6. Conclusions ... and Openings

6.1. A Recapitulation

We have come a long way on our tour of the life cycle of robots, since our first step into the current New Age of Robotics. Before recapitulating the insights we gained along the way, let us jump back to the beginning and return to mind why we went on this journey.

We started off with the realization that robots, as well as their technological and fictional ancestors, have been a part of our world for a long time, and that today they are more in the focus of attention than ever. Robots are at the center of a hype and the subject of a controversial and emotional discourse across political, economic, academic, and other public spheres. On the one hand, robot technology is hailed as a utopian panacea, as a solution for a range of fundamental problems our society faces. On the other hand, robots are also inextricably embedded in a dystopian, at times even apocalyptical, narrative of humans losing control, of robots in competition to us, even of our very humanness being in peril.

This is rooted in the fact that robots, although they are new in our direct physical environment, are not new to us at all: Across both extremes of the discourse, technological reality is inherently intermixed with decade-, even century-old, fictional narratives. Even as robot technology is just now starting to step out of its factory cages – into our everyday lives and direct physical environment, our jobs, our households – the realm where robots are present most vividly is still our imagination. The rich cultural history of literary and cinematic robot fiction fundamentally influences our relationship with the robot technology in our real life – where the so-called New Age of Robotics brings forth technologies that appear to move closer and closer to what we know from fictional narratives. These new technological artifacts are not only used in physical proximity to us but are also embodied and mobile, act au-

tonomously, and sense and manipulate their environment. Some are even socially interactive and have a human-like design.

In this newly close relationship with robot technology, we can observe instances of humans perceiving them and treating them like more than "just" inanimate objects. Are they to be considered machines? Creatures? Something in between? Something else entirely? It seems that robots do not belong in the ontological category of the "inanimate object". But apparently they do not belong with the "living beings" either. In this, robot technology has been redrawing the attention of academia to the phenomenon of attributions of animacy to technological artifacts.

Humans have always attributed characteristics of living beings to inanimate objects – a human quirk that, historically, has often been viewed as either "primitive" behavior or, in the context of science, as methodological misconduct. With the advent of the "New Age of Robotics", however, a new type of inanimate object has been pushing into humans' lives: machines with features traditionally reserved for animate beings, technological artifacts appearing to sway between the natural and the artificial (cf. Haraway, 1991, p. 152).

This unique ontological challenge posed by robots and other autonomous technologies inspired the overarching research question of this book. Crucially, this question is not which ontological category robots "really" belong to. Rather, we followed Lucy Suchman's suggestion to shift the discussion from "whether humans and machines are the same or different to how and when the categories of human or machine become relevant, how relations of sameness or difference between them are enacted on particular occasions, and with what discursive and material consequences" (2007, p. 2).

Inspired by this, we explored the range of discursive and non-discursive manifestations of in/animacy attributions to robotics, as well as their conditions, functions, and consequences. We made four major observations:

First, attributions of animacy to robots are not an isolated phenomenon, not just a perceptual quirk of the human mind, and they are not only present in the direct, physical interaction of humans and robots. On the contrary, we found a broad range of manifestations, on a discursive, non-discursive and material level along the whole life cycle of robots: Visions of a future shared with socially interactive robot companions – both in fictional narratives and in the sociotechnical futures, which are present in the roadmaps and guidelines of the robotics industry, and of academic and policy institutions (Chapter 1). Practices of making robots "lifelike" in human-robot-interaction research

(Chapter 2). Robots being perceived as unpredictable but valuable team members in robotics research and development (Chapter 3). Complex narrative scenarios and robot personas staged for science communication, marketing, and demonstration practices (Chapter 4). And constant references to fictional narratives of robots as competition to humans in media discourse (Chapter 5).

Second, attributions of animacy to robots are not a static, inflexible practice. Instead, across all the explored contexts we found a constant switching of perspectives, of robots being perceived and represented alternatingly as inanimate objects and animate beings.

Third, these attributions of in/animacy to robots – and the switching between them – are not just an involuntary reaction to certain features of the robot, such as its design or its behavior, but in fact have context specific constructive qualities. Robot technology is embedded not only in a rich fictional and cultural history but also in a quite controversial public discourse. In this context, attributions of in/animacy help us grasp and embrace the unique challenges that robot technology brings to our lives. They are powerful in that they shape our perception of, and our relationship with intelligent technologies. In this sense, they can be a way of navigating the complex environment of our technologized society.

Fourth, the ubiquitous attributions of in/animacy to robots are persistently accompanied by a critical discourse. This criticism is predominantly aimed at the deceptive, overly opportunistic, or unthinking use of these attributions, which are feared to cause systemic, lasting, and potentially problematic consequences on policy, governance, and legislative decisions. The present chapter will discuss this in more depth (Section 6.3).

But first, Section 6.2 will take a step back to reflect on these major observations from a more cross-contextual perspective, identifying the underlying conditions, motives, and forms of in/animacy attributions permeating the whole life cycle of robots.

Section 6.4 will step even further back and break out of the robot life cycle. We will have look around in the vicinity to see how practices of in/animacy attribution and the related critical discourse, as well as the more general question of what is "natural" and what is "artificial", are discussed in our technologized society.

Finally, in Section 6.5, I will step down from the observer position and make some suggestions for how the insights of this book can be applied constructively by different actors.

6.2. The Constructive Quality of In/Animacy Attributions

Across the different contexts of robotics, we encountered attributions of animacy to robots in a variety of discursive, performative, and material forms. From playful to opportunistic, from reflective to unthinking; from images of humanoids and other comparisons with the human physiology, to references to science fiction stories in which robots have life-like characteristics, to the creation of narrative robot personas, down to the linguistic level, for example in the use of gendered pronouns. Constantly switching perspectives, to and from attributions of inanimacy, were observable on different contextual levels: In the location of presentation (website vs. social media; headline vs. article body; physical work practice vs. academic publications), in the narrative perspective (robot vs. human as narrative persona), and in the task or goal specific to the context (e.g. educating and informing vs. attracting attention). Through the whole spectrum of contexts, across the various manifestations of attribution and the different levels of perspective changes, there was one aspect we found again and again: the underlying constructive function of in/animacy attributions.

Embracing Robots' Agency

Chapter 3 found that, in the particular context of robotics research and development (R&D), it is part of a robot's job to act unpredictably. Especially in the "troubleshooting" phases of robotics R&D, when the complex system of the robot clashes with the complex physical environment, roboticists are dependent on the robot as a "feedback-giver" – the robot's behavior serving as a crucial source of information. As we observed, this can lead to a perceived distribution of agency between the roboticist and the robot – reflected in roboticists perceiving their robots as something akin to a research companion or team member. This perception is reinforced by the often very heterogeneous and interdisciplinary structure of roboticist teams: The robot platform serves as a central object of focus, taking the role of a boundary object. In this context, attributions of animacy are a reflection of the robot's central and active role. At the same time, however, roboticists are required to take a strictly "professional" (i.e. technical) perspective on their robots, focusing on them as the inanimate machines they are, and refraining from openly expressing attributions of animacy – for example when writing research papers or technical documentation, or when presenting robot demonstrators to academic

peers. The practice of constantly switching perspectives – from the robot as an animate research companion, to the robot as an inanimate object – is a constructive way of dealing with the multiple, constantly changing demands of a roboticist's professional environment.

Making Robot Behavior Explainable

In the context of robotics research and development practice (Chapter 3) we also encountered a phenomenon that earlier research in HRI and communication studies also found for the direct interaction of users and robots (cf. Chapter 2, Section 2.1): Observing a robot acting unpredictably appears to make humans think and talk about it as if it were alive. Crucially, this is not only the case for lay users, for whom the complex technical processes responsible for the robot's behavior are hidden inside a "black box". Also roboticists with professional insight and understanding of the technical details are prone to think and talk about "their" robots as if they were animate – even though they are perfectly aware that the robots are in fact inanimate objects. In practice, this is reflected in roboticists' practices of assigning names and gender to robotic platforms, in the "joking" framing of technical components as body parts, of technical processes as physiological functions, and in the attribution of liveliness and personality to robots.

In this context, in/animacy attributions are constructive in that they are a way of dealing with the strangely contradictory situation of an inanimate object acting in a way usually unique to animate beings. Allowing oneself to think and speak about a robot as a quasi-animate being makes its behavior appear somewhat more explainable (cf. Frey & Jonas, 2002). Crucially, this animate perspective is not static. Rather, human users and observers of robots appear to be able to effortlessly, playfully balance attributions of animacy and inanimacy. This allows them to keep a flexible perspective that is able to do justice to the robot's unique behavioral characteristics and its category-defying ontological status.

Making Robots Tangible

Another function of in/animacy attributions we encountered in several of the explored contexts is that of making robot technology tangible. Robots are a complex technology and, for most people, not yet a part of everyday life. This can make it difficult to imagine what the application of robot technology in

our direct environment could look and feel like. Luckily, for those trying to bring across the vision of a life with robots, while real robots are still quite exotic for most, there is a rich cultural reservoir of popular science fiction narratives to draw from. By referencing well-known fictional robot characters, science communicators and media professionals try to make robot technology imaginable and tangible for their audience (cf. Chapters 4 and 5). These fictional narratives are so popular and engaging because they deal with the topics of robotics in relation to ourselves. Classic narratives center on robot characters in juxtaposition to us "real" humans, on robots as the "other", as potential companions or competition, and on their struggle of "wanting to be like us". It is these narratives that media discourse draws from in an effort to make complex topics – such as the consequences of increasing automation on our lives – tangible to the audience. For a lay audience it is simply easier to visualize an army of humanoid robot overlords physically coming to steal their jobs from under their noses, than to disentangle difficult to grasp concepts such as "artificial intelligence", "machine learning" or "algorithms", and the systemic effects they will have on the labor market.

Even when communication efforts do not explicitly refer to science fiction narratives, comparisons with the human body, with its physiology, its sensory and emotional experiences, are ubiquitous. We find an abundance of illustrations of humanoid and human-like robots all over robotics innovation roadmaps, policy documents, research institutions' social media accounts, and newspaper articles – regardless of whether the robot technology in question is actually humanoid, or even strictly speaking a robot. Importantly, it is not the goal of these communication efforts to convince the audience that the robot technology in question is actually animate or human-like. In fact, most of the communication activities also present quite a "technical" perspective, focusing on the robot as a – clearly inanimate – machine. Wherever it fits the purpose, however, wherever the technology needs to be made tangible for the audience, we encounter a switch to attributions of animacy – and back to inanimacy. This does not only include references to the physical shape of the human body, such as illustrations of complete humanoid robots. We also find cartoons adding cute eyes to a space probe, descriptions of robots as a "he" or "she", and even complete social media accounts from the first-person perspective of a robot persona, reporting on exiting adventures, sensory experiences, and social interactions with "friends", "family", and "colleagues".

Making Robots Desirable

Across several contexts along the life cycle of robotics, we were able to observe that attributions of animacy to robots are utilized to make robots desirable – figuratively and literally.

Robots are made desirable in the figurative sense, for example in the contexts of science communication, marketing, and media discourse (cf. Chapters 4 and 5). Here, instances of animacy attribution serve to attract the attention of the audience – of potential customers, investors, and sponsors, but also of the broader public. Actors in these contexts therefore present robots in exciting interactive scenarios, embed them in narratives referring to desires and struggles relatable to the audience, in stories known from popular culture. Likewise, a headline referring to the robot revolution or a picture featuring the Terminator simply draw more attention, more readers, more clicks, to a news article than a headline or picture depicting robot technology as a "boring" and difficult to understand technical object (cf. Chapter 5).

Attributions of animacy are also employed to make robots desirable in a more literal sense. Staging robots with a quirky personality and an exciting life story, letting them have funny and adorable interactions, even letting them speak for themselves as a persona in the first person perspective, makes them engaging and likeable. In the context of science communication (cf. Chapter 4), this is utilized to draw positive attention to both robotics research and development, and to projects using robot technology to achieve scientific goals. We observed personas like Roboy, who takes his audience along to events and invites students to "hack him". We followed the travels of spacecraft and planetary rovers, who explore the universe, interact with their team of human engineers, and share information on their newest discoveries. Crucially, these narratives of animacy are "switched on and off" whenever it appears useful. In some cases, a robot's first-person-perspective Twitter account is accompanied by a website on which the robot's technical details are described with scientific distance. In other cases, social media posts "by the robot itself" take turns with posts by the engineers behind the project. Always, the goal is to convince the audience that the work done in the context of the respective projects is interesting, successful, and worth the taxes the audience might have contributed to the efforts.

In the context of commercial marketing (cf. Chapter 4) we found that attributions of animacy are switched on and off in an even more opportunistic way. Here, the goal is quite obviously to sell robots as a product. When iRobot

presents its Roomba vacuum cleaners as dutiful, dedicated, pet-like cleaning companions, it does so in order to make potential customers wish to own one. Here as well, attributions of animacy are only one half of the effort: One click further, on the company's website, the robots are presented as thoroughly inanimate products for sale, their technical details in the focus of the presentations, instead of their cute personality.

Chapter 2 (Section 2.1) discussed how human-robot interaction research (HRI) approaches the issue of animacy attributions to robots. In taking this phenomenon seriously, and in striving to use it in a constructive way, the field of HRI studies is progressing away from the historically mistrusting perspective on the phenomenon (cf. Section 2.3). Again, we were able to observe that animacy attributions to robots are viewed as having a certain function and are researched and employed with an overarching goal in mind. We can find a large number of HRI studies trying to identify and quantify features of robots (e.g. design, behavior) and users (e.g. personality, cultural background) that reliably trigger animacy attributions. Drawing heavily on research in the cognitive sciences, these studies strive to make animacy attributions controllable and predictable. The goal of these research efforts is usually to facilitate or optimize human-robot interaction and, in the long-term, to promote the introduction to interactive robot technology in society at large.

Making Robots Imaginable

Another function of animacy attributions we were able to observe is that of making robots imaginable. This can be understood on two levels: On the one hand, in the sense of making robot function imaginable for the future use of the robot; and on the other hand, in the sense of making robot technology imaginable for a future society.

In robotics demonstrations (cf. Chapter 4), be it in academic or commercial contexts, animacy attributions are employed purposefully to "prove" that a robot is functioning as claimed, now and in the future, with as little human intervention as possible. We observed that demonstrations are therefore carefully scripted, rehearsed, and – in the case of video demonstrations – also heavily edited performances. In order to make a robot appear as autonomous as possible, any "undesirable" human intervention is usually either subtly backgrounded or actively concealed. Additionally, some demonstrations embed their performance in a scenario inspired by fiction-inspired visions of a robotized future. The overall result is a performance in which the

robot appears as autonomous as possible. Sometimes, a robot is even staged as more autonomous than it actually is, leaving the audience with the impression that the robot possesses something akin to animacy. Especially non-expert audiences cannot realistically assess the current state of technology. They can therefore easily misjudge a robot's performance and overestimate its autonomy – such as in the case of Boston Dynamics' videos, which regularly go viral and have viewers express the belief that a rise of robot overlords is imminent.

The goal of proving a robot's functionality, of making it tangible and desirable not only for the present moment but also for the future, is also observable in other contexts along the life cycle of robots. In interaction studies (cf. Chapter 2, Section 2.1), in demonstrations, and in science communication (cf. Chapter 4), robots are embedded in scenarios meant to illustrate desired futures. Countless robotics research and development projects present their results in scenarios in which robot technology is advanced enough to be deployed seamlessly in everyday life. Frequently, these scenarios make use of narratives of robot animacy. They paint a picture of smooth interaction and of companionship with the robot by subtly or blatantly referencing narratives that the audience is well acquainted with from science fiction. This makes the application of robots in desired futures not only imaginable, but also paints a picture of the robot's undisputable relevance, even necessity.

Not only in science communication and demonstrations, also in media discourse (cf. Chapter 5), attributions of animacy to robots are inherently connected to visions of and predictions for our technological future. In these contexts, the future we seem to "know" from science fiction – a future populated with agentic, intentionally acting human-like robots – is treated not simply as an entertaining story, but almost as a prediction. Especially in media discourse, references to robots as quasi-animate beings serve to paint a picture of robotic inevitability and are a way of commenting on the seemingly unstoppable advance of autonomous technologies in our everyday lives. Here, too, attributions of animacy are not used consistently, but switched on and off wherever they serve their purpose. This is the reason why we find so many fact-focused technical articles, the main text body focusing solely on the technology's clearly inanimate features, accompanied by pictures of humanoid robots, by flashy headlines and punchlines referencing robot overlords, the rise of the robots, the robot revolution, robots stealing jobs – over and over.

6.3. Critical Discourse: Individual and Systemic Issues

The way robot technology is often portrayed in the media – with its flood of Terminator pictures and constant references to an inevitable robot apocalypse – faces considerable criticism. In fact, many other practices of attributing animacy to robots are discussed controversially as well. On our tour along the life cycle of robots, we were able to identify similar points of criticism in almost all contexts. The following sections will revisit and consolidate them.

The critical discourse is directed, on the one hand, at attributions of animacy as potentially being problematic on the level of the individual. Here, the focus is on whether representations of robots as quasi-animate beings constitute a form of deception, in that they create misconceptions of current and unrealistic expectations for future robot technology. On the other hand, there is also a controversial discussion of animacy attributions having long-term consequences on a more systemic level, in that these misconceptions can influence political and legislative decisions – but also in that they draw away attention from equally, if not more important, social and ethical issues in the context robot technology.

The Individual Level

One overarching point of dispute is whether actively making robots appear animate – be it though their design, their behavior, or by setting them in a certain narrative frame – is a form of deception. Even when the intention is benign, such as when a robot's humanoid design is supposed to facilitate interaction, the question remains whether it is a harmless form of manipulation or "deceit through humanization" (Butnaru, 2018; cf. Zawieska, 2015). Critical voices caution that, with increasing complexity of the technology, the "connection between input (the programmer's command) and output (how the robot behaves) will become increasingly opaque to people, and may eventually be misinterpreted as free will" (LaFrance, 2016), and that this might cause people to believe that "somebody is at home" in a robot (Scheutz, 2012, p. 3). Some interpret this as the deliberate induction of a false mental model of the robot in the user, exploiting the fact that, for non-experts, a robot can be a black box – or even "indistinguishable from magic"[1] (Clarke, 1973, p. 38).

1 The complete quote by science fiction author Arthur C. Clarke reads: "Any sufficiently advanced technology is indistinguishable from magic" (1973, p. 38).

This criticism is backed by current examples, such as a chat bot being misunderstood as being alive and sentient (Mitsuku Chatbot, 2019), or a computer generated video of a "robot uprising" being mistaken as real (Koebler, 2019).

Even when users are aware that a robot is in fact not alive, and merely attribute animacy to it on a playful, metaphorical level, this can have palpable emotional consequences. There are many examples of people grieving about "dying" robots – such as the Philae asteroid lander (Feltman, 2014; cf. Chapter 4), EOD robots destroyed on duty (e.g. J. Carpenter, 2016; Garreau, 2007), the service robot Jibo (Carman, 2019), or the robot dog Aibo (Griffin, 2015). On a practical level, attributing animacy to robots might make humans hesitant to "abuse" or "kill" robots (Bartneck et al., 2007; A. M. Rosenthal-von der Pütten et al., 2013; Sandry, 2018; James Vincent, 2019), or to deploy them into dangerous situations (J. Carpenter, 2016, p. 44; Sandry, 2015b, p. 106). The "unidirectional emotional bonds" (Scheutz, 2012) between humans and robots could also be exploited – for example by pressuring users into buying updates for their robot in order to keep it "alive". Some authors view these phenomena as expressions of an overall loss of authenticity in our technologized society, warning of illusory experiences replacing genuine relationships (Sparrow, 2002), or diagnosing a "Culture of Simulation" (Turkle, 1997), populated by machines designed in such a way that they make us "fool ourselves" (Turkle, 2011a, p. 20).

Moreover, a mental model of robots as animate – even just on a metaphorical level – sets high expectations for their physical and interactive abilities. Potential robot users have been found to have a strong expectation bias towards how robots are represented in the media. This includes the expectations that "representative robots" (T. Nomura et al., 2005, p. 125) have humanlike cognitive abilities (Kriz et al., 2010), are capable of fluent cooperative behavior (Oestreicher & Eklundh, 2006), and have a humanoid physical appearance (T. Nomura et al., 2005). It appears there is a "mismatch" or "conflict between the expectations of the users (that are primarily shaped by movies and fiction), the goals of HRI research, and the needs of the users" (Sandoval, Mubin, & Obaid, 2014, p. 61). When confronted with the actual current state of robot technology – which cannot yet hold up to these high expectations – disappointment can be the consequence. For example, many customers who bought the humanoid Pepper – advertised as a highly interactive entertainment and customer service robot – were so disappointed by its performance that they "fired" their Pepper robot (e.g. Forrest, 2018). Pepper is not the only example of misleading forms of animacy attribution having tangible economic conse-

quences: Robotics and AI businesses have been accused of fostering a "bullshit-industrial complex" (Mallazzo, 2019), deceiving investors and investors into "[throwing] disproportionate amounts of money [at] business ideas that are flat-out unfeasible and incorrectly ambitious" (Montani, 2017).

Critics warn that the narratives and practices we explored in the previous chapters – routinely treating robot technology as quasi-animate beings over and over – push certain ideas about the role of robotics in our future: The idea that robots inevitably will play a crucial role at all (cf. Bischof, 2017a, p. 137); the idea that robots will reach a certain sophistication within a certain time span, like the prediction that by 2050 a team of robots will be capable of winning the human soccer World Cup (Robocup.org, n.d.); the idea that robots will be malicious, even destroy human life – which, a study claims, is held by more than two thirds of UK adults (Business Wire, 2017); and, crucially, the idea that those robots will have a humanoid form (The Royal Society, 2018). Altogether, critics warn, these biased expectations might "affect public confidence and perceptions [and] contribute to misinformed debate" (The Royal Society, 2018, p. 4).

Another concern is that attributions of animacy to robots propagate misconceptions about the current state of robot technology by backgrounding the contribution of human actors to the actions of robots, while at the same time inflating robots' ability to act autonomously. This effectively ignores, or even negates, the complexity and social thickness of the construction of technological systems (Jasanoff & Kim, 2015, p. 2; cf. Bijker, Hughes, & Pinch, 1987; R. Williams & Edge, 1996). There are numerous examples: Science communicators staging space probes as autonomously acting "explorers", thereby taking away well-deserved credit from human scientists and engineers behind the mission (cf. Clancey, 2006); reports framing robots as the perpetrator of accidents, shifting away the blame from the human who actually made a mistake in the programming or control of the robots; reports blaming a medical service robot for delivering news of a terminal illness to a patient, shifting away the blame from the doctor who made the decision to convey the news via the remote-controlled telepresence robot (cf. Becker, 2019); or the news embedding increased automation in a narrative of "robots are coming to take away our jobs", thus not only omitting that it is humans who make the decision to replace human workers with technology, but also fostering the idea of an inevitable robotized future, which humans only can await passively and helplessly (cf. Merchant, 2019).

The Systemic Level

A worry that pervades all these points of criticism is that misconceptions about the current state of the art in robotics, as well as biased expectations about the future of robot technology, may not only have an impact on how individuals perceive and interact with robot technology. Policy makers and academic experts, too, fall for the biased representations and science fiction-inspired narratives in science communication and the media. Critical voices thus warn that even far-reaching policy decisions are at risk of being made based on misconceptions about robot technology.

Indeed, practices of animacy attribution to robotics can be found in political contexts as well. Obvious at first glance is a strong propensity for using humanoid robots to visualize not only robotics topics, but also neighboring areas such as artificial intelligence. For example, the EU Parliament does not only use the expression "rise of the robots" on their news website (e.g. European Parliament News, 2016), it also features pictures of fictional androids in marketing materials, like those for a hearing on the legal and ethical aspects of robotics and artificial intelligence (see Figure 10). The U.S. Department of Defense features cute humanoid cartoon robots in the logo for their Algorithmic Warfare Cross-Functional Team (see Figure 10) – whose work does in fact not focus on robots, but on computer vision (Pellerin, 2017; G. L. Scott, 2018). And the German Bundestag's Enquete Commission on Artificial Intelligence uses pictures of humanoid robots to illustrate news articles on all kinds of AI topics on their website (see Figure 10).

A strong humanoid bias can also be observed in the context of political events. Especially the practice of featuring commercial humanoid robots as "guest speakers", staging them as sentient, autonomously acting beings, has drawn considerable criticism. For example, a Pepper robot was "invited" to "speak" in the UK parliament as a "witness expert" for robotics and AI topics (UK Parliament, 2018; see Figure 11). The event was quickly criticized by the robotics community as a publicity stunt and even potentially illegal practice (Bryson, 2018a; Volpicelli, 2018).

Hanson Robotics' Sophia robot is probably the most famous – or rather, infamous – robot in the political arena. The robot has been staged as a "speaker" at various political events – among them the 2018 Munich Security Conference (see Figure 11) and several United Nations conferences (ECOSOC, 2017; UNDP, 2018). Sophia has also been at the center of several marketing stunts in the political context: She was named the United Nations Devel-

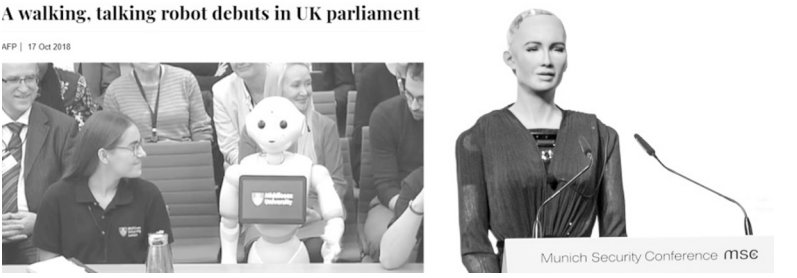
opment Programme's first-ever non-human Innovation Champion (UNDP, 2017), was made a honorary citizen of Saudi Arabia (Sini, 2017), and was issued an Azerbaijani visa (Armstrong, 2018).

Figure 10: Top left – Poster for a Hearing on Legal and Ethical Aspects of Robotics and Artificial Intelligence in the European Parliament (2016). Top right – News article on the logo of the US Department of Defense's Algorithmic Warfare Cross-Functional Team (2018). Bottom – News posts illustrated with humanoid robots on the website of the German Bundestag's Enquete-Commission on Aritificial Intelligence (2019).

Sources: http://www.europarl.europa.eu/committees/en/juri/events-hearings.html?id=20160421CHE00181 (top left, accessed on 2019-10-12) | https://www.inverse.com/article/45423-project-maven-logo-department-of-defense-google (top right, screenshot taken on 2019-12-07) | https://www.bundestag.de/ausschuesse/weitere_gremien/enquete_ki (bottom, screenshots taken on 2019-10-12).

At all of these, and the many other events where robots "spoke", it was never clearly disclosed who authored the robots' statements and to which extent the companies providing the robots were involved (Cuthbertson, 2018). Instead, the robots were presented as if they were animate and speaking for themselves.

Figure 11: Left – Newspaper article about a Pepper Robot in the UK Parliament (2018). Right – Robot "Sophia" at the 2018 Munich Security Conference.



Sources: https://www.dawn.com/news/1439541 (left, screenshot taken on December 7, 2019). | https://commons.wikimedia.org/wiki/File:Wolfgang_Ischinger_mit_Roboter_Sophia_MSC_2018.jpg (right, accessed 2019-10-12). Author: MSC/Kuhlmann. Image available under the CC BY 3.0 DE license (https://creativecommons.org/licenses/by/3.0/de/deed.en).

The way these commercial humanoid robots are "paraded around" at political events draws consistent criticism, especially from the robot and AI ethics community. Critics fear that marketing stunts like Sophia's performances, by presenting a biased and distorted image of the current state of technology, make it difficult for government and policy actors to ground their decision on sound facts (e.g. Sharkey, 2018). Fernaeus and colleagues (2009) even warned of a "robot cargo cult" (cf. Feynman, 1974), in which unproven ideas are presented as facts. Crucially, misconceptions about the current state of the art caused by this kind of robot "marketing" are not limited to non-experts. Also funding decisions for robotics research and development, made by reviewers who should be aware of the current state of technology, are sometimes heavily influenced by the ubiquity of narratives of animate (appearing) robots in public discourse. Robotics professor Tony Belpaeme (2018) reported that "an EU project reviewer express[ed] disappointment in [Belpame's team's] slow

research progress, as the Sophia bot clearly showed that the technical challenges [they] were still struggling with were solved" – the reviewer apparently having fallen for Hanson Robotics' well-staged demonstrations.

Attributions of animacy to robots are not only discussed critically for their potential to cause misconceptions about current technology. Another major area of concern is that systematically biased representation of robots as animate "may sustain and trigger unrealistic visions", that "not only the general public, but also researchers may maintain an unrealistic, even fantasy-based, perspective of what robots are and could be" (Fernaeus et al., 2009, p. 279), and that not making it clear that even a human-like robot's behavior is controlled by humans "might lead us to design legislation based on the form of a robot, and not the function, … a grave mistake" (Richards & Smart, 2013, p. 21).

In the context of political discourse, including the discourse surrounding funding initiatives for robot technology development, as well as robotics legislation, references to robot animacy are not only observable in the ubiquity of humanoid robot illustrations and marketing stunts. "Science fiction and fantasy are increasingly invoked by policy elites in service of arguments about the real world" (C. Carpenter, 2016, p. 53), serving as either guiding visions or deterrent scenarios. As discussed in Chapter 1, research agendas and roadmaps for robotics innovation, for example by the European Commission, often draw motivation and justification from science fiction-inspired visions of the future, featuring scenarios of interactive, even human-inspired, robot companions and coworkers.

However, not only utopian science fiction narratives can be observed in political discourse. Dystopian scenarios, such as the Terminator movies, are frequently exploited in the controversial discussion of armed conflicts and the development of autonomous weapons (Sharkey, 2018). These narratives are kindled by prominent activists of the anti-autonomous weapons movement, such as Stephen Hawking or Elon Musk, "to signal other broadly recognized meanings, such as the perceived potential impending crisis of an enormous magnitude if these systems are widely used" (J. Carpenter, 2016, p. 24; cf. Gibbs, 2014; Mick, 2014). Critical voices caution that this could "mislead the public on the actual dangers of artificial intelligence" (Shead, 2019). References to science fiction also fall on fertile ground in the military community: A representative of the US Pentagon explicitly stated he is concerned about robots becoming like "a Terminator without a conscience" (Silver, 2016).

Criticism of a misguided use of science fiction narratives in the context of legislation is especially present in the discourse around the legal status of robots. In discussions on who should carry the legal responsibility for accidents caused by robots, Isaac Asimov's (1950) Three Laws of Robotics[2] are almost routinely used as a base for discussion, or even as an explicit model for robot legislation (Murphy & Woods, 2009). The European Parliament's "Resolution on Civil Law Rules of Robotics" makes explicit reference to Asimov's Three Laws having to be upheld (2017, p. 4). Critics caution against using a literary plot device as a basis for legislative decisions. The Three Laws are, after all, formulated deliberately vague so they can be broken to drive the story forward (P. W. Singer, 2009, p. 520). Additionally, Asimov's Laws are – at least as of yet – technologically impossible to "install" in a robot (ibid.). For the robotics community itself, they are "little more than an imaginative literary device" (McCauley, 2007, p. 159). Even a study commissioned by the European Parliament itself criticized the explicit references to the Three Laws in a EU policy document (2016, p. 12) and noted that "when we consider civil liability in robotics, we come up against fanciful visions about robots. Here we must resist calls to establish a legal personality based on science fiction" (ibid., p. 5). The UNESCO World Commission on the Ethics of Scientific Knowledge and Technology warns against "call[ing obots] 'persons' as long as they do not possess some additional qualities typically associated with human persons, such as freedom of will, intentionality, self-consciousness, moral agency or a sense of personal identity" (COMEST, 2017, p. 46). An Open Letter signed by "Artificial Intelligence and Robotics experts, industry leaders, law, medical and ethics experts" criticizes the "bias based on an overvaluation of the actual capabilities of even the most advanced robots … and a robot perception distorted by Science-Fiction and a few recent sensational press announcements" (Robotics-Openletter.eu, n.d.).

The focus of public, political, and legislative discourse on a narrative of robots as futuristic, animate-appearing, and humanlike – be it inspired by

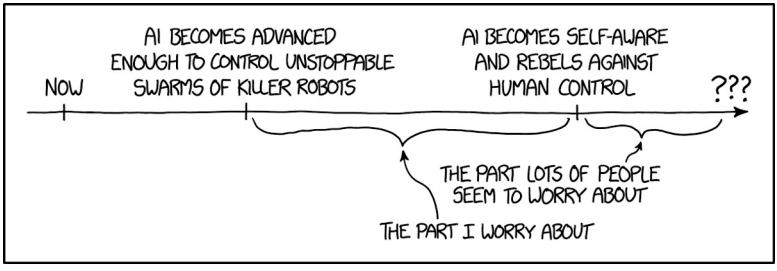
2 The Three Laws of Robotics: "(1) A robot may not injure a human being or, through inaction, allow a human being to come to harm. (2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. (3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws and (0) A robot may not harm humanity, or, by inaction, allow humanity to come to harm" (Asimov, 1950).

science fiction or by other representations of robots – has another problematic consequence. Critics warn that it shifts away attention from other equally, if not more, important aspects of robotics and automation. In this context, Chihyung Jeon (2016) described robot technology as a "technofuturistic escape": By promoting idealized scenarios of a future in which today's pressing problems have been solved by robot technology, policy-makers are able to evade having to address current problems. The ubiquitous references to a dystopian robot future are criticized for deflecting attention from the fact that robots, but also other non-embodied autonomous technologies, are already inherently embedded in, and have impact on, our current lives:

> "For all the fears of world where robots rule with an iron fist, we already live in a world where machines rule humanity in another way. … We're embedded in a matrix of technology that increasingly shapes how we live, work, communicate, and now fight." (P. W. Singer, 2009, p. 515)

A 2018 cartoon commented on the apparent lack of interest in short- and medium term consequences of artificial intelligence and robotics by creating an imaginary timeline of the infamous AI apocalypse (see Figure 12).

Figure 12: Cartoon "Robot Future" (XKCD, 2018).



Source: https://xkcd.com/1968 (accessed on 2019-10-13). Image used in accordance with the artist's guidelines (https://xkcd.com/license.html).

The prominent narrative of "robots are coming to take away our jobs" is criticized not only for omitting that it is humans who make the decision to replace human workers with technology, but also for fostering the idea of an inevitable robotized future that humans only can await passively and helplessly (cf. Merchant, 2019). Moreover, the constant discourse on robots as au-

tonomous and agentic is feared to leave the impression that it is they who are responsible for developments such as technological unemployment: "It is easier and more compelling to imagine humanoid robots than to consider the evolution of the consequences for business models, organizations and labour" (Craig, 2019, p. 40). In reality, of course, it is humans who make the decision to automate traditionally human tasks. In other words: "'Robots' Are Not 'Coming for Your Job'—Management Is" (Merchant, 2019). Neither are accidents caused by robots the fault of the robot individual. After all, "robots are simply tools of various kinds, albeit very special tools, and the responsibility of making sure they behave well must always lie with human beings" (Boden et al., 2017, p. 125).

Of course, it is easier to focus on tangible technologies, such as humanoid robots, than on more abstract concepts, such as "algorithms", "big data", or "machine learning". However, it is these complex, non-embodied technologies, which already play an important role in our lives: "There has been too much talk about interesting but irrelevant future questions, and not enough about harder current ones" (Mulgan, cited in Highfield, 2019). This sentiment is shared by many commenters from the AI and robot ethics community:

> "The 'robot invasion' is not something that will transpire as we have imagined it in our science fiction, with a marauding army of evil-minded androids either descending from the heavens or rising up in revolt against their human masters. It is an already occurring event with machines of various configurations and capabilities coming to take up position in our world through a slow but steady incursion." (Gunkel, 2020, p. 1)

6.4. In/Animacy: Beyond Robotics

In the larger context of our current technologized society, among the many perspectives one can choose to explore how technology influences our private and professional lives, robots are an especially tangible and engaging, often even spectacular example. With their long cultural history and their shining roles in fictional narratives, robots are a constant presence. This is only heightened by the current "robotics hype", which places robots at the center of not only significant economic developments but also academic and public discourses. The underlying issues, however, which drive our fascination with robots and feed the ongoing discourses, are not necessarily unique to

the topic of robot technology. Some of the more specific aspects observed in this book, such as the constructive function of in/animacy attributions, can be found mirrored in other contexts of our technologized society as well.

As discussed in Chapters 1 and 5, not only is there no professional consensus on what actually "counts" as a robot, also public discourse tends to group robots with what is perceived as neighboring technologies. This includes whole fields, such as artificial intelligence or machine learning, but also specific pieces of technology, such as autonomous cars, drones, or smart home appliances. In the context of these technological fields and artifacts we can observe a similar juggling of ontological categories and, moreover, a thematic overlap with the public and critical discourse on robot technology.

Many of the observations we made specifically for robotics can be transferred to the context of artificial intelligence – including much of the critical discourse (Kurenkov, 2019; Marcus, 2013; Schwartz, 2018; Togelius, 2017). Users of virtual assistants like Alexa[3] or Siri[4] can develop emotional connections to the software personas (Aronson & Duportail, 2018), going so far that they can be more open and willing to disclose personal feelings to virtual humans, compared to real humans (Lucas et al., 2014). Outside the context of deliberately socially interactive AI, we can observe further practices of personification. Typical phrases with umbrella terms insinuating cognitive processes, such as "teaching an AI to do something" or "the AI thinks that", reveal practices of animacy attribution. The backgrounding of human involvement, too, is an issue in the context of artificial intelligence, for example in the case of "pseudo-AI". There are reports of companies charging customers for the services of "AI assistants" – which are in fact nothing but human workers pretending to be the AI by communicating in a "robotic" style (e.g. Shane, 2018; Solon, 2018). Newspapers were accused of "faking" after they published an article presumably written by a neural net, while in fact human journalists were involved in the editing process (e.g. Seabrook, 2019; Lowndes, 2020). We can also observe cases of agency and intentionality being attributed to algorithms, in order to shift away the blame for questionable practices from human developers and management – such as the discussion about a credit algorithm "deciding" that female customers were less credit-worthy (Heinemeier Hansson, 2019; Vigdor, 2019).

3 https://developer.amazon.com/en-US/alexa (accessed 2020-01-08).

4 https://www.apple.com/siri (accessed 2020-01-08).

Outside the realm of autonomous technologies we find further examples of similarly constructive attributive practices scraping, even crossing the borders of ontological categories like "human" and "nonhuman", "artificial" and natural".

Lucy Suchman (2011), for example, discussed the concept of the model organism – meaning any nonhuman species serving as a biological research platform, with the crucial expectation that scientific insights made in the model organism can in some way be transferred from this organism to another one (cf. Fields & Johnston, 2005). Typical model organisms are the *E. coli* bacterium, laboratory rats and mice, or the common fruit fly *Drosophila*. In the work of researchers like Robert Kohler (1994) or Lynda Birke and colleagues (2004) we can find several parallels to the case of robot technology. In their unique role as models for other organisms, these species are an example for human characteristics being mapped onto nonhuman entities, in that – at least in public discourse – the crucial differences between the animal and human organism is played down. Preclinical research on mice and rats is frequently "hyped" when making its way into public discourse, reports misleadingly making the results appear directly transferable to human organisms (Heathers, 2019). Model organisms are also an example for living entities being reconstructed, at the same time, as research tools and as active participants in the knowledge making process. In a similar vein, Karin Knorr-Cetina (1997, 1998) described how cytogeneticist Barbara McClintock perceived herself to be "among the chromosomes" during her work, and how "she not only identif[ied] with them, she enter[ed] their environment, in which she bec[ame] situated as 'one of them'" (Knorr-Cetina, 1997, p. 24). Here, similarly to what we observed for the case of robots, the practice of identifying with the chromosomes, indirectly giving the objects of research the status of a persona, was constructive for McClintock's work process.

On a more general and abstract level, we can observe more parallels to the practices and discourses discussed in this book. The issue of ontological categorizations at the border of the "technical" and the "natural" is a central point of discussion in the context of how we perceive and represent robot technology. In the case of robotics, the central question is whether an objectively technical, inanimate object shares enough characteristics with "traditionally animate" entities to be sorted in the same ontological category, or merits the creation of a completely new category, which would effectively break the traditional dichotomy of "animate" and "inanimate".

In other contexts, we can find similar ontological questions. One example is the case of so-called biofacts – artificially created biotic artifacts, such as cloned animals, artificially grown body tissues, or genetically modified fruit. Biofacts are (or were at a certain point) phenomenologically animate[5], but their development and growth processes are technologically controlled (Gill, Torma, & Zachmann, 2018). In the case of robot technology, the artificial-technological aspects of the artifact are relatively obvious on a phenomenological level. With the exception of extremely realistic androids, robots are usually identifiable technological artifacts at first glance. In the case of biofacts, on the other hand, their artificiality and the technological influence that shaped them are invisible – sometimes even down to the molecular level (Karafyllis, 2003). Both robot technology and biofacts, however, are situated at or on the border of the natural and the technological, making their assignment to traditional ontological categories difficult or even impossible. As discussed in the introductory chapter (Section 1.1), robots are sometimes assigned to a completely new category, somewhere between "animate" and "inanimate". Likewise, biofacts are discussed to be a new category for themselves, ontologically located between "artifact" and "animate being" (Karafyllis, 2003, p. 16).

Robots are not the only technology that, in the words of Donna Haraway, "has made thoroughly ambiguous the difference between natural and artificial" (1991, p. 152). However, robots are a highly instructive case to observe how our technologized society constantly forces us to question seemingly long-established ontological boundaries. In this sense, the constant switching of attributions we observed across all contexts along the life cycle of robots also is a manifestation of the negotiation of these ambiguous boundaries between us and technology, and a negotiation of how much control and closeness we are willing to allow technology – both on a practical and on an emotional level.

6.5. Speaking Clearly: A Take-Home Message

Over the course of this book, on our long and winding journey along the life cycle of robots, we encountered different practices of talking and thinking about robots – these peculiar machine-beings that seemingly only recently

5 Here, the notion of animacy is associated with the existence of biological growth processes.

stepped from factories and fiction into in our physical lives. We found that all kinds of people – from experts who work closely with state-of-the-art robot technology, to lay people who only know robots from science fiction movies – routinely and effortlessly balance and play with robots' complex and confusing ontological status. It appears that humans are able to see robots as both, inanimate machines and animate beings, and able to express both perspectives in the way they talk about and interact with robots, without it feeling contradictory. We saw that the practice of balancing and playing with robots' in/animacy has crucial, constructive, and useful functions. However, we also found that this practice – if practiced unthinkingly or too opportunistically – is perceived as causing problematic consequences.

At the end of this book, let me take a step back from the position of the scientific observer and consider how some insights of my research can be applied constructively to our current and future lives in a robotized society.

The first insight is that attributions of animacy to robots – and to other autonomous (appearing) technologies – are a ubiquitous, persistent, and very old phenomenon. It is deeply ingrained in our cognitive-perceptual system and, in all likelihood, we will not be able to "stop it". In other words: "Just telling people not anthropomorphize robots won't work" (De Graaf, 2017). In fact, we do not have to stop: As we saw, we are doing an excellent job at juggling, on the one hand, our rational knowledge about the inanimacy of a robot and, on the other hand, the playful metaphorical attributions that help us interact with robots and communicate with other humans about robots in a meaningful way. For us, "these seemingly contradictory features – a thing and a living creature – [can] unproblematically coexist" (Alač, 2016, p. 12). Our ability to switch effortlessly between different ontological perspectives on technology serves us well in the technologized society we live in. In this sense, attributions of in/animacy are a cognitive, practical, and discursive tool that helps us make sense of complex autonomous technologies and the different contexts where we encounter them. Maybe we can "cut each other, and ourselves, some intellectual slack when it comes to [these] familiar, relatively benign, kinds of self-indulgence … [they] can co-exist with ordinary honesty and commitment to truth" (Blackford, 2012, p. 50).

This commitment to truth is crucial, however. Knowing now how powerful even "only" metaphorical attributions of animacy can be, which very concrete consequences they can have – such as when legislative decisions are based on them – we should take them seriously. For those of us who talk about, who write about, who present robot technology, this does not mean that we

should not use those playful metaphors and references to science fiction. As we learned, they are inevitable in a way, they are useful – and they are fun. That is, they are as long they are applied mindfully and responsibly.

We, the readers and the writer of this book, are now experts for the issue of in/animacy attributions to robots. But not everyone is. Not everyone is able to assess whether an extremely realistically behaving humanoid robot is remote controlled or whether it has "real" intentionality, intelligence, and animacy. Not everyone knows that "killer robots" do not look like the Terminator. It is for these people we need to make an effort not to let opportunism turn playful attributions of animacy into deception. What does this mean in practice?

On a fundamental level, it means being aware that technology and society are always entangled. It means knowing that robotics, like any other technology, can never be "neutral" and unbiased, as its production is always inherently connected to its societal context (e.g. Jasanoff, 2004).

On a practical level, it means accurately describing the current state of robot technology; clarifying what is fact and what is fiction, and separating the present from an imaginary future; refraining from making non-experts believe that a robot is more intelligent and autonomous than it actually is; being upfront about the influence of humans on robot's behavior, and about the limitations of a robot's capabilities (Kurenkov, 2019). It also means using realistic pictures as illustrations for articles – or at the very least, providing explanatory image captions; clarifying that the technology described in an article does not look like the humanoid robot in the picture, but that the picture is from a science fiction movie.

More generally, it also means not letting the "exoticness" of robot technology distract us from other equally, or even more, pressing issues of technological innovation. So, instead of only wondering whether Boston Dynamics' robots will bring about the robot apocalypse, let us also talk about them being funded by the Department of Defense, about the use of their robots for surveillance and in law enforcement (Cuthbertson, 2019; Schwartz, 2018; Sullivan, Jackman, & Fung, 2016). While worrying about the dangers of Terminator-like autonomous weapons, let us also consider the dangers of embedded forms of intelligence in "smart homes" and "smart cities" (Craig, 2019, p. 40; König, 2019). And when following the exciting adventures of cute space robots, let us also consider the ethical, environmental, and economic consequences of space exploration and planetary exploitation (L. Wright, 2016).

Overall, it means taking a pragmatic approach to existing and future technologies – robotic and otherwise (cf. von Gehlen, 2018). To promote one's own

and others' technology literacy[6] by being open for emerging technologies, but also to keep in mind their limitations (cf. Renn, 2011).

And finally, it means to not let science fiction make us think that one or the other robotic future is inevitable, that the machines are in control, but to be aware of our power to influence the presence of robot technology in our present and future lives.

> "The machine is not an it to be animated, worshipped, and dominated. The machine is us, our process, an aspect of our embodiment. We can be responsible for machines; they do not dominate or threaten us. We are responsible for boundaries; we are they." (Donna Haraway, "A Cyborg Manifesto", 1991, p. 81)[7]

6 A loose translation of the German term "Technikmündigkeit", "Mündigkeit" meaning maturity or majority. Hat tip to Ilja Sperling for suggesting this translation.

7 Hat tip to Beth Singler (2019) for using this quote in a blog post.