object and addressee of this responsibility [2] (pp. 171–181). Examples of responsibility networks include climate responsibility [2], responsibility in the global financial market system, and responsibility in road traffic.

Within, for instance, the responsibility network "responsibility in road traffic", numerous potentially responsible parties are involved, such as the human drivers, the owners of the (autonomous) cars, the companies that sell (autonomous) cars, the programmers, designers, lawyers, driving instructors, pedestrians, but also the public of a society with a common sense of moral norms. Regarding the object of the responsibility in road traffic, it is not possible to ascribe responsibility to one or a small number of subjects for 'the' road traffic as a whole, since this object is too huge and complex for one or a few persons to be fully responsible for. However, we can divide several spheres of responsible acting within the responsibility network "responsibility in road traffic", structured by different sets of norms that define equivalent responsibility types, such as moral, legal, and political norms. For all of these areas of responsibility, the road traffic serves as the overall object of responsibility, but is necessarily differentiated in smaller and less complex objects of responsibility that different parties are answerable for in different ways. Responsibility for the road traffic might, for instance, refer to the economic and moral responsibility for getting safely, efficiently, and as quickly as possible from A to B, or to the aesthetic responsibility for a pleasing design of roads and sidewalks, or to the moral responsibility for preparing children and young drivers for the ethical challenges to be met in participating in road traffic. Within these and further responsibilities as part of the overall responsibility network "responsibility for road traffic", numerous authorities, addressees, and normative criteria are to be defined.

To this day, an autonomous driving system that is to be identified as an artificial agent that is responsible only in a weak sense (as an artificial operational agent) cannot fill the subject position of a responsibility within the responsibility network "responsibility for road traffic", because there are several more qualified potential (human) subjects of responsibility. However, such an artificial system could be identified as object or even addressee of one or more responsibilities, and through this could be included in the responsibility network. To conclude, in this manner it is possible to integrate robots as moral patients in responsibilities—even in challenging situations that require the complex structure of a responsibility network.

5.3. Inclusive Approaches in Robot Ethics: Ascribing Responsibility Relationally

In Section 4, the inclusive approaches were presented as alternatives to the Aristotelian understanding of ethics, to the (anthropological) essentialism that accompanies it, and to the (moral and epistemic) anthropocentrism that is often included in it. Inclusive approaches are concerned with questioning the traditional understanding of the moral agent to extend it to nonhuman beings. In addition, and this is particularly relevant for this chapter, competences that have been ascribed essentialistically to individual subjects of action within the framework of the usual exclusive approaches (Sections 5.1 and 5.2) should now be understood relationally as realising themselves in the interaction of several human and nonhuman agents. In the following, a relational understanding of responsibility in a narrow sense will be developed against the background of the inclusive paradigm. I am therefore talking about a relational concept of responsibility in the narrow sense, since the phenomenon of responsibility is already a relational concept with five relational elements from the outset (see Section 2). This "grammatical" relationality, which responsibility shares with many other linguistic concepts (a theft, for example, is a relational concept with at least two relata, requiring the definition of a subject and an object) and therefore represents a relationality in the broad sense, is not meant here. Relationality in the narrow sense focuses on responsibility as something that takes place exclusively in the interaction between beings and cannot be attributed as an attribute to a single being. Inclusive robot-ethical thinkers therefore do not only focus on nonhuman beings (especially robots) as potential subjects of action. Rather, they seek agency and the competences associated with it (here above all responsibility) as relational in the narrow sense in interaction or, to speak with Lucy Suchman and Karen Barad, in intra-action and interference of human and nonhuman agents.

The inclusive concept of responsibility is relational in the narrow sense for two reasons. On the one hand, the acting subject must learn to see itself as not monadic and not self-sufficient, as a subject that always already interacts and is entangled with other human and nonhuman beings. The narrative of the self-sufficient agent is a social, legal, and political illusion—perhaps necessary for the functioning of our society, but nevertheless a construct—with very concrete, real and material consequences. Responsibility cannot be attributed to a single agent, nor can judgement, autonomy, and other competences. On the other hand, responsibility arises from and in the interaction with human and non-human beings. For the so-called object of knowledge cannot be understood as independent of the observer, it is not simply 'found' in reality at some point in time, but is fundamentally created by the observer (see also Varela's et al. concept of enaction [123]). Responsibility arises in this process and is carried by the entire apparatus of human and nonhuman 'subject–objects'. This, of course, does not mean that individuals cannot be called responsible for their actions, and it does also not mean that all individual actions become excusable through 'the circumstances'. However, inclusive approaches of responsibility, such as Donna Haraway's relational concept of responsibility, in the narrow sense of the word, does imply that the circumstances and the respective situation must be included in the assessment of what someone has 'done'.

The step from a critical-posthumanist approach, such as Donna Haraway's, with responsibility as a relational concept in the narrow sense towards a techno- and robot-ethical theory can be taken with the text "Beyond the skin bag: on the moral responsibility of extended agencies" (2009) by F. Allan Hanson, which will be presented exemplarily in the following. Here, he contrasts the traditional position of a methodological and ethical individualism with the theory of an "extended agency" [105] (p. 91). For several centuries we have been accustomed, according to Hanson, to understanding a subject of action as an autonomous, monadic entity, even if this notion is neither historically uniform nor particularly old. For it is based on the idea of individuality, which emerged only after the Middle Ages in the so-called Western cultural space and which cannot claim a global status up to the present day (e.g., with regard to the so-called Asian cultural space; [105] (pp. 91–93)). On the basis of this methodological individualism, "the moral responsibility for an act lies with the subject that carried it out" [105] (p. 91). If one is prepared to deviate from this view of the agent, also changes the "concept of responsibility" [105] (p. 91) associated with it. Hanson, too, is moving from a deconstruction of the classical understanding of the agent to rethinking the competences and abilities essentialistically attributed to her or him. "The basic reasoning behind this extension of agency beyond the individual is that if an action can be accomplished only with the collusion of a variety of human and nonhuman participants, then the subject or agency that carries out the action cannot be limited to the human component but must consist of all of them" [105] (p. 92).

Referring to Andy Clark, Donna Haraway, John Law, and other thinkers, Hanson explains his concept of "joint responsibility" (ibid.), which corresponds to the extended agency as an "interaction agency". This idea is not completely new, but can already be found in a similar form in supra-individual ways of ascribing responsibility, for example, in the responsibility that we ascribe to collectives, corporations, and entire systems. The traditional understanding of responsibility against the background of a methodological and ethical individualism is based on the attribution of certain, above all cognitive, competences such as "intentionality, the capacity to act voluntarily, and awareness of the consequences of what they do" [105] (p. 93). As should have become clear in the Section 5.1 and especially Section 5.2, most robot ethicists are not prepared to attest inanimate entities in general and robots in particular to these and other competences relevant to the attribution of responsibility. If, however, according to Hanson, we imagine, for example, two people who intend to kill someone, one chasing the victim onto the street, where she or he is then run over by a car, responsibility only arises in and out of the interaction of the two people and the car. If the latter did not exist, the murder could not be carried out in this way. Not "the car has moral responsibility by itself" [105] (p. 95), but the extended agency as an apparatus (as Karen Barad would say) of 'subject-objects' has to answer for the murder.

A question that certainly follows from Hanson's position is whether the participation of a human agent is a condition for the attribution of joint responsibility. Hanson himself asserts this, since "it [the extended agency; J. L.] must include a human component because that is the locus of the will, which is necessary to intentions" [105] (p. 97). By finally attributing responsibility to essentialistically ascribable competences such as intentionality and (human) will here, Hanson's argumentation appears inconsistent and falls back into an implicit methodological individualism, which he had previously criticized of his colleagues in the debate. On the other hand, he expresses himself much more explicitly up to this point in the present paper on the attribution of responsibility in a relational (narrow) sense than many other robot ethicists. Understanding the subject of action "more as a verb than as a noun" [105] (p. 98) results in a reformulation of responsibility less as an attribute and property, but rather as a way of interaction and connection between 'subject–objects'.

6. Conclusions

Against the backdrop of Wallach's and Allen's approach of functional equivalence competences, supplemented with an algorithmic scheme, one might complement the positions of an anthropo-, patho-, bio-, and physiocentrism with a mathenocentrism and thus locate creatures with the capacity to learn within the moral universe. I assume that the ability to learn requires programming with non-determined (non-deterministic) algorithms. Such creatures would be ascribed functional morality and would possess an intrinsic value. Robots, on the other hand, which primarily function on determined (non-deterministic) algorithms, would be ascribed operational morality. They possess a high instrumental value.

If one cannot identify the necessary prerequisites for ascribing responsibility as traditionally understood, then the human parties involved (the designers and users) bear responsibility, at least as long as we claim that humans are the only genuine moral agents who possess Darwall's four types of autonomy. If one day we can identify some very complex machines as artificial agents which are ascribed functional responsibility, it would be conceivable to understand their relation to their human 'parents' as comparable to that of almost grown-up children to their parents. In the case of an accident, these exceptional functionally responsible agents might be able to partly excuse their human 'parents' from their responsibility, although not excuse them completely from bearing their responsibility. Until that day, artificial systems might be part of responsibility networks and fill the positions of objects and maybe even addressees of responsibility.

Inclusive approaches generally accuse exclusive theories, i.e., those that identify robots either as moral agents or as moral patients, of not adequately dealing with the exceptionally problematic (implicitly discriminatory, sexist, heteronormative, racist, etc.) foundations of their thinking. Thus, these first premises of their arguments are at least implicitly confirmed and perpetuated. Sometimes even the attempt is made to justify these basic assumptions as objective and universal. The exclusive positions, on the other hand, usually accuse the inclusive thinkers (such as Haraway, Hanson, Coeckelbergh, Gunkel, and Suchman) of a general blurring of their actual concerns, a darkening of the circumstances by an unnecessary and confusing softening of conceptual categories such as responsibility and species boundaries. After all, not even the decision-making authority is set over inclusion and exclusion into or out of the moral universe.

In any case, we can try to look beyond the horizon of centrist approaches that confront us with major philosophical challenges in order to establish a new practice of responsible acting–also with regard to nonhuman beings.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Matthias, A. The responsibility gap. Ascribing responsibility for the actions of learning automata. *Ethics Inf. Technol.* **2004**, *6*, 175–183. [CrossRef]
2. Sombetzki, J. *Verantwortung als Begriff, Fähigkeit, Aufgabe. Eine Drei-Ebenen-Analyse*; Springer: Wiesbaden, Germany, 2014.
3. Heidbrink, L.; Langebhn, C.; Loh, J. (Eds.) *Handbuch Verantwortung*; Springer: Wiesbaden, Germany, 2017.
4. Lenk, H.; Maring, M. Verantwortung. In *Historisches Wörterbuch der Philosophie, Band 11*; Ritter, J., Ed.; Schwabe: Basel, Switzerland, 2007; pp. 566–575.
5. Bayertz, K. Eine kurze geschichte der herkunft der verantwortung. In *Verantwortung, Prinzip oder Problem?*; Bayertz, K., Ed.; Wissenschaftliche Buchgesellschaft: Darmstadt, Germany, 1995; pp. 3–71.

6. McKeon, R. The development and the significance of the concept of responsibility. *Revenue Internationale Philosophie* **1957**, *11*, 3–32.
7. Lenk, H.; Maring, M. Wer soll verantwortung tragen? Probleme der verantwortungsverteilung in komplexen (soziotechnischen-sozioökonomischen) systemen. In *Verantwortung, Prinzip oder Problem?*; Bayertz, K., Ed.; Wissenschaftliche Buchgesellschaft: Darmstadt, Germany, 1995; pp. 241–286.
8. Loh, J. Roboterethik. *Inf. Philosophie* **2017**, *1*, 20–33.
9. Hilgendorf, E. (Ed.) *Robotik im Kontext von Recht und Moral*; Nomos: Baden-Baden, Germany, 2014.
10. Lin, P.; Abney, K.; Bekey, G. (Eds.) *Robot Ethics: The Ethical and Social Implications of Robotics*; MIT Press: Cambridge, MA, USA, 2012.
11. Lin, P.; Abney, K.; Bekey, G. Robot ethics: Mapping the issues of a mechanized world. *Artif. Intell.* **2011**, *175*, 942–949. [CrossRef]
12. Anderson, M.; Anderson, L. (Eds.) *Machine Ethics*; Cambridge University Press: Cambridge, UK, 2011.
13. Loh, J.; Loh, W. Autonomy and responsibility in hybrid systems—The example of autonomous cars. In *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*; Lin, P., Abney, K., Jenkins, R., Eds.; Oxford University Press: Oxford, UK, 2017; pp. 35–50.
14. Duff, R.A. *Responsibility. Routledge Encyclopedia of Philosophy*; Craig, E., Ed.; Routledge: London, UK, 1998; pp. 290–294.
15. Kallen, H.M. Responsibility. *Ethics* **1942**, *52*, 350–376. [CrossRef]
16. Loh, J. Strukturen und relata der verantwortung. In *Handbuch Verantwortung*; Heidbrink, L., Langbehn, C., Loh, J., Eds.; Springer: Wiesbaden, Germany, 2017; pp. 35–56.
17. Schütz, A. Einige äquivokationen im begriff der verantwortlichkeit. In *Gesammelte Aufsätze, Band 2. Studien zur soziologischen Theorie*; Schütz, A., Ed.; Nijhoff: The Hague, The Netherlands, 1972; pp. 256–258.
18. Sombetzki, J. Historische beiträge zu einer minimaldefinition von 'verantwortung'. Etymologie und genese der verantwortung vor dem hintergrund der verantwortungsforschung. *Archiv Begriffsgeschichte* **2014**, *56*, 197–219.
19. Fischer, P. *Politische Ethik. Eine Einführung*; Fink: Munich, Germany, 2006; p. 105.
20. Nunner-Winkler, G. Verantwortung. In *Lexikon der Wirtschaftsethik*; Enderle, G., Ed.; Herder: Freiburg im Breisgau, Germany, 1993; pp. 1185–1192.
21. Schlink, B. Die zukunft der verantwortung. *Merkur Deutsche Zeitschrift Europäisches Denken* **2010**, *738*, 1047–1058.
22. Waldenfels, B. Antwort und verantwortung. *Friedrich Jahresheft* **1992**, *10*, 130–132.
23. Ropohl, G. Das risiko im prinzip verantwortung. *Ethik Sozialwissenschaften Streitforum Erwägungskultur* **1994**, *5*, 109–120.
24. Zadeh, L. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [CrossRef]
25. Werner, M.H. Verantwortung. In *Handbuch Ethik*; Düwell, M., Hübenthal, C., Werner, M.H., Eds.; Metzler: Stuttgart, Germany, 2006; pp. 541–548.
26. Fauser, P. Kann die schule zur verantwortung erziehen? *Friedrich Jahresheft* **1992**, *10*, 7–9.
27. Fischer, J.M.; Ravizza, M. *Responsibility and Control: A Theory of Moral Responsibility*; Cambridge University Press: Cambridge, MA, USA, 1998; pp. 208–210.
28. Nida-Rümelin, J. Politische verantwortung. In *Staat ohne Verantwortung? Zum Wandel der Aufgaben von Staat und Politik*; Heidbrink, L., Hirsch, A., Eds.; Campus: Frankfurt am Main, Germany, 2007; pp. 55–85.
29. Witte, E. Verantwortung in erziehung und bildung. In *Handbuch Verantwortung*; Heidbrink, L., Langbehn, C., Loh, J., Eds.; Springer: Wiesbaden, Germany, 2017; pp. 667–680.
30. Neuhäuser, C. Roboter und moralische verantwortung. In *Robotik im Kontext von Recht und Moral*; Hilgendorf, E., Ed.; Nomos: Baden-Baden, Germany, 2014; pp. 269–286.
31. Schwartländer, J. Verantwortung. In *Handbuch philosophischer Grundbegriffe, Band 6, Transzendenz—Zweck*; Krings, H., Baumgartner, H.M., Wild, C., Eds.; Kösel: Munich, Germany, 1974; pp. 1577–1588.
32. Birnbacher, D. Grenzen der verantwortung. In *Verantwortung, Prinzip oder Problem?* Bayertz, K., Ed.; Wissenschaftliche Buchgesellschaft: Darmstadt, Germany, 1995; pp. 143–183.
33. Van de Poel, I. The relation between forward-looking and backward-looking responsibility. In *Moral Responsibility: Beyond Free Will and Determinism*; Vincent, N.A., van de Poel, I., van den Hoven, J., Eds.; Springer: Wiesbaden, Germany, 2001; pp. 37–52.

34. Weischedel, W. *Das Wesen der Verantwortung. Ein Versuch*; Vittorio Klostermann: Frankfurt am Main, Germany, 1972.
35. Müller, C. Verantwortungsethik. In *Geschichte der Neueren Ethik*; Pieper, A., Ed.; Francke: Tübingen, Germany, 1992; p. 114.
36. Holl, J. *Historische und Systematische Untersuchungen zum Bedingungsverhältnis von Freiheit und Verantwortlichkeit*; Forum Academicum: Heidelberg, Germany, 1980.
37. Picht, G. Der begriff der verantwortung. In *Wahrheit, Vernunft, Verantwortung. Philosophische Studien*; Picht, G., Ed.; Ernst Klett: Stuttgart, Germany, 1969; pp. 318–342.
38. Kaufmann, M. (Ed.) Die Grenzen der zurechnung. In *Zurechnung als Operationalisierung von Verantwortung*; Lang: Frankfurt am Main, Germany, 2004; pp. 288–289.
39. Honnefelder, L. *Was Soll Ich Tun, wer will Ich Sein? Vernunft und Verantwortung, Gewissen und Schuld*; Berlin University Press: Berlin, Germany, 2007.
40. Albs, B. *Verantwortung übernehmen für Handlungen und deren Folgen*; Verlag Dr. Kovač: Hamburg, Germany, 1997; p. 26.
41. Ryffel, H. Verantwortung als sittliches phänomen. Ein grundzug der moderne. *Staat Zeitschrift Staatslehre Öffentliches Recht Verfassungsgeschichte* **1967**, *6*, 275–292.
42. Steigleder, K. Deontologische theorien der verantwortung. In *Handbuch Verantwortung*; Heidbrink, L., Langbehn, C., Loh, J., Eds.; Springer: Wiesbaden, Germany, 2017; pp. 171–188.
43. Williams, G. Responsibility. Available online: http://www.iep.utm.edu/responsi/ (accessed on 4 June 2018).
44. Birnbacher, D. Teleologische ethik: Utilitarismus und verantwortung. In *Handbuch Verantwortung*; Heidbrink, L., Langbehn, C., Loh, J., Eds.; Springer: Wiesbaden, Germany, 2017; pp. 189–204.
45. Goodin, R.E. *Utilitarianism as a Public Philosophy*; Cambridge University Press: Cambridge, MA, USA, 1995.
46. Weber, M. Politik als beruf. In *Wissenschaft als Beruf: 1917/1919. Politik als Beruf: 1919*; Mommsen, W., Ed.; Mohr: Tübingen, Germany, 1992; pp. 156–252.
47. Böcher, W. *Selbstorganisation, Verantwortung, Gesellschaft. Von subatomaren Strukturen zu politischen Zukunftsvisionen*; Westdeutscher: Opladen, Germany, 1996.
48. Bierhoff, H.W. Verantwortungsbereitschaft, verantwortungsabwehr und verantwortungszuschreibung. soziopsychologische perspektiven. In *Verantwortung, Prinzip oder Problem?* Bayertz, K., Ed.; Wissenschaftliche Buchgesellschaft: Darmstadt, Germany, 1995; pp. 217–240.
49. Schmiedl-Neuburg, H. Verantwortung in der tugend- und wertethik. In *Handbuch Verantwortung*; Heidbrink, L., Langbehn, C., Loh, J., Eds.; Springer: Wiesbaden, Germany, 2017; pp. 205–220.
50. Allen, C.; Wendell, W.; Iva, S. Why machine ethics? *Intell. Syst. IEEE* **2006**, *4*, 12–17. [CrossRef]
51. Anderson, M.; Susan, L.A. Machine ethics: Creating an ethical intelligent agent. *AI Mag.* **2007**, *4*, 15–26.
52. Bendel, O. (Ed.) *Handbuch Maschinenethik*; Springer: Wiesbaden, Germany, 2018.
53. Edgar, S.L. *Morality and Machines. Perspectives on Computer Ethics*; Jones & Bartlett Learning: Boston, MA, USA, 2003.
54. Misselhorn, C. *Grundfragen der Maschinenethik*; Reclam: Stuttgart, Germany, 2018.
55. Moor, J.H. The nature, importance, and difficulty of machine ethics. *Intell. Syst. IEEE* **2006**, *4*, 18–21. [CrossRef]
56. Rath, M.; Krotz, F.; Karmasin, M. (Eds.) *Maschinenethik. Normative Grenzen Autonomer Systeme*; Springer: Wiesbaden, Germany, 2019.
57. Loh, J. *Roboterethik. Eine Einführung*; Suhrkamp: Frankfurt am Main, Germany, 2019.
58. Aristotle. *Nikomachische Ethik*; Rowohlt: Reinbek, Germany, 2006.
59. Beauchamp, T.L.; Frey, R.G. (Eds.) *The Oxford Handbook of Animal Ethics*; Oxford University Press: New York, NY, USA, 2011.
60. Cohen, C. The case for the use of animals in biomedical research. *N. Engl. J. Med.* **1986**, *14*, 865–870. [CrossRef]
61. De Gracia, D. *Taking Animals Seriously. Mental Life and Moral Status*; Cambridge University Press: New York, NY, USA, 1996.
62. Donaldson, S.; Kymlicka, W. *Zoopolis. Eine politische Theorie der Tierrechte*; Suhrkamp: Frankfurt am Main, Germany, 2013.
63. Regan, T. *The Case for Animal Rights*; University of California Press: Berkeley, CA, USA, 1983.
64. Schmitz, F. (Ed.) *Tierethik. Grundlagentexte*; Suhrkamp: Berlin, Germany, 2014.

65. Singer, P. *Animal Liberation. Die Befreiung der Tiere*; Harald Fischer: Erlange, Germany, 2015.
66. Wolf, U. *Das Tier in der Moral*; Vittorio Klostermann Verlag: Frankfurt am Main, Germany, 2012.
67. Wolf, U. (Ed.) *Texte zur Tierethik*; Reclam: Stuttgart, Germany, 2008.
68. Coccia, E. *Die Wurzeln der Welt*; Hanser: München, Germany, 2018.
69. Gunkel, D. *The Machine Question. Critical Perspectives on AI, Robots, and Ethics*; MIT Press: Cambridge, MA, USA, 2012; p. 5.
70. Asaro, P.M. What should we want from a robot ethic? *Int. Rev. Inf. Ethics* **2006**, *6*, 9–16.
71. Bekey, G. *Autonomous Robots. From Biological Inspiration to Implementation and Control*; MIT Press: Cambridge, MA, USA, 2005.
72. Brey, P.; Adam, B.; Waelbers, K. (Eds.) *Current Issues on Computing and Philosophy*; IOS Press: Amsterdam, The Netherlands, 2008.
73. Capurro, R.; Nagenborg, M. (Eds.) *Ethics and Robotics*; IOS Press: Heidelberg, Germany; Amsterdam, The Netherlands, 2009.
74. Lin, P.; Jenkins, R.; Abney, K. (Eds.) *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*; Oxford University Press: New York, NY, USA, 2007.
75. Stahl, B.C. Information, ethics, and computers: The problem of autonomous moral agents. *Minds Mach.* **2004**, *14*, 67–83. [CrossRef]
76. Sullins, J.P. When is a robot a moral agent? *Int. Rev. Inf. Ethics* **2006**, *6*, 23–30.
77. Turing, A. Computing machinery and intelligence. *Mind* **1950**, *59*, 433–460. [CrossRef]
78. Versenyi, L. Can robots be moral? *Ethics* **1974**, *3*, 248–259. [CrossRef]
79. Wallach, W.; Allen, C. *Moral Machines: Teaching Robots Right from Wrong*; MIT Press: Cambridge, MA, USA, 2009.
80. Mainzer, K. *Leben als Maschine. Von der Systembiologie zur Robotik und Künstlichen Intelligenz*; Mentis: Paderborn, Germany, 2010.
81. Misselhorn, C. Maschinenethik und 'Artificial Morality': Können und sollen Maschinen moralisch handeln? *APuZ* **2018**, *68*, 29–33.
82. Remmers, P. *Mensch-Roboter-Interaktion*; Berlin Logos: Berlin, Germany, 2018.
83. Sombetzki, J. Roboterethik. In *Zur Zukunft der Bereichsethiken. Herausforderungen durch die Ökonomisierung der Welt*; Maring, M., Ed.; KIT Scientific Publishing: Karlsruhe, Germany, 2016; pp. 355–379.
84. Canguilhem, G. Maschine und organismus. In *Die Erkenntnis des Lebens*; August Verlag: Berlin, Germany, 2012; pp. 183–232.
85. Strandh, S. *Die Maschine. Geschichte—Elemente—Funktion*; Herder: Freiburg im Breisgau, Germany, 1980.
86. Jordan, J. *Roboter, aus dem Englischen von Manfred Weltecke*; Berlin University Press: Berlin, Germany, 2017; p. 50.
87. Misselhorn, C. Robots as moral agents. In *Ethics in Science and Society: German and Japanese Views*; Rövekamp, F., Bosse, F., Eds.; Iudicium Verlag: Munich, Germany, 2013; p. 43.
88. Mataric, M.J. *The Robotics Primer*; MIT Press: Cambridge, MA, USA, 2007.
89. Asimov, I. *The Complete Robot. The Definitive Collection of Robot Stories*; Harper Collins Publications: London, UK, 1982.
90. Drux, R. (Ed.) *Menschen aus Menschenhand. Zur Geschichte der Androiden. Texte von Homer bis Asimov*; Metzler: Stuttgart, Germany, 1988.
91. Floridi, L.; Sanders, J.W. On the morality of artificial agents. *Minds Mach.* **2004**, *14*, 349–379. [CrossRef]
92. Duffy, B.R. Anthropomorphism and the social robot. *Robot. Autonom. Syst.* **2003**, *42*, 177–190. [CrossRef]
93. Gerdes, A. The issue of moral consideration in robot ethics. *ACM SIGCAS* **2017**, *45*, 247–279. [CrossRef]
94. Johnson, D.G. Computer systems. Moral entities but not moral agents. In *Machine Ethics*; Anderson, M., Leigh, S., Eds.; Cambridge University Presss: New York, NY, USA, 2011; pp. 168–183.
95. Levy, D. The ethical treatment of artificially conscious robots. *Int. J. Soc. Robot.* **2009**, *1*, 209–216. [CrossRef]
96. Sparrow, R. The Turing triage test. *Ethics Inf. Technol.* **2004**, *6*, 203–213. [CrossRef]
97. Tavani, H.T. Can social robots qualify for moral consideration? Reframing the question about robot rights. *Information* **2018**, *9*, 73. [CrossRef]
98. Darling, K. Extending Legal Protection to Social Robots. Available online: https://spectrum.ieee.org/automaton/robotics/artificial-intelligence/extending-legal-protection-to-social-robots (accessed on 3 November 2019).

99. Krebs, A. Naturethik im überblick. In *Naturethik. Grundtexte der gegenwärtigen tier- und ökoethischen Diskussion*; Krebs, A., Ed.; Suhrkamp: Frankfurt am Main, Germany, 1997; pp. 337–379.
100. Christian, G.; Hügli, A.; Kiefhaber, M.; Romberg, R.; Konersmann, R. Mensch. In *Historisches Wörterbuch der Philosophie. Band: L–Mn*; Ritter, J., Ed.; Schwabe: Basel, Switzerland, 2007; pp. 1059–1105.
101. Diels, H. *Die Fragmente der Vorsokratiker*; Weidmann: Hildesheim, Germany, 1966.
102. Nagel, T. What is it like to be a bat? *Philos. Rev.* **1974**, *4*, 435–450. [CrossRef]
103. Churchland, P.M. *Matter and Consciousness*; MIT Press: Cambridge, MA, USA, 1999.
104. Coeckelbergh, M.; Gunkel, D. Facing animals: A relational, other-oriented approach to moral standing. *J. Agric. Environ. Ethics* **2014**, *5*, 715–733. [CrossRef]
105. Hanson, F.A. Beyond the skin bag: On the moral responsibility of extended agencies. *Ethics Inf. Technol.* **2009**, *11*, 91–99. [CrossRef]
106. Haraway, D. *Staying with the Trouble: Making Kin in the Chthulucene*; Duke University Press: Durham, NC, USA, 2016.
107. Coeckelbergh, M. The moral standing of machines: Towards a relational and non-Cartesian moral hermeneutics. *Philos. Technol.* **2014**, *27*, 61–77. [CrossRef]
108. Damiano, L.; Dumouchel, P. Anthropomorphism in human-robot co-evolution. *Front. Psychol.* **2018**, *9*, 1–9. [CrossRef] [PubMed]
109. Suchman, L. *Human-Machine Reconfigurations. Plans and Situated Actions*; Cambridge University Press: Cambridge, MA, USA, 2007.
110. Nadeau, J.E. Only androids can be ethical. In *Thinking about Android Epistemology*; Ford, K., Glymour, C., Hayes, P.J., Eds.; The MIT Press: Cambridge, MA, USA, 2006.
111. Both, G.; Weber, J. Hands-free driving? Automatisiertes fahren und mensch-maschine interaktion. In *Robotik im Kontext von Recht und Moral*; Hilgendorf, E., Ed.; Nomos: Baden-Baden, Germany, 2013; pp. 171–187.
112. Hevelke, A.; Nida-Rümelin, J. Intelligente autos im dilemma. *Spektrum Wissenschaft* **2015**, *10*, 82–85.
113. Hötitzsch, S.; May, E. Rechtliche problemfelder beim einsatz automatisierter systeme im strassenverkehr. In *Robotik im Kontext von Recht und Moral*; Hilgendorf, E., Ed.; Nomos: Baden-Baden, Germany, 2014; pp. 189–210.
114. Maurer, M.; Gerdes, J.C.; Lenz, B.; Winner, H. (Eds.) *Autonomes Fahren. Technische, rechtliche und gesellschaftliche Aspekte*; Springer: Wiesbaden, Germany, 2015.
115. Anderson, M.; Anderson, L.; Armen, C. An approach to computing ethics. *Intell. Syst. IEEE* **2006**, *4*, 2–9. [CrossRef]
116. Russell, S.; Norvig, P. *Artificial Intelligence. A Modern Approach*; Prentice Hall Press: Upper Saddle River, NJ, USA, 2003.
117. Frankfurt, H. Freedom of the will and the concept of a person. *J. Philos.* **1971**, *68*, 5–20. [CrossRef]
118. Rawls, J. *Justice as fairness: A restatement*; Harvard University Press: Cambridge, MA, USA, 2001.
119. Darwall, S. The value of autonomy and autonomy of the will. *Ethics* **2006**, *116*, 263–284. [CrossRef]
120. Brooks, R.A.; Breazeal, C.; Marjanović, M.; Scasselatti, B.; Williamson, M.M. The Cog Project, building a humanoid robot. In *Computation for Metaphors. Analogy, and Agents*; Nehaniv, C., Ed.; Springer: Wiesbaden, Germany, 1999; p. 70.
121. Brooks, R.A. Intelligence without Reason. 1991. Available online: https://people.csail.mit.edu/brooks/papers/AIM-1293.pdf (accessed on 3 November 2019).
122. Wilhelms, G. Systemverantwortung. In *Handbuch Verantwortung*; Heidbrink, L., Langbehn, C., Loh, J., Eds.; Springer: Wiesbaden, Germany, 2017; pp. 501–524.
123. Varela, F.; Thompson, E.; Rosch, E. *The Embodied Mind: Cognitive Science and Human Experience*; MIT Press: Cambridge, MA, USA, 1991.