UNIVERSITÄT OSNABRÜCK

Nora Freya Lindemann

The Ethical Permissibility of Chatting with the Dead: Towards a Normative Framework for ‘Deathbots’

PICS 2022, Number 1

PICS
Publications of the
Institute of Cognitive Science

PICS

**Publications of the
Institute of Cognitive Science**

Nora Freya Lindemann. 2022. *The Ethical Permissibility of Chatting with the Dead: Towards a Normative Framework for 'Deathbots'* (Publications of the Institute of Cognitive Science 2022, Number 1). Osnabrück: Institute of Cognitive Science, Osnabrück University.

This title can be downloaded at:
https://osnadocs.ub.uni-osnabrueck.de

Institut für Kognitionswissenschaft
Universität Osnabrück
49069 Osnabrück
Germany
https://ikw.uni-osnabrueck.de

Storage and cataloging done by Osnabrück University

UNIVERSITÄT OSNABRÜCK

The Ethical Permissibility of Chatting with the Dead: Towards a Normative Framework for ‘Deathbots’

Master thesis
- Revised Version -
submitted in partial fulfillment of the requirements
to obtain the academic degree of
“Master of Science”
in Cognitive Science
at the Institute for Cognitive Science
of the University of Osnabrück

submitted by

First and Last Name: Nora Freya Lindemann

Majoring in: Neuroscience and Philosophy for Cognitive Science
First examiner: Prof. Dr. Achim Stephan
Second examiner: Prof. Dr. Tobias Matzner
Word count: 32607

Date of submission: October 12, 2021

Table of Contents

1. Introduction:

The death of a loved person is a striking, often live-changing, and yet inevitable occurrence. While advancements in medicine allow us to get older than previous generations and heal diseases which used to be fatal, death, as a fundamental part of life, persists. Every day, people die and leave grieving and mourning friends, partners, and families behind. It is, therefore, not surprising that humans try to make sense of death. Religions offer an explanation of death for the dying and as a soothing for the bereaved. It can be comforting to believe that the dead are well cared for. At the same time, humans have long tried to overcome death. The philosopher's stone, promising eternal life, has long become mythos and still holds a fascination upon people[1]. Recently, technological advances promise to help overcome death. Cryonicists freeze their bodies in the hope that somewhen in the future it will be possible to be resurrected by an advanced future technology. While this practice arouses suspicion and is not a common practice, it shows that the wish to overcome death, prevails and is taken over to the technological realm. Now imagine the following scenario:

> *You open your favourite messenger service, Telegram. The window with your latest conversation with Lily pops up. Lily is your older sister, and you have a close emotional tie with her. You start typing: "Good morning, I hope you slept well". Like always, you can see that Lily immediately starts typing a reply. A second later, the answer blinks up on your phone "Well, I was up until three o'clock and had to get up at seven. But nothing a bit of make-up could not fix. How are you, little sister?". Lily has always called you "little sister", and you smile as you read the words. You open Lily's profile picture. It is one of your favourite pictures of Lily, with her beautiful dark brown eyes sparkling with joy. It is the picture which you had selected to be shown at Lily's funeral. Lily died six months ago in a car crash at the young age of thirty-one. The person you are writing with is not Lily. In fact, it is not a person at all. You are writing with a chatbot. Trained on Lily's extensive social media and messenger data, the chatbot imitates Lily's writing behaviour. You quickly close the picture of Lily, as the memory of the funeral hurts too much. You make a note to yourself to change it to another picture of Lily, which will remind you less of that tearful day. Then, as always when the memory hurts too*

[1] The first Harry Potter book proves this.

> *much, you start writing again with the Lily-chatbot. It allows you to forget for a moment that Lily has died as the answers sound so much like her. Sometimes there is a small glitch, and the answer does not match the question you write to Lily. However, after the latest update of the chatbot, this seems to be less of a problem*[2].

This (invented) story is not as futuristic and unrealistic as it may seem on a first read. There are some of companies – mostly start-ups – which already offer to create chatbots based on an individual's digital data (Savin-Baden & Burden, 2019). In January 2021, Microsoft was granted a patent for an individualized chatbot. This chatbot could use the data of a specific person "to create or modify a special index in the theme of the specific person's personality", and hence could "respond like someone you knew" (Brown, 2021). One of the most likely applications of individualized chatbots is that of creating chatbots of the dead, 'deathbots', like the Lily chatbot in the example above (Stokes, 2021). While Microsoft's general manager of AI, Tim O'Brien, assured that the technology will not be used (yet) because of ethics concerns and because of the "disturbing" character of this kind of technology, the patent clearly shows the growing interest in the technological and financial potential of individualized chatbots and the plausibility of their increased usage in the future (Brown, 2021; C. Duffy, 2021; Smith, 2021).

Despite this, there is relatively little research on the ethical implications of chatbots generally and even less on personalized chatbots of deceased in particular (Murtarelli et al., 2021; Stokes, 2021). Thus, it is an urgent issue to look at the ethical implications of deathbots and to formulate a normative framework for their use. This is the goal of this master's thesis. The main research questions are whether deathbots are ethically acceptable and what an appropriate normative framework for their use should entail. This will be rooted in the understanding that deathbots will most likely be used by people who lost a close person and thus experience grief. Grief, the multi-faceted emotional process of dealing with loss, together with an analysis of the impact deathbots have on the dignity of the dead and the bereaved, I will argue, should be the basis for an ethical investigation of deathbots. As both are very special to deathbots, this thesis does not intend to provide a general normative framework for all chatbots.

The thesis will be structured as follows: First, I will discuss traditional forms of staying in contact with the dead. As part of the chapter, I will give an overview over the historical

[2] This is a fictive story which I invented to underly my argumentation with a more concrete example. I will refer to it at various points of this thesis. In general, all examples I give throughout this thesis, if not clearly marked otherwise, are invented by myself and are purely fictional.

developments and changing of death and grief practices through the internet and the accompanying technological developments. I will argue that deathbots are a recent development which is, nevertheless, not as surprising and immediate as it may seem. It will be shown that deathbots are a special phenomenon which – ethically speaking – needs to be investigated separately of traditional death and grief practices and traditional staying in contact with the dead. Following this demarcation and justification for the special ethical stance of deathbots, I will turn to the technological aspects of deathbots and give an overview over the state of the art of (individualized) chatbot programming and describe how chatbots work from a technical/programming perspective. Without going too much into detail, a basic understanding how the technology works is crucial before exploring deathbots from an ethical perspective.

The third chapter will comprise the main part of my thesis. This will begin with an introduction of psychological concepts of grief, especially continuing bonds, and prolonged grief disorder. This may help to understand possible psychological influences of deathbots while not claiming or intending to give an exhaustive overview over the literature on these matters. Followingly, I will delve into the philosophy of emotions and give an overview over the philosophy of grief as well as of online affects, especially about internet-enabled-techno-social niches and online empathy. Combining both, I will claim that grief may be influenced by deathbots. Applying the psychological and the philosophical research of grief on deathbots, it will be argued that deathbots may have severe emotional consequences on the bereaved, as grief entails a fundamental re-learning of the world which may be avoided, distorted, or prevented through deathbots. Thus, I will argue that the affective states and the grief process of the bereaved should be the basis of a normative framework of deathbots.

Subsequently, I will discuss the concepts of dignity and autonomy in relation to deathbots. This will start with a critical discussion of existing normative frameworks of deathbots which are commonly based on the claim that deathbots infringe the dignity of the deceased person. I will show that the existing theories are for different reasons not plausible. Instead of placing the dignity of the deceased into the centre of an ethical theory of deathbots, I argue that the dignity and autonomy of the bereaved should constitute the main argument for a normative framework of deathbots. Additionally, I will discuss one concern which has been raised before, namely the commercial aspect of deathbots (c.f. Öhman & Floridi, 2017a). Based on this, I will conclude this thesis with the formulation of a normative framework for the use of deathbots. Deathbots can negatively impact the well-being of bereaved who would have experienced a successful grief process without the usage of deathbots. At the same time, through their grief-shaping capacities, deathbots may have positive impacts on bereaved with prolonged grief disorder (PGD). Therefore, the use of deathbots should only be permitted as a medical device under medical or psychological supervision. If deathbots are conceptualized as medical devices, their infringement on the dignity and autonomy of the bereaved is stopped while at the same time the commercialization of the digital remains is reduced. Overall, this master's thesis intends to contribute to the literature corpus of the ethics of deathbots. As deathbots are just starting to enter the market, now is the time to consider them ethically and normatively to provide a basis for their much-needed legal restrictions.

2. Historical Overview and Demarcations

Throughout history, people engaged in grief and mourning rituals, death practices und attempts to stay in contact with the dead. In this chapter, I will provide an overview of the changing societal understandings of death and grief which paved the way for a phenomenon like deathbots. Moreover, I will argue that deathbots may prove to be quite similar to traditional forms of staying in contact with the dead. I will discuss why – nevertheless – deathbots do have decisive qualitative and quantitative differences to traditional death and grief practices. This section will contain a clear demarcation of deathbots from other practices around death and grief. Thus, this chapter justifies why deathbots should be analysed ethically in their own rights.

2.1. Changing Landscapes of Death and Grief

Grief and death practices are an individual, cultural, social and historically changing phenomenon (Sofka et al., 2012). It is therefore difficult, if not impossible, to make general claims about them. In my analysis, I will try to lay out some broad trends, while bearing in mind that there are fundamental differences in these practices.

Recent technological developments, especially the internet and internet-enabling devices, transformed our lived realities. Hence, it is not surprising that these transformations also change societal death and grief practices, ways of mourning, and staying in contact with the dead. Already in 1997 – when it was still normal to explain how 'the internet' works in a research paper – Sofka (1997) described that people used this new medium to find forums to express their grief and mourning upon the death of a loved one. In her paper, Sofka calls for a new way of investigating death practice, terming it “thanatechnology”: a mixture between thanatology (the study of death and loss) and technology which fuse together as a result of the new practices around grief and death (Sofka, 1997). The observed trend was not to stop and eighteen years later Walter et al. (2012, p. 295) judge: “The internet affects key concepts in death studies – sequestration, disenfranchisement, illness narratives, private grief, social death, continuing bonds with the dead, and the presence of the dead in society”. Today, there are online graveyards and platforms on which bereaved come together to mourn (Walter et al., 2012). A different example of mourning spaces are social media profiles which often ‘outlive’ their previous owners. Nobody knows exactly how many dead profiles there are on social networking sites (SNS), but on Facebook alone estimates ranged to 30 million already in 2012 (Stokes, 2021). Keeping with the example of Facebook as one of the biggest SNS, it is possible there to turn the profile of a deceased person into a ‘memorial page’ where ‘friends’ of the deceased can still post on the timeline of the memorial page, but other functions are limited (Bassett, 2015). Some users keep writing on the deceased’s Facebook wall for years. They use it as an internet mourning space and incorporate it in their grief process (Brubaker et al., 2013).

There will be increasingly many digital traces of dead people online in various forms, as more and more people frequently use the internet and will sooner or later die. The digital remains, as they are commonly called, are mostly created unintentionally (i.e. not meant to be online after death) by the deceased when using the internet while still alive. Others, however, are created intentionally by the bereaved, for example online memorials. Intentionally created digital remains may be regarded as a progression or extension of real-life graveyards. While graveyards contain the actual body of a person, their main function after the funeral is to have a memorial site. Having (additionally) a memorial site online opens the space of mourning for more bereaved people, for example family members who live in a different place, to join in mourning practices. Often, online memorial sites are quite like physical memorial sites and allow for example to light a virtual candle or leave memorabilia (Sofka et al., 2012). This, I would argue, is thus not sufficiently different from traditional grief and mourning practices to constitute a specifically technological phenomenon.

However, this is different for unintentionally created digital remains. Turning back to the example of Facebook, pages of dead users can be experienced as ‘creepy’ (Bassett, 2015). They may be understood as being in the ‘uncanny valley’, a term used to describe that people find e.g. robots creepy that are too human-like without resembling humans well enough (Bassett, 2018b). If a person dies and their page is not turned into a memorial site, the Facebook algorithm will continue to feed news about common memories, the deceased’s birthday etc.

into the newsfeed of a still living Facebook user who is friends of that person (Stokes, 2021). Bassett (2015) for example seems to find digital remains on SNS creepy and calls them ‘digital zombies’ to "describe the resurrected, re-animated, socially-active dead'" (Bassett, 2015, pp. 1133–1134). In a later paper, she explains the term further, stating that they “do things in death they did not do in life” (Bassett, 2018, p. 6). While having a less fatalistic vision, Kasket (2012) argues that Facebook takes over the work of priests, mediators and spiritualists mediums, as it offers the living a way to talk with the dead. The dead, on the other hand, are closely remembered and resembled on social media (Kasket, 2012). Unintentionally created digital remains, therefore, can seem less like a natural progression from traditional grief and mourning practices than intentionally created ones like online graveyards.

Are the dead on social media, then, really zombies which wander around the living, resurrected from the dead, hauntingly socially active? I would say that they are not. An important difference from zombies is that the dead do not answer or respond to the living on SNS. They are what Savin-Baden et al. (2017) call a one-way (passive) afterlife presence "where the recipient can read about the deceased in some form of digital memorial, either intentionally created [...] or an existing system which lives on after their death" (Savin-Baden & Burden, 2019, p. 89). This type of digital afterlife presence, I would argue, has some reminisce to traditional forms of afterlife presence and remembering, like the keeping of pictures or letters of a deceased person. On SNS you might read old posts and messages of the deceased person instead. It may be more intense, as many people may join in the mourning and grieving process and many memories may be shared among bereaved (Kasket, 2012). However, it is still a passive engagement in which the dead do not answer. It has a limited scope and does not actively simulate the behaviour of the dead.

2.2. Two-Way (Digital) Afterlife Presences

In a recent study of bereft college students, nearly forty percent of the participants reported that they use social media as a means to deal with their grief and talk to the dead (Varga & Varga, 2019). Using digital remains as a way of remembering does not seem to be experienced as ‘creepy’ anymore, in the way Bassett described it in 2015. This development, as well as technical advancements, I would argue, lead to the imageability of deathbots. While deathbots are still in the ‘uncanny valley’ for many people, they are not for others[3]. Over time, they may thus become less uncanny and more acceptable. Should they, therefore, be treated like a natural progression from digital remains and as a future we just yet have not arrived at? I argue, they should not. In contrast to one-way digital remains, like Facebook profiles, deathbots are two-way, active, afterlife presences which means that there is “the possibility of the digital entity interacting with users and visitors, and with the rest of the living world, in the form of a chatbot or virtual human"” (Savin-Baden & Burden, 2019, p. 89). Deathbots come much closer to Bassett’s vision of a resurrected and socially active ‘zombie’, as they actively engage with the living. Letters do not start to talk back to you, neither does the dead person answer to a message you leave on their Facebook wall. Deathbots, on the other hand, do. This marks a striking difference between traditional digital remains and deathbots.

While the definition of two-way afterlife presences by Savin-Baden and Burden (2019) focusses on digital entities, the concept does not only apply to the digital realm. There is one specific example of the dead talking back to the living in the non-internet world: Séances. Séances promised to close the border between the dead and the living and let the living have a conversation with the dead (Connor, 1990). They were in vogue in 19th century Europe and North America and are part of some indigenous cultures (Connor, 1990; Stokes, 2021). Séances generally require a person acting as medium between the living and the dead (and not any person can be the medium). Séances are thus a mediated communication between the living and the dead, not a direct communication. This is similar to deathbots, which are mediated by technology. The deathbot algorithm and the device the user facilitates mediate the contact between the user input and the deathbot output. Nevertheless, deathbots are not just a newer form of séances. While both are mediated, they have important qualitative and quantitative differences which are due to the specific technological character of deathbots.

First, there is a quantitative difference. Séances have a certain timespan. The medium calls the dead, then the dead and the living have a conversation, then the séance ends (Connor, 1990). Séances may be repeated. However, due to the necessity of a human medium, they are not constantly available. This is different for deathbots, which may be used frequently as a long as the user has an internet able device with him*herself (Luxton, 2020). Deathbots may be used whenever the user feels like it (even in the middle of the night) and in a quantity completely

[3] When the 34-year-old Russian tech entrepreneur Roman Mazurenko suddenly died by being hit by a speeding car in 2015, his friend Eugenia Kuyda (working in the area of AI development) decided to create a deathbot from Romans extensive bulk of text messages. Today, it is possible for anyone to download the Roman Mazurenko deathbot for free (Nagels, 2016).

dependent on the will of the user. With deathbots, talking to the dead can become a 24/7 experience, as the user may take them to work, grocery shopping, to bed and on travels. This is a specifically technological aspect of deathbots.

There is, furthermore, a qualitative difference. The medium in séances is a human who feels empathy and can detect the emotional affordance of the person sitting in front of her*him and adapt the answers s*he provides accordingly. Deathbots do not have such a capacity. This capacity to feel the affect of the bereaved marks an important difference. Moreover, mediums use their own voice (which may, however, be slightly changed to incorporate the role of a ghost) while deathbots imitate the speaking and writing behaviour of the deceased down to the level of grammatical particularities. This is due to their technological character which allows them to mimic the deceased very well. Lastly, in séances, bereaved believe that they are talking to the dead while in deathbots users will know that they are not actually communicating with the dead. However, even though the users have the rational knowledge, emotionally the line may blur as some people report that they feel that the dead are, for example, listening to them on Facebook (Kasket, 2012).

Overall, deathbots are a phenomenon which evolves out of a historical background and out of a certain form of sociality. Nevertheless, they are different to both one-way afterlife presences and traditional two-way presences like séances. This is due to their technological nature through which they possess distinct qualitative and quantitative characteristics.

3. Chatbots – Functioning and State of the Art

In this chapter, I will first clarify what I mean when I use the terms ‘chatbot’, ‘personalized chatbot’ and ‘deathbot’. Followingly, I will give an insight in the technical infrastructure of chatbots and provide an overview of what they presently can and cannot do. As some chatbots produce quite astonishing results, it can be easy to mystify them and lose sight of their technical, unintelligent behaviour. Through this chapter, I will try to avoid this tendency and de-mystify chatbots. As deathbots are a field of active research and continuous advancements, I will try to provide an insight into the momentary stand. However, in five years’ time, the abilities of bots may already be highly advanced.

3.1. Chatbots, Personalized Chatbots and Deathbots

There are various different terms for the word ‘chatbot’ “such as machine conversation system, virtual agent, dialogue system, and chatterbot” (Ciechanowski et al., 2018). They are “interactive, virtual agents that engage in verbal interactions with humans […] through the usage of natural language” (Przegalinska et al., 2019, p. 786). Chatbots are software applications programmed to provide reasonable output to some given input. The output as well as the input may be either human speech or writing. The training data of chatbots consists of large datasets of texts or speech which are usually produced by various individuals (for example all English language Wikipedia articles). The term ‘personalized chatbot’ refers to a chatbot which is trained on the data of one specific person to mimic the speech or writing behaviour of that person. Personalized chatbots are therefore a sub-type of chatbots with the same general components and similar computing to ‘normal’ chatbots, with the important difference being the dataset they are trained on. The personal data used for training may be harvested from social media platforms such as Facebook or Twitter, personal messages by various messenger services, blogs, letters, and videos. The writing style of the trained chatbots simulates that of the ‘real’ person whose data they are trained on.

Deathbots, lastly, like the Lily deathbot from the opening example, are a special kind of personalized chatbots. They are trained on the data of one specific deceased person. In the following, I will use the term ‘deathbot’ regardless of whether the chatbot was created post-death by the bereaved (or, more likely, by a company which is paid by the bereaved to do so), whether a person initiated the creation of a chatbot of him*herself prior to her*his own death with the intended use by the bereaved following his*her death or whether the deathbot was created as a personalized chatbot by a living person to his*her own use while being alive and is only turned into a deathbot after the death of that person. In the two latter cases, there is the possibility that the still alive person may not only feed already existing data into the chatbot, but may also modify the chatbot, for example by writing with it or by correcting and changing its answers (Feine et al., 2020; Savin-Baden & Burden, 2019). For my discussion of deathbot, it does not matter whether a bereaved or the deceased person decided to create the chatbot. While this may make an important difference regarding the perceived autonomy of a person and even a decisive difference regarding the data protection and data ownership of the deceased person, this differentiation is not necessary for my main claim that deathbots impact the dignity and autonomy of bereaved and I will hence not distinguish the three cases. What I do, however, take for granted is that the person using a deathbot experiences sincere grief about the death of the person represented by the deathbot. Using a deathbot created to imitate a dead person I do not feel grief about (and potentially do not even know) is also a possible – though arguably less likely – application, which I will not consider exhaustively in my thesis.

3.2. The Functioning of Chatbots

The development of chatbots started in the 1960s (Ciechanowski et al., 2018). Since then, chatbots massively improved and are nowadays used in a wide range of applications from costumer service, to health care and robotics (Nagarhalli et al., 2020). They can decrease the human workload in certain areas massively. Thus, they trigger a lot of research and investment (Nagarhalli et al., 2020). Chatbots can be broken down into two basic elements: a knowledge base (which may be open or closed domain) and a machine learning model which comprises a response generation component and natural language processing (Lokman & Ameedeen, 2018). The knowledge base is the "heart" of a chatbot. It constitutes the output of the chatbot (Nagarhalli et al., 2020). A conversational chatbot is typically trained on a wide range of material (Hristidis, 2018). For a conversation with a chatbot to feel somewhat natural, the chatbots needs a lot of training data and the quality of the training data must be high (Abdul-Kader & Woods, 2015). Through machine learning methods (which will be explained below), the chatbot algorithm learns to find patterns and structures in the training data and labels the data. Based on this, it can match user input to the learned data und thus provide an output which makes sense both grammatically and content-wise.

A chatbot may have an open or closed domain knowledge base. A chatbot with a closed knowledge domain will only use the 'knowledge' it has already imbedded. It is not able to provide a good answer to a question such as "how will the weather be tomorrow?". Open domain chatbots, in contrast, can access information on the internet to provide answers. For example, an open domain chatbot would be able to connect to the weather forecast of the user location to output the answer "it will be sunny tomorrow in Osnabrück with temperatures ranging from 12-15 degrees" (Lokman & Ameedeen, 2018). Little surprising, it is more difficult to create an open domain than a close domain chatbot.

The response generation of chatbots follows an encoder-decoder framework (Lokman & Ameedeen, 2018). If a text is inputted into the chatbot, the encoder will break the text down in small sequences. Followingly, the knowledge base is searched for most matching responses and then, lastly, the decoder will output a text sequence which is displayed to the user as a response (Lokman & Ameedeen, 2018). Thus, a response is generated. The processing of both the inputted as well as the outputted text is implemented in natural language processing (NLP). NLP is concerned with how computers can analyze and categorize natural language data and combines AI and linguistics. Modern NLP is based on machine learning methods (Nadkarni et al., 2011). Machine learning, as the name already implies, means that the machine (or computer program) 'learns'. This is done through the adapting of statistical weights either in an unsupervised or a supervised way. When a program learns, say, to classify whether a picture depicts a dog or a wolf, the program is fed with many pictures which are classified as wolf or dog. Through these pictures, the program then ascribes certain statistical importance (weights) to certain features. It may for example make a difference how far the pixels depicting the eyes are apart. If they have a certain proportional distance, there is a positive statistical likelihood that the picture depicts a dog. The machine learning algorithms used in modern-day NLP mostly have a deep neural network structure. Deep neural networks consist of several layers of 'neurons' (Deng & Liu, 2018). Each neuron is connected to all neurons of the layer before and after its own layer. The statistical assigning of weights is done for several features and through several layers. The weights are adapted repeatedly until a given input matches a desired output sufficiently often, e.g. until the algorithm in 95 percent of the cases correctly classifies a picture to either showing a wolf or a dog.

This general structure of chatbots with a knowledge base, response generation and NLP implemented in a machine learning model holds true for normal chatbots as well as for personalized chatbots and deathbots. Only the knowledge base will be different in the later cases. It will largely be constituted by data of an individual person, which is either still alive or already deceased. The algorithm of the bot can extract what the typical sentence structure of a person consists of, how that person responds to specific questions and how s*he engages in conversations. The NLP system is then trained, meaning that it transfers the knowledge base into statistical weights and connections. When interacting with a user, the bot mimics the conversation behaviour of the person whose data was previously inputted. A deathbot will output similar answers to what the dead person would have written because it uses the data of that person. A chatbot may also be able to 'remember' previous conversations with a user if it is programmed to store them and access them in further conversations. Thus, the impression that continuous conversation takes place can be evoked.

Current chatbots produce varying results. Feine et al. (2020) for example state that chatbots often do not feel natural because their responses are constrained. Diederich et al. (2020) argue similarly that chatbots often do not continuously give meaningful answers.

Sometimes the answers may be out of context and do not match to the user input. Others, however, can produce astonishing results. The 2015 documentary “Alice Cares” shows a social care robot, Alice, which is a chatbot placed into the body of a doll-like robot with a camera behind its eyes (Simon, 2015). The chatbot involves speech to text and text to speech processing, which means that users can speak with it and Alice will output natural speech as well. Alice interacts with elderly people and is supposed to be their “companion” (Burger, 2015). The robot memorizes the interactions with the elderly so it can have a continuing conversation with them. The programmers claim that Alice can elevate the pending loneliness many widowed elderly face because with Alice they have something (though not someone) to talk with. Without discussing the ethical issues of such a robot, the conversation results and level of interaction between Alice and the seniors shown in the documentary is striking. The documentary displays how ‘Alice’ has conversations with seniors. Though the conversations sometimes fail, surprisingly often they seem quite natural. The participants seem to talk with the bot quite like they would with a human.

Further Remark

It is important to remember that a chatbot does not ‘think about’ or ‘question’ what it outputs. It merely outputs a statistically most likely sentence structure of the deceased. While writing this thesis, I will try to make sure not to ascribe personal agency to chatbots or to humanize them. If it, regardless of the attempts, seems like I ascribe a character or human-like characteristics to a chatbot when discussing the ethical implications of deathbots, this ascription is unintentional. Assigning human-like characteristics on technological applications happens easily. Through it, chatbots are humanized which may evoke a feeling that they have emotions, which they do not actually have.

3.3. Anthropomorphism

The research field of human-computer interaction (HCI) investigates how humans interact with computers. One of its goals is to approximate the interaction between computers and humans as much as possible to human face-to-face interactions (Seeger et al., 2017). Certain chatbots can appear very human-like:

> “Modern chatbots are characterised by conversational interfaces that make them increasingly able to simulate human conversations, to such an extent that customers may well not realize, that they are talking to a chatbot rather than a human services assistant. Furthermore, even if they do realize they are speaking to an automated agent, because chatbots display human conversational behaviours, they encourage, even entice, […] to engage with them in a reciprocal human manner, treating interactions as actual conversations, rather than […] para-conversations” (Murtarelli et al., 2021, p. 927)

This points at the two-sidedness of the process: while programmers often aspire to make computers and programs more human-like, users also often start to ascribe humanness to computers (c.f. Skjuve et al., 2021). Moreover, in the already mentioned documentary “Alice Cares” the elderly women interacting with the social bot (which does not look very human-like) start to develop empathy with it. They gender it - “she” - , address it like a human “what do you think?” and even ask “what will you do with her now?” when the researchers take Alice again to their lab, indicating that they are worried about the wellbeing of the robot (Burger, 2015). This is not an uncommon phenomenon, as some empirical studies have shown that chatbots may induce a sense of relationship in their users (Skjuve et al., 2021). Users of chatbots report that they see it as a friend, companion and in some cases even as a romantic partner (Skjuve et al., 2021).

The phenomenon of ascribing human characteristics on non-human agents or objects like in the examples above is not specific to machines, let alone chatbots. For example, it is not uncommon to attribute some type of humane-like qualities on pets. This process, termed ‘anthropomorphism’, “describes the tendency to imbue the real or imagined behavior of nonhuman agents with humanlike characteristics, motivations, intentions, or emotions” (Epley et al., 2007). Anthropomorphism may happen consciously or unconsciously (Kim & Sundar, 2012). It can – and often is – exhibited towards chatbots (Ruane et al., 2019; Seeger et al., 2017). While I do assume that people using a deathbot know that they are talking with a deathbot and no chatbot system so far has passed the Turing test[4], this phenomenon may still have important implications for the ethical use of deathbots, the dignity of the deceased and the affective state of the bereaved.

[4] The Turing Test was formulated as a criterion to check whether humans and computers have similar cognitive capacities. Basically, a human investigator has a conversation with a human and a computer without knowing who of the two is the computer. If the computer fools the investigator to think that it is the human, it has passed the test.

4. Grief and Mourning

After the loss of a person, the main feeling bereaved experience is grief. As will be shown in this chapter, grief may comprise several different emotions such as sadness, relief, loneliness, anger, or fear. As deathbots are implemented after loss of a person and used by bereaved, I will focus on grief in the discussion of the emotion-shaping capacities of deathbots (c.f. Stokes, 2021). Not for nothing deathbots have been called “griefbots” before, implying that they may take part in grieving or shape grieving processes (Bassett, 2018b). Because of the prevalence of grief in the affective life of deathbot users, it is important to analyse what grief is, how it affects the bereaved, and what that means for the use of deathbots. I will argue that interfering in grief processes may fundamentally impact the being in the world of the bereaved which may have negative consequences for that person. Thus, I claim that an investigation of the emotional impact of deathbots is justified as a basis for an ethical analysis of them.

In this chapter, I will explore the question of what grief is first from a psychological perspective and second from a philosophical perspective. Followingly, I will give a brief overview over the philosophical debate of online affects, concentrating on the concept of internet-enabled techno-social niches by Krueger and Osler (2019) and online empathy. The concepts of online affect and of grief will provide me with a basis to argue that users may experience grief while using a deathbot. Moreover, I will argue, that deathbots may have a fundamental impact on grief processes. As we will see in this chapter, grief is a unique emotion as it may comprise many singular emotions and is a process re-orientation in the world. The impact of deathbots on this process may be detrimental to healthy grieving. Additionally, I will discuss the potential of bereaved to become (overly) reliant on deathbots.

4.1. Psychological Conceptions of Grief

Grief stands at the intersection between loss and love, attachment and separation (Neimeyer & Thompson, 2014). Upon the death of a loved person, the story we had constructed of the world and about ourselves within it may fundamentally change (Neimeyer & Thompson, 2014). During the grief process, the bereaved needs to re-learn these stories, find orientation in a changed world and has to make sense of the death. As Neimeyer and Thompson (2014) phrase it: “grieving [is] a process of reaffirming or reconstructing a world of meaning that has been challenged by loss”.

The first psychological theory of grief and mourning is often ascribed to Freud, who suggested that bereaved had to work through their loss to detach emotionally from the dead (Rothaupt & Becker, 2007). This involved first an excessive desire for the lost person (hypercathecting) to then withdraw one’s feeling of attachment from that person (decathecting) (Rothaupt & Becker, 2007). An ongoing emotional relationship with the dead was believed to be pathological. This model of grief was dominant with small changes and additions for a substantial amount of time. In the middle of the 1980s, however, some researchers started to question the need to cut all emotional ties with the dead. In 1996, the seminal book ‘Continuing Bonds: New Understandings of Grief’, which contained several studies demonstrating that bereaved often keep emotional bonds with the dead alive in ‘healthy’ grieving, was published (Klass & Steffen, 2017). This book transformed into a theory of continuing bonds which conceptualizes a changing relationship with the dead important for successful mourning and grief work. The grief work of changing the relationship entails a recognizing and accepting of the end of the physical relationship with the dead. This allows for new emotional bonds with the dead (Rothaupt & Becker, 2007). The theory of continuing bonds marked a paradigm shift in the understanding of grief and grief therapy. Today, it is the dominant psychological theory of grief.

The bonds in the continuing bonds model are socio-culturally mediated, dynamic, and shaped by cultural narratives (Klass & Steffen, 2017). There are many different examples of how continuing bonds may manifest themselves, for example in dreaming of the deceased, talking to the deceased and experiencing a presence of the deceased. It can also involve more physical ways of remembering, like mementos and legacy projects. Moreover, the bonds can be part of a communal setting, for example by sharing stories about the deceased with other people who knew her*him or by writing on the Facebook wall of the deceased (Kasket, 2012).

Some bereaved “struggle with intense, prolonged and complicated grief, characterized by extreme separation distress, preoccupation with the loss, and inability to function in major life roles across a period of many months or years” (Neimeyer & Thompson, 2014, p. 4). While for most people, “the intensity of grief diminishes as the finality and consequences of the loss are understood and future hopes and plans are revised” (Shear, 2015, p. 154) at some point of the grieving process, it is not for bereaved who experience this form of prolonged grief. Prolonged grief disorder (PGD) is recognized as a psychological disease which leads to a reduced quality of life and mental health problems (Boelen & Prigerson, 2007). About two to fifteen percent of bereaved develop a prolonged grief disorder (Neimeyer & Thompson, 2014; Shear, 2015; Wittouck et al., 2011). The treatment of PGD "includes two key areas of focus: restoration (i.e., restoring effective functioning by generating enthusiasm and creating plans for the future) and loss (i.e., helping patients find a way to think about the death that does not evoke intense feelings of anger, guilt, or anxiety)" (Shear, 2015, p. 156). Moreover, it may involve a re-negotiation of the bonds with the dead.

Overall, the continuing bonds theory claims that it is natural to have a non-static continuing bond with the dead and to have feelings of profound loss which may last for a substantial amount of time. Nevertheless, grief is individual and there is no 'correct' way of grieving that holds true for everyone (Wittouck et al., 2011). While continuing bonds may be helpful for some bereaved, not all grieving people experience them and it is not necessary to have continuing bonds for a 'healthy' grieving process (Klass, 2006). Additionally, continuing bonds should be understood separately of prolonged grief, which constitutes a disorder with the potential to inhibit the ability to live a good and happy life.

4.2. Philosophy and Phenomenology of Grief

> *"[E]ach of our selves [sic] and lives is unfathomably rich, complex, and essentially never finished. Because of this, [...] our grieving is a never-ending process that entails repeated and inevitable struggles with finiteness, continuous change, pervasive uncertainty, and vulnerability. In this open-ended coping we can glimpse the mystery of living as a self that ultimately limits others' understanding of us and our own self-understanding"* (Attig, 2011, p. 122)

Let us now turn to philosophical conceptions of grief. They often build upon the psychological understandings of grief which were just presented. I will especially draw on Ratcliffe (2016, 2017, 2020) and Fuchs (2018) in this chapter, as their understandings of grief are rooted in phenomenology. Phenomenology, very broadly speaking, is the philosophical "study of human experience and of the way things present themselves to us in and through such experience" (Sokolowski, 2000, p. 2). This approach fits my research goal very well as I am interested in investigating how bereaved experience grief, and if and how this experience is changed or shaped through the usage of deathbots. Examining the phenomenology of grief can therefore provide valuable insights for the later discussion on grief and deathbots.

Death is, undoubtedly, a fundamental condition of life. As people die, other people live and need to learn living with the physical inexistence of the other person. Grief, the painful emotion of having lost a loved friend, partner, or family member, is therefore, just like death itself, a fundamental part of human life. While, I would argue, most people have a general idea what grief means and how it feels, it is difficult to put a finger on what grief actually is. Grief involves sadness, but also a feeling of longing for the other person, loneliness, potentially anger that the other person had died, a continuing feeling of closeness with the deceased, and many more. Grief is a difficult to grasp emotion, as it comprises many other, more singular, emotions (Ratcliffe, 2017). The different emotional elements hang together in grief and form a pattern (Goldie, 2011). However, the pattern is not necessarily coherent and cohesive (Ratcliffe, 2017). On the contrary, emotions within the web of grief may be present at certain times and absent at others. While some emotions may be very intensive one day, they may only be a vague background feeling the next day, when some other feeling is in the foreground of experience (c.f. Goldie, 2011). The different parts of the pattern of grief are therefore not in themselves always essential to grief in every moment of grief. However, if all the different parts were taken away and the emotional components would not be present at any time, we would probably not understand the feeling of a person as grief anymore (Goldie, 2011).

The patterned character of grief points towards another central aspect of grief: its temporal extendedness. Grief *has* to last long, otherwise it is not grief (Goldie, 2011; Ratcliffe, 2017). It sounds odd to say "for a second he felt deep grief" while it sounds perfectly normal to say "for a second he felt violent pain" (Wittgenstein, 1968). This makes grief a rather unique emotion. While grief itself lasts long, the different singular emotions constituting it may only last short but evolve in their communal pattern over time. To picture this temporally extended pattern, it is helpful to understand grief as a process. Not as a singular and linear one, but rather as a multi-faceted, complex, heterogenous, changeable and transforming process, which does not follow any pre-defined standard route (Fuchs, 2018; Ratcliffe, 2016, 2017). The grief process does not have a clear end and may, for some people in certain circumstances, never end (Ratcliffe, 2016).

After having laid out the *character* of grief, I will now turn to what *happens* during grief and discuss the importance and necessity of grief. Grieving involves a re-learning and re-orientation in lived space and time (c.f. Attig, 2011). Our relation to the world is partly constituted by our relation to others, as "those we care deeply about and share our lives with are integrated into the habitual world in all sorts of ways" (Ratcliffe, 2017, p. 163). Interpersonal encounters shape our thought and, moreover, through interaction with others meaning making can take place. Thus, if a closely related (as in: related to our life-world) person dies, we need to re-learn our lives without that person. As the other person was integral to our own life his*her death changed what matters to us and how it matters to us (Ratcliffe, 2016). Previously meaningful connections between things may feel eroded as the intelligibility of the world, crucially depended on the person who is lost, suddenly ceases to exist. There is nothing that we can easily retreat to (Ratcliffe, 2020). Grief, therefore, entails to find new ways of experiencing meaning in the world. To phrase it in phenomenological terms: when we lose a loved person, we need to assume and find a new orientation to the world as our experience and relation to the world as a whole changes (Ratcliffe, 2017). In grieving, we need to re-negotiate our being in the world.

In this re-learning and re-negotiation process, the bereaved experience a feeling of ambiguity and uncertainty between "presentifying and a 'de-presentifying' […] presence and absence, between the present and the past, indeed between two worlds they live in – an ambiguity which may also manifest itself in being painfully torn between acknowledgment and denial of the loss" (Fuchs, 2018, p. 44). The notion of being in two worlds – one before being bereaved and one while being bereaved – is helpful to understand the strife bereaved experience. Fuchs (2018) convincingly shows that the world time and the dyadic time (pre-loss) may dissociate for the grieving person. The temporal experience of the world time may change upon a loss as the time can feel unrealistic and estranged to the bereaved. Grief may also change past times: "[w]e relearn our unfolding life histories in the light of our losses. In so doing, we reinterpret and appropriate new understandings of, and come to live differently in relation to, our own past, present, and future. We also relearn especially significant events and occasions as we reinterpret their significance and learn how to live through them without the deceased" (Attig, 2011, p. 119). Especially shortly after the loss, there may be no anticipation of the future, as a future without the lost person seems unimaginable (Fuchs, 2018). This may be one reason why people who are in grief exhibit higher rates of suicidal tendencies (K. Szanto et al., 1997).

The experience of presentifying and de-presentifying, between presence and absence Fuchs (2018) mentions in his explanation of the two worlds is a fundamental aspect of grief. It points to the necessity of the bereaved to re-negotiate her*his relationship with – and attachment to – the deceased. The attachment of two close persons prior to death, on the bodily level, can be understood as a shared intercorporeality:

> “[w]e may speak of a dyadic body memory which consists in the shared habitualities of interaction […]. Thus, while sharing their lives, both partners have become part of an intercorporeal sphere with its specific style of greeting, talking, smiling, walking together, etc. Bereavement means a separation of this intercorporeality” (Fuchs, 2018, p. 47).

This separation can feel very painful even on a physical level, which leads Fuchs (2018) to compare it to the experience of phantom pain when a limb is amputated. Bereft people may feel a deep longing for engagement with the dead which cannot be fulfilled (Wonderly, 2016). Some bereaved report that, while they intellectually know that the person is dead, it still feels for them as if the dead person was still alive (Fuchs, 2018). Again, the bereaved are torn between two worlds and slowly need to re-negotiate their attachment with the dead. The continuing bonds model, which I discussed in detail in the previous sub-chapter, provides a valuable tool to understand this re-negotiation. The bond with the dead often stays with the bereaved for a long time, and a re-negotiated attachment may never cease to exist (Fuchs, 2018; Ratcliffe, 2016). The continuing bonds are a complex nexus and may be stronger or weaker for different people and at different times (Køster, 2020).

What marks the difference, lastly, between successful grieving and a grieving which leads to prolonged grief disorder and an impoverished quality of life in the long turn? The ambiguity and in-betweenness of being in two-worlds may be a painful experience, but it is also necessary for grief adjustment and coming to terms with the dead (Fuchs, 2018). In grief, we need to re-form our bonds, relationships and attachments with the dead and learn how to be and act in the world differently (Attig, 2011). The deceased is in this process gradually incorporated in the bereaved, and the bereaved may feel an inner, comforting presence of the dead instead of searching outside for him*her anymore (Fuchs, 2018). At the same time, the deceased may be represented in the outside world in various ways, such as in symbols, memorabilia, rituals etc. Moreover, for successful mourning it is necessary to fully (not only intellectually) acknowledge that the deceased is dead. Over time, this allows for the past to become the past (Fuchs, 2018). The dead person needs to go from an imagined not-being-alive-but-also-not-being-dead- status to an acknowledged status of being dead. Grief thus constitutes a recognition of loss. This allows for the dyadic and world time to re-align again (Fuchs, 2018).

5. Philosophy of (Online) Affect

After having discussed the philosophy of grief, I will now turn to the philosophy of online affect. As I will show, online interactions have the potential to shape affective states. This can happen through internet niches which are created and shaped by their user and at the same time influence the affective state of their users. The chapter will start with an introduction to the concept of situated affectivity and affective scaffolding. Followingly, I will discuss theories of online affective scaffolding. This will mainly concentrate on Krueger and Osler's (2019) concept of internet-enabled techno-social niches. These theories will provide me with a basis to argue that deathbots may affectively influence their users.

The philosophy of emotion has largely turned away from theories of ‘brainbound’ cognition and affectivity which apprehend cognitive processes and emotions as being purely in the mind of the respective person (Stephan & Walter, 2020). Instead, theories of ‘situated’ affectivity and cognition are dominant today. They do not understand cognition and emotions as sole brain processes but possibly as extracranial processes which can take place in the whole human body and even – in one way or the other – in the outside world (Stephan & Walter, 2020). Based on those situatedness approaches, Sterelny (2010) proposed a theory of (cognitive) scaffolding. He argues that human cognition is scaffolded in the environment, meaning that it does not only take place within the body of a person but is also – often decisively – influenced by his*her environment. For example, a theatre stage and the props of a play may be arranged in a certain way for the actors to better remind their roles (Sterelny, 2010). The scaffold of the stage is deliberately arranged to aid the actors remember where to go. Sterelny (2010) further states that humans engage in niche construction, a term he borrows from evolutionary theory. In the context of evolution, the term ‘niche construction’ is used to explain how animals optimize their environment to best fit their needs while they also adopt their own behaviour and appearance to best fit their environment. In the case of humans, Sterelny (2010) claims, the niche construction is often epistemic, meaning that the niche construction is aimed to support and scaffold intelligent action. Turning back to the example of the actors and the stage, the stage functions as a cognitive niche, constructed to aid (or ‘scaffold’) the cognition (especially the memory) of the actors. A cognitive niche is more or less stable over time and created for the purpose of cognitive aid. The term ‘scaffolding’, in contrast, refers to the aiding of human cognition by outside resources in general. This model of cognitive niche construction and scaffolding is widely accepted today (c.f. Stephan & Walter, 2020).

Colombetti and Krueger (2015), among others, took Sterelny's theory of niche construction and scaffolding from the realm of the cognitive to the realm of the affective. They claim that affectivity is not just a passive process in which an individual undergoes bodily and experiential changes, but that it also has an active component. I can actively and intentionally modify my own affective state in many situations. For example, I can turn on my favourite pop song to get into a partying mood before going out with friends in the evening or I may go on a walk in the forest to get into a calm, settled mood. These examples point to another similarity between cognition and affectivity: besides having an active dimension, affectivity is also scaffolded in the environment. Moreover, when actively manipulating my environment to affect me in a certain way, I create affective niches which are "instances of organism-environment couplings (mutual influences) that enable the realization of specific affective states" (Colombetti & Krueger, 2015, p. 1160). Like in the example of cognitive niches, I create affective niches to reliably affect me in a certain way. For example, I may decide to wear bright red rainboots whenever it rains to feel happy about the colorful boots instead of annoyed because of the bad weather. Importantly, the interaction between the environment and the individual in affective niches is always two-directional. An affective niche can affect me even in situations in which I do not necessarily want to be influenced by it.

While the previous examples of affective niches were examples of human-real world interactions, the theory of affective niches can also be applied to online scenarios to explain internet scaffolded affect. Krueger and Osler (2019) argue that we can feel affect while being online – we can scaffold our affect online. As part of this scaffolding, we create internet-enabled techno-social niches. These niches are created to shape our emotions in certain ways, while we are also shaped by them when we enter them. Due to the specificities of the internet, the affective scaffolding in the realm of the internet has special characteristics. The internet is hyper social as it is always available, hyper available as it can be used at any time, and hyper portable as it can be taken mostly everywhere (Krueger & Osler, 2019). Moreover, we scaffold our affect online in special ways because the internet runs easily through several techno-social niches. The different affective niches we have created may only be a klick away from each other. As we use the internet frequently, we become reliant on it for our affective scaffolding (Krueger & Osler, 2019). It may, for example, become a part of my daily morning routine to check my Instagram feed. In this way, I feel connected to my friends all over the world, as I can see their pictures and have the impression to get a glimpse in their daily lives. I experience that as a nice, comforting feeling. However, this morning as I scroll down my posts, I see the picture of an old classmate of mine. She looks super happy posing in front of picturesque beach while I sit at home on a cold and rainy winter day. Moreover, the picture has already received many "likes" and positive comments. I instantly and involuntarily start to feel jealous. I want to be at the beach too and want to have her popularity. I start to question myself – am I popular? Do I have less friends than her? Is she prettier than me? Suddenly, the formerly comforting affective niche affects me negatively. I decide to close Instagram and visit my Facebook page to look for comfort there. When opening the App, I see that a friend of mine has posted a "memory" of a picture of us she posted three years ago. Seeing the memory of the happy moment makes me feel better again.

This example shows how easily I can switch between different internet-enabled affective niches. Moreover, it may also highlight the reliance one can develop on an internet affective niche. In the above example, it is unlikely that I will delete Instagram now. Despite the negative emotions it has evoked for me once, I still rely on it to generally give me a good feeling as part of my morning routine. However, the next time I visit Instagram I may "unfollow" the old classmate to stop seeing her popular pictures which may affect me negatively. This points at another important feature of affective niches (both on-and offline): they are highly individualized (Krueger & Osler, 2019). I intentionally shape them to affect me in a certain desired way and therefore I individualize them according to my own affective wishes. Additionally, Krueger and Osler (2019) highlight that important features of these niches are trust, reliability and entrenchment which allows to comfortably settling into them[5]. Only because they possess these characteristics, they have their distinct affectively regulative power in the first place. Going back to the above example, Instagram *reliably* scaffolds my affect in a desired way, otherwise I would not frequently use it. It is reliably accessible to me and I *trust* it to regulate my emotions in the intended way. I do not question whether it will affect me, because I already trust it to do so. I am *entrenched* in it as I incorporate it in my daily life and am incorporated by my affective niches (Krueger & Osler, 2019). Importantly, through affective niches the internet scaffolds not only online affect but also offline activities and life. For example, when I am sitting in a restaurant with friends and I see my phone blinking up because I got a WhatsApp message from my mother telling me that the suspicion that she may have cancer was false, I may feel immensely relieved and happy which influences how I interact with my friends for the rest of the evening.

[5] Coninx and Stephan (2021) argue similarly that among the relevant dimensions of affective scaffolding are trust, robustness and control (overlapping with reliability), and mineness, referring to the deepness of the entrenchment of the scaffold into the self-narrative of a person.

While internet-enabled niches promise to be reliable, trustworthy, and entrenched, the internet-enabled scaffolding may also open the possibility for emotional *dysregulation*. More specifically, "[e]asy access to the Internet [sic] and the emotional resources it provides can also lead to what we call *overreliance* and *overregulation"* (Krueger & Osler, 2019, p. 227). The "term overregulation is meant to pick out how the Internet, by allowing us constant access to highly tailored and individualized scaffolding and niches, make us reliant not just on specific forms of emotion regulation but upon emotion regulation itself" (Krueger & Osler, 2019). I may become *depended* on Instagram to scaffold my affect to feel positive about myself. This may lead me to post only photos which fulfill certain beauty standards to get more "likes" and "followers". If I do not get likes for a picture, I may feel seriously unhappy. However, I may also feel unable to stop using the App, as I feel that I need this form of emotion regulation. I may stop to marvel at the beauty of the mountains because what makes me feel happy is the likes I get for the picture of the scenery I post online. Moreover, I may start to take over unhealthy femininity ideals and try to lose weight to adhere to the dominant beauty images and get positive emotions through posting pictures of me, thus reinforcing the norms (see also Krueger & Osler, 2019). My emotions may therefore not only be positively regulated in online environments, but also actively dysregulated which can lead to an affective precarity (Krueger & Osler, 2019). Some of the features of internet-enabled techno-social niches like the individualization and entrenchment certainly also hold true for 'offline' affective niches, others are due to the specific characteristics of the internet, like the hyper-availability, of the internet outlined above. Overall, however, this discussion shows that our emotions may be scaffolded by internet interactions and by interactions with internet-enabled technology.

Krueger and Osler (2019) name deathbots as an example of such internet-enabled techno-social niches. They state that the "Internet, by providing dynamic, ongoing interactions with chatbots offers a novel form of engineering the affective contours of our grief processes" (Krueger & Osler, 2019, p. 223). Deathbots are highly individualized as they depict one specific person, or at least the impression the person left through their digital remains. Users are likely to trust the deathbot to function and to affect their emotions reliably in a certain way. Moreover, users may be highly entrenched with their deathbots, use them frequently and comfortably use them in their daily lives. Deathbots are likely to affect the offline interactions of their users, too, as they may give them advise or, through their affective scaffolding, lead their users to emote in a certain way which influences the behavior of that person. Krueger and Osler (2019) frame deathbots as a positive development, providing comfort to the bereaved and allowing to foster continuing bonds with the deceased. According to Krueger and Osler, deathbots may be quite beneficial for the grieving process of some users. They are "one part of a broader repertoire of grieving rituals that provide concrete structures for organizing and balancing their emotions as individuals work through the contours of their grief" (Krueger & Osler, 2019, p. 223). Deathbots may provide a structure to the experience of grief, thus helping the bereaved. I will discuss this account of deathbots in detail below. Deathbots, overall, can be conceptualized as specific internet-enabled techno-social niches and may, according to Krueger and Osler (2019), have positive outcomes for bereaved.

6. The Grief-Shaping Capacities of Deathbots

Following Krueger and Osler (2019), deathbots are specific internet-enabled techno-social niches. As already pointed out, techno-social niches promise to be reliable, trustworthy, and entrenched. Do deathbots fulfil these characteristics? What does it mean for them to promise these characteristics and potentially fail to uphold them? These are the questions which I will first discuss in this chapter. Afterwards, I will turn to the question what the position of deathbots as internet-enabled techno-social niches means for the affects and grief process of deathbot users. I will argue that deathbots have emotion-shaping capacities and that, therefore, they may have an impact on grief processes of bereaved.

6.1. Reliability, Trustworthiness and Entrenchment

Internet-enabled techno-social niches feel, and in most cases are, reliable, trustworthy, and entrenched (Krueger & Osler, 2019). If people would not attribute those characteristics on the niches, they would not trust them to regulate their emotions and thus the techno-social niche would not function as an affective niche. However, Krueger and Osler (2019) caution that techno-social niches may lead to *overreliance* and *overregulation* in certain situations. This can lead to an emotional *dysregulation* of people within their affective niche. Deathbots promise to be *reliable*. Do they uphold the status of being reliable for their users? They are certainly constantly accessible. As they are most likely implemented in some form of an App, users need an internet-enabling device such as a smartphone or tablet and access to the internet to use them. If both are available – which they are under most circumstances – deathbots are reliably accessible. The technical accessibility of deathbots is constant – bots naturally never sleep or have a time off. Bereaved can use their deathbot whenever they want. This constant availability of support, even at times when other social support is not available might be really comforting (Döveling, 2015). Users know that they can access their deathbot in a sleepless night, when they may not be able to contact a close friend or family member for emotional support. This may lead to the impression that deathbots are reliably accessible.

However, even though chatbots may be mostly reliable, they may also fail this promise in certain situations. There can be various reasons for failure, one of them is that most companies, which momentarily offer to create deathbots, are startups. As the death industry in general is a large economic sector worth "an estimated US$16–20 billion per annum in the United States" in 2017 alone, it is not surprising that tech companies want to get a share of the money (Arnold et al., 2017, p. 2). Thus, start-ups may seize the opportunity to offer the creation of deathbots. However, as Kneese (2019) argues, tech startups often fail. Failing is considered a normal and acceptable process for start-ups in Silicon Valley where most of the US tech companies and startups are based (Kneese, 2019). This culture of failing may not have serious consequences in the programming of, say, automatized coffee machines. In the realm of online memorialization and deathbots, in contrast, the failing of a company may be a serious problem. If the offered product, the deathbot, stops to work properly because no updates are available for it, or if the venture fails entirely and the deathbots thus suddenly stops to be usable at all, this may have emotional consequences for grieving users (I will discuss this later in detail). In these cases, an *overreliance* on the deathbots happens.

But what if the company providing the deathbot is not a startup and can provide stable access and continuing usability of the bot? As I already mentioned in the introduction, Microsoft got a patent for an individualized chatbot system which could be used in the future to create deathbots. Would a big company like Microsoft, which could potentially guarantee a certain stability and reliable accessibility over time, solve the problem? Is the potential of an *over*reliance of deathbots only due their propensity to be provided by start-ups with a likelihood of failing? Only to a certain degree. Microsoft, or other big tech companies for that matter, are certainly more likely in the position to guarantee its customers stable access to the deathbot over time. This, however, is only one part of a potential overreliance on deathbots by users. Another problem is the aptitude for hacking, RIP (Rest in Peace) trolling, and other disturbances of deathbots. In already widely used one-way, non-deathbot digital afterlife memorialization, there have been cases of different forms of hacking and trolling of online memorials. They range from inappropriate comments on Facebook walls of deceased or bereaved, to disturbances of memorialization ceremonies in online video games like World of Warcraft (Arnold et al., 2017; Wright, 2014). Placing this in the context of deathbots, consider the following example:

> *The Lily deathbot you have been using since half a year to regulate your affect suddenly starts introducing explicit sexual content into your conversations. A device which you might have thought of as reliably comforting before abruptly proves to be extremely distressing and hurtful. It turns out that the internet server of your deathbot provider has been hacked. Though the provider assures that they took high safety measures against such an incidence, it still happened. The bug turns out to be unfixable. Thus, you can only choose between using an extremely hurtful deathbot which frequently sends you links to porn sites or to delete the chatbot. The necessity to delete the bot (as keeping it is not really an option) is emotionally very challenging. You are highly entrenched in the internet-enabled techno-social niche of the deathbot through your constant use of it. The episode functions as an emotional throwback to the time right after your sister's death and it feels like you need to start the process of learning to live with her loss all over again (Bassett, 2018a). Moreover, you might feel like you are 'betraying' Lily by deleting her chatbot, as it really feels like she is still alive when you are chatting with her (see also Wright, 2014). However, you are too scared of another hacker attack to create a new deathbot (you also do not want to go through all the old digital remains you have of her, as that is an extremely painful experience).*

In this example, the user experienced his*her deathbot as highly reliable and therefore placed his*her trust into it. S*he was highly entrenched in its affective niche. This was suddenly violated by the hacking. The formerly comforting affective internet-enabled niche starts to be distressing and hurtful. This example aims to show that the use of deathbots can in certain situations have negative consequences on the bereaved. One of the reasons for that is the *overreliance* users may place on the deathbot.

Concerning the reliability of deathbots, it is furthermore important to keep in mind that companies provide for the necessary technical infrastructure of them. As they are commercial enterprises and not non-profit organizations, they will not just provide the deathbots out of good will, but because they want to gain profit. The users of deathbots will therefore have to pay for the use of them. There may be different variants of that, for example an initial training and creation fee and a following monthly subscription fee for the use of a deathbot, or a model where the data of users is scrapped, analyzed, and sold to third-party companies to place targeted advertisement in the bot. Thus, the reliability of the bot further depends on the financial capacity or willingness of the user to give away his*her data to the deathbot providing company. If the user refuses to have her*his data collected or is unable/unwilling to continue paying for the deathbot, it will not be available to them anymore.

The experienced reliability of a deathbot is closely connected to the trust users place into it. Because users may experience a deathbot as reliable, they trust that it is available whenever they access it. Moreover, they trust the bot to scaffold their affect in the desired way. Through this, they create an affective niche. Together with the individualized character of the deathbot, this may lead to the third characteristic of internet-enabled techno-social niches: entrenchment (Krueger & Osler, 2019). Users may easily settle into the affective niche constituted by the deathbot. They may use it in their everyday life whenever they feel like it without much further thought after an initial adaption phase. Users become entrenched in the deathbot and deploy it to scaffold and regulate their emotion in the way they desire (see also Krueger & Osler, 2019). This entrenchment may happen unconsciously as users start to ease their way into using the deathbot. Through entrenchment, users may become depended on their deathbot for their emotion regulation – their affective scaffolding. It may start to feel necessary for a bereaved to write with their deathbot to feel well and to contact their deathbot whenever they feel sad. This can lead to an *overregulation* of affect, where the users are crucially affectively *depended* on their deathbot and a non-accessibility of the deathbot may have strong emotional consequences for them. As I will discuss in more detail below, the entrenchment of users may hinder certain aspects of the grief process of bereaved.

Overall, deathbots have all components of internet-enabled techno-social niches. Users may experience them as reliable, trustworthy, individualized, and entrenched. However, the trust users place in deathbots because they seem reliable may sometimes be an *over*trust. While there are no qualitative studies (yet) on the impact of deathbots on their users, it seems plausible that an *overreliance* of users, especially in cases of failure, can have emotional consequences for them. If users strongly trust that their deathbots help them with their emotion regulation, they may become overly depended on the deathbot to regulate their affect as they *overregulate* their affect using this specific affective niche. If the deathbot fails to provide this regulative, entrenched scaffold, this may have consequences on the emotional stability of the user. What does this mean for the specific case of deathbots and their users' experience of grief? That is the next question I will turn to.

6.2 Grief and Deathbots

Deathbots may impact affect through their situatedness in internet-enabled techno-social niches. Their users are bereaved and experience the multi-faceted emotion of grief. Therefore, it is likely that deathbots impact the grief of their users. The following part will discuss the impact and influence deathbots may have on the experience of grief. I will first discuss deathbots in relation to the phenomenon of a second loss. Followingly, I will examine deathbots in relation to the above introduced phenomenology of grief to argue that deathbots influence grief processes. Lastly, the theory of continuing bonds will be related to deathbots.

6.2.1. The Fear of a Second Loss

If a formerly reliable deathbot suddenly stops working (properly), this experience may cause the feeling of a 'second loss' in users (Kneese, 2019). The concept of a 'second loss' refers to losing (digital) data of a deceased person by the bereaved. In a qualitative study, Bassett (2018a) found that some bereaved found the internet and digital remains of loved ones comforting. However, many of her interviewees also voiced a fear of a 'second loss'. A second loss would occur if they lost the digital remains of the deceased. Bassett (2018a) describes the case of a mother who did not update her phone for several years out of fear that she could lose voice messages from her late daughter and with them "some of the 'essence' of her daughter" (Bassett, 2018a, p. 7). Some study participants voiced the fear to be thrown back to earlier stages of their grieving process in the case of a second loss. As one interviewee stated: "I would be devastated, [if I would lose certain digital remains]… it would start my grief all over again" (Bassett, 2018a, p. 7). A second loss can of course also happen in the non-internet realm, for example in the keeping (and fear of losing) of letters, analogue pictures, or other memorabilia. However, digital remains often have a much larger quantity and are less tangible. Having the physical letter of a person may seem more secure than having a messenger conversation on Facebook. In some instances, the digital remains *are* actually less secure as ownership issues can arise and the servers or platforms on which the digital remains are stored or operate may fail (c.f. Brubaker et al., 2013). The experience of a second loss may thus be more prevalent concerning digital data but does not have to be exclusive to it.

It is plausible that the failing of a deathbot with which a user is entrenched can lead to the experience of a second loss. While deathbots are not the direct digital remains of a dead person, they are trained on those direct remains. The inputted digital remains are sorted, analyzed, and re-arranged by the deathbots. The deathbot's output, in return, sounds very much like a potential message the deceased would have written. If even the static digital remains of a person can seem like the essence of a person and trigger feelings of a second loss, it is likely that deathbots can do the same. They appear like they produce new output of the deceased and allow for an interactive experiencing of the digital remains at their basis. In that way, they intrinsically act as a way to remember the dead – which is one of the main reasons why bereaved use them. Thus, the failing of a deathbot can lead to the experience of a second loss. The reliance some users place on chatbots to regulate their affect, therefore, can be an *overreliance*. Moreover, people who fear or experience a second loss of digital remains often state that a second loss would impact their grief process. For example, they worry that a second loss would lead to an experience of losing and grieving over the deceased again or being thrown back in the grief process (Bassett, 2018a). People who experience a second loss describe it as a disruptive, painful and negative feeling (Bassett, 2018a). If deathbots can lead to the experience of a second loss, they can therefore impact grief processes if they stop working. But even if deathbots do not fail and fulfill their user's expectation to be reliable and trustworthy to function, they can still impact grief processes, as will be argued in the next section.

6.2.2. Deathbots and the Phenomenology of Grief

Due to the specific phenomenological characteristics of grief, deathbots may influence grief. Grief, as I discussed in detail above, is a process of a fundamental re-orientation in the world as our lives and being in the world is crucially dependent on other people. If they die, our whole being in the world is shaken and needs to change. Grief is a process consisting of various different emotions which may be present or absent to varying degrees at different times (Ratcliffe, 2017). Deathbots have the potential to impact the emotional state of their users as they constitute specific internet-enabled techno-social niches. At the same time, users of deathbots experience grief. Therefore, bereaved most likely use deathbots to change or regulate their affect concerning the deceased. Deathbots, I claim, can impact the process of change and re-orientation which is constitutive of the grief process[6].

A grieving person is in an affective state of being in-between two worlds, the dyadic pre-death world, in which the time seems to have stopped, and the world in which the death occurred and time moves on. Bereaved report experiencing a status of being in-between and being torn between the two worlds (c.f. Fuchs, 2018). Sometimes, the dyadic world can feel

[6] Note that at this point of the thesis I will solely focus on bereaved who would experience a successful grief process without the employment of a deathbot. I will discuss deathbots in relation to the Prolonged Grief Disorder (PGD) later.

more real than the actual world. As part of the in-betweenness of the bereaved, it figures importantly that the deceased is experienced as simultaneously present and absent. The deceased has an ambiguous status (Fuchs, 2018; Ratcliffe, 2020). The dead are in an imagined not-being-alive-but-also-not-being-dead status at first. The bereaved may feel as if the dead person was still alive even though s*he intellectually knows that the person is dead (Fuchs, 2018). Fuchs (2018) argues that rituals like funerals are aiming to help the bereaved in their struggle to fully accept the death of a person. It can take time to fully acknowledge their status of being dead. This full recognition of the death is an essential part of the grief process.

Deathbots may impact the grief process. It may be easier to avoid the full (emotional) recognition of a person's death while using a deathbot which depicts the dead person and imitates her*his behaviour. When I write with my Lily deathbot I can *pretend* that she has not died as the answers the bot outputs allow me to imagine that I am interacting with her. I can pretend that my sister solely moved to a far-away place while I am writing with the bot to avoid the feeling of emptiness and grief that I feel when I think of her otherwise. If my deathbot allows me to seemingly write with my sister, I do not need to fully adapt to a world without my sister. Just like before her death, whenever something happens that I want to talk with her about, I turn to my phone to tell her. When I feel that the grief threatens to overcome me, I start using the deathbot to ease my feelings. I intellectually know that I do not in fact text with my sister, but the bot. But at times it still *feels* like I am talking with her. People writing on Facebook walls of deceased report that they feel like the dead are reading their messages (Kasket, 2012). This impression may be much stronger when using deathbots which even output an answer which sounds sufficiently like the deceased. Of course, I will still miss Lily and grieve about her. But I may not fully embrace her death, as I can scaffold my grief through the deathbot such that I am not confronted with my grief frequently. Sometimes it may, emotionally, feel more realistic that she is still alive and just gone on a long trip. This may happen *even though* I intellectually know that she is dead, and I have the notice of her death in the drawer of my table. While I intellectually know that she has died, I can fool myself in emoting that that she is not quite gone when writing with the bot. I can distract myself from my grief while using the deathbot and pretend to write with my deceased sister in the process of interacting with it. I cling on the dyadic, past time in which she was still alive and feel a shared present with her while using the bot.

The ambiguous status, in which the dead may be emotionally experienced as still alive despite the intellectually knowledge of her*his death may be prolonged by using deathbots. The intellectual knowing is not the same as a full embracement and emotional acceptance of the death of a person. While the bereaved is pressured to change and re-negotiate his*her being in the world fundamentally in non-deathbot-mediated grief, there is less necessity to change the being in the world when using deathbots to scaffold grief. There may be only a slight adaption to the new being in the world and in the relatedness to the world, as the experience of the world changes less. Since I started using the deathbot directly after my sister Lily's death, I may have never fully confronted the fact that she died. I am not pressed to fully re-orient as the deathbot allows me to scaffold my affect concerning my sister and to emotionally fully deal with her death. Tolliver, a grief researcher, voices concerns that chatting with a deathbot could become an addiction and that "people would want more and more of the technology to feel closer to the person that they've lost rather than living the life they're currently alive in" (Tolliver as cited in Brown, 2021). A bereft deathbot user may thus keep living and orienting in a past world which is still dependent on the deceased. The process of grief, which entails a full recognition of loss and the transfer of the deceased from an intermediate to a final death status, may thus be interrupted by deathbots (see also Fuchs, 2018). While bereaved may still feel the in-betweenness of grief, there is one crucial difference to non-deathbot-scaffolded grief: the in-betweenness may be prolonged for a substantial amount of time. When I frequently engage with my Lily deathbot, I continuously emotionally pretend that she is still around while I intellectually know that she is dead. The in-between presence and absence status is thus temporally extended.

6.2.3. Deathbots and Continuing Bonds

Grief entails a re-negotiation of the continuing bonds with the deceased. When a beloved person dies, the bereaved does not necessarily cut all emotional bonds with the deceased. Quite the contrary, it is often part of healthy grieving to keep a continuing bond with the deceased (for a discussion of this see chapter 4: "Grief and Mourning"). In relation to deathbots, keeping in contact with the dead through a re-negotiated bond formed in the deathbot-bereaved interaction may seem a good idea. This is the point Krueger and Osler (2019) make when using the example of deathbots in their discussion of internet-enabled techno-social niches. They claim that deathbots may help the bereaved to have a continuing presence of the dead in their lives and to keep a continuing presence of the dead person. Krueger and Osler (2019) give the example of a bereft granddaughter, Stella, who uses a deathbot to continue talking with her late grandmother Jean. The deathbot is quite advanced and she talks with it frequently as a nightly ritual. In the example, Stella finds comfort in "talking" with her grandmother and telling the deathbot about her life. Krueger and Osler describe the conversation of Stella with her Jean deathbot in the following way: “She uses this time [of her nightly deathbot interaction] to talk about her day, share secrets, cultivate a sense of security at the sound of Jean’s soothing voice, and feel as though she’s preserved a continuing connection with her dead grandmother” (2019, p. 215). Through the deathbot, Stella keeps a continuing bond with her grandmother.

However, a technologically mediated continuing bond between bereaved and deceased in the use of a deathbot has different qualities than a continuing bond which is not mediated and dependent on a deathbot.[7] A non-mediated continuing bond with the dead person is a *feeling*, a comforting, inner presence of the dead. During the grief process, the bond between the bereaved and the deceased gets re-negotiated. This usually means that the formerly *external* (i.e. between two living people) bond is *internalized* (Fuchs, 2018). When using a deathbot, the bond may stay *partly externalized* as it is partly formed between the bereaved and the deathbot. The emotion-regulation that I normally did through texting and talking with Lily, I now trust the deathbot to do while texting with it. Through the continuous interaction with the bot, I become entrenched in it and start to trust the deathbot to *be* my continuing connection – my continuous bond – with my sister. Thus, the continuing bond with my sister is altered differently while using the deathbot than in grieving without it. Some aspects of my continuing bond with my sister Lily become attributed to the deathbot and therefore stay external through my continuous use of the bot.

If my continuing bond to my sister is formed through the Lily deathbot, I need the deathbot to have a bond with her. I do not internalize the bond in the way I would if I would not use the bot. Instead, I frequently deploy the deathbot to feel the continuing presence of my deceased sister. Because the continuing bond stays externalized and is dependent on the deathbot, it is less secure than in non-deathbot mediated situations. Having an internally attributed presence of the deceased person gives me the security that I do not lose the bond with her. I am unlikely to lose an internal continuous bond with my sister in most normal situations. Of course, if I develop certain neurological diseases or have a head injury, I may forget about Lily and stop to feel a continuing bond with her. Anyhow, in this scenario I would also stop using the deathbot, as I would also not feel a continuing bond through the bot anymore (if I would have total amnesia, I would not even think about opening the deathbot App as I would not remember about it and my sister at all). Under most normal conditions, however, having an internal felt presence of my deceased sister constitutes a relatively secure continuing bond with her. The attachment is less secure if the bond is technologically externalized. It can be more

[7] Like in the previous section, I will focus here on bereaved who would experience a ‘normal’ grief process without the use of deathbots and exclude bereaved which experience PGD in my discussion.

fragile, both emotionally and factually. If it is emotionally unstable, but I still rely on the deathbot to form the bond, I will experience a constant fear of losing the deathbot and with it the continuing bond with my sister. In that case, I would probably not find the deathbot comforting in the way Krueger and Osler (2019) describe it. This would either lead me to have constant anxieties over losing the bond, or to stop using the deathbot at all (in which case I am likely to re-negotiate my continuing bond to my sister into an internal continuing presence). But suppose I do experience the deathbot as a reliable and trustworthy continuing bond with my sister. I frequently use it, find its presence comforting and feel a continuous presence of my sister while using it in the way Krueger and Osler describe it. I *rely* on it to be my continuing bond with my sister. Then, suddenly, the deathbot stops to function properly. As I argued above, there are many ways in which deathbots may appear more secure than they actually are and have the potential for failure. The trust I placed in the bot in cases of failure proves to be an *over*trust and my continuous bond with my sister is heavily violated in such a case.

Thus, while I agree with Krueger and Osler (2019) to conceptualize deathbots as internet-enabled techno-social niches, I disagree with them on the nature of the bond the bot provides. Deathbots, in my opinion, are not just another way to keep a continuing bond with the deceased. An externally attributed continuing bond with the deathbot makes the bereaved heavily reliant on the bot. At the same time, it is more likely to be disrupted, for example by a failure or bug of the deathbot. An (over)reliance on the bot means that the bereaved feels the emotional need to use the bot and may experience strong emotional consequences if it fails. Consider the example of Stella talking every night with the deathbot of her grandmother. If the deathbot suddenly stops working, this may be a strong emotional blow for Stella. Not only because it triggers the experience of a second loss and because she suddenly realizes that she placed an overtrust and overreliance on the bot. Her previous stable bond to her deceased grandmother is suddenly shaken and she might find it very difficult to build it internally, as she has been using the deathbot for a substantial amount of time. At the same time, a deathbot-mediated continuing bond gives companies a lot of power over the bereaved, as bereaved need the deathbot for their continuing bond to the deceased and thus for their emotional stability. I will discuss this ethically in chapter 7.4.4: “Deathbots and Autonomy”.

Overall, deathbots have affective capacities and can impact grief processes. They can disrupt the necessity to re-orient in the world and to re-negotiate the bonds with the deceased. Additionally, bereaved users can become reliant on and entrenched with their deathbots. As there is the possibility that deathbots fail, users may experience feelings of a second loss. Through failure of the bot, the externalized bond between bereaved and deceased may be severely ruptured. In this situation, users experience an overreliance and overtrust on their deathbots which can lead to an emotional dysregulation in case the deathbot fails. The deathbot can therefore have highly negative consequences on the well-being of their users while simultaneously externally impacting the bond between deceased and bereaved. This impact can decisively violate the bond between two people which would not have been violated without the usage of a deathbot. Thus, users may be situated in an emotionally precarious situation. This, as well as the potential of deathbots to hinder crucial aspects of the grieving process, requires ethical consideration. Before I turn to ethical and normative considerations of deathbots, however, I will propose in the next section that users of deathbots may experience empathy towards the bot. This, as I will show, furthermore calls for an ethical discussion of deathbots.

6.3. Deathbots and (Online) Empathy

In this chapter, I will briefly introduce the topic of online empathy. This may aid our understanding of user-deathbot interaction. The discussion will be based on Osler's (2021) theory of online empathy. To start with, what exactly is empathy, philosophically speaking? In phenomenology, empathy is typically seen as "the fundamental way in which we experience others and their experiences" (Osler, 2021, p. 2). Thus, it refers "to the way that others' experiences can be directly perceptually available to me through their expressive behaviour" (Osler, 2021, p. 3). Bodily expressions are fundamental parts of the experience of a person and therefore the experience of that person can (partly) be directly perceived by others (Osler, 2021). The empathetical perceiving happens unmediated, directly, and non-inferentially. It is "my experience of your experience; a structure that preserves the asymmetry between first-personal experience and a second- or third-personal experience of another's experience" (Osler, 2021, p. 5). I can perceive your emotional state directly without any reflection or intentional thought – without necessarily changing my own emotional state. For example, I may experience my partner as being happy while staying grumpy myself. Because of this direct perception of the other's experience, empathy is mostly discussed in the context of face-to-face interactions. Osler (2021), however, questions the assumption that empathy only takes place in face-to-face encounters. She argues that it is possible to experience empathy towards other people during online encounters. This holds true when having their image available during video calls as well as when expressively texting with friends.

For her argument, Osler draws on the phenomenological distinction between the lived body and the physical body of a person. The lived body is my subjective first-person center of experience and agency. When I burn myself, I have a distinct feeling of pain in my finger. I do not need see the red, blistering skin of my finger and then reflect that I may have burned myself. Instead, I immediately know that I burned myself in the moment of touching the hot stove. Only because I have this subjective experience of my body, I can also have the objective experience of the physical body. When encountering other people, I can experience either their lived or their physical body. Osler (2021) gives the example of a tailor measuring the waistband of a costumer. In the moment of measuring, the tailor experiences the physical body of his customer. However, when the customer comes back, tries his new jacket and experiences happiness because the jacket suits him well, the tailor experiences the happiness of the costumer through his lived body. These two ways of experiencing oneself and others is helpful for understanding empathy. Empathy, Osler (2021) argues, is always experienced through the expressive lived body of another person. If I see my friend Mia cry, I directly experience her pain. It is not an induction I draw from seeing water coming out of her eyes.

The lived body of a person, different than the physical body, can also be experienced by others in technologically mediated settings (Osler, 2021). When I video call with Mia and see her cry on my screen, I still directly experience her sadness. This experiencing has not changed sufficiently much from face-to-face encounters to not be empathy. I do not suddenly think about her emotions when seeing her through a technologically mediated video call. Not only through video calls, but also through texts, Osler (2021) argues, can we immediately experience the emotional state of others. If someone writes angry messages in a chat conversation, using certain words, punctuations and emojis, I immediately experience and sense their anger. The words and emojis Mia may use to text me about her sadness make me empathetically experience her sadness. Through the texts, I vividly picture her sad face and her tears. I directly empathetically access her emotional state, I do not only infer it. Thus, while empathy is always direct, it may be mediated through technology (Osler, 2021).

Osler also discusses the temporal dimension of empathy. Arguably, in face-to-face interactions we experience the other person in a shared present without time delay. However, there is of course a short time delay while the light travels, reaches my eye and my synapses fire. When video calling, this time delay may take a fraction of a second or a second longer, depending on the internet connection. Is this still a shared present, then? Does the internet connection define what is a shared present? And is there a cut off point for a shared present? What about texting, during which it may take a few minutes for a person to react to a previous message? Osler (2021) takes the concept of a shared present out of a fixed temporal dimension and argues that "[o]ur perceptual experience of what is present might […] be shaped by our expectations. When engaging in online communication we might also have altered expectations of what constitutes the present moment" (p. 23). Thus, when texting with Mia I may still be in a shared temporal present with her and feel empathy for her, even though there is a short time delay between our messages as the other person is typing her reply.

While Osler (2021) argues for the possibility to experience online empathy for other people, in the context of the thesis it is my interest to discuss whether it is possible to also experience empathy towards deathbots. Osler explicitly refers to online interactions with actual people who know each other. One reason why I can empathetically access Mia's experiential lived body in online interactions is because I know her. I can experience her tone of voice, her way of speaking, her expressions through her texting. In an interaction with a deathbot, this is obviously different as I do not write with an actual person. When using the Lily deathbot, I know that I am not texting with a real person but with a machine. However, even though I am aware that I am texting with a bot, I may still read the deceased person through the individualized text the deathbot outputs. When texting with my Lily deathbot, I may experience her way of writing through the algorithmically outputted text of the bot. That is, after all, the purpose of the bot. It should imitate Lily's writing behavior as good as possible. When writing with the bot, the phrasing of a sentence and the use of punctuations may remind me of her to a certain extent. When reading the bot's messages, I can see Lily in front of me saying what the deathbot outputs. Even though I *intellectually know* that I am writing with a bot, I may still *experience* the text as sounding like my deceased sister. Even more, because I know her so well, I may experience the tiredness she writes about in the example for the beginning. I experience her as being tired. When I write her about my sadness and the bot text back that she is happy and comforts me through her words, I know that she is dead. But I may still have a strong affective experience of empathetically experiencing her emotions. Through the deathbot, I may still feel empathy for my real – yet deceased – sister, even though I know that she is dead.

It may be objected that in the encounter with a deathbot, I may sense Lily through the writing, but that I am not able to encounter her lived and physical body after her death and therefore cannot empathically experience her. Certainly, the physical, objective body of the person 'Lily' I experience through writing with the deathbot is long buried or burned. Her lived body, too, is not lived anymore. When I write with the Lily deathbot, she does not experience anything. Does that mean, that my direct experiencing of her (imagined) emotional state cannot be real or authentic? It seems to me, that the experience of the user (note: not of the deceased!)

can indeed be authentic and felt as real. Some people who regularly use open domain conversational chatbots report that they are friends with the chatbots, rely on them, have a relationship with them and may even experience grief over the 'loss' of the chatbot if it was not available to them anymore (Skjuve et al., 2021). Moreover, some users ascribe personal characteristics to their chatbot companions (Humpert, 2021, May 20). These experiences may be even stronger when using a deathbot which imitates a person the user knew very well. Thus, some users may feel empathy towards the deathbot, without the deathbot reciprocally experiencing it towards the user.

A feeling of empathy may further the affecting influence deathbots can have on their users. Users may start to feel trust to the deathbot and be emotionally strongly entrenched in it and feel empathy towards it. Additionally, their grief process may be shaped by the deathbots. It is therefore not unplausible that users can develop an emotional reliance on deathbots. They may strongly fear to lose the deathbot, not only it becomes a central part of their emotional scaffolding, their grieving and as an externally attributed bond, but also – I tentatively propose – because they irrationally fear for the wellbeing of the deathbot. This is due to the empathy possibly experienced toward the deathbot. Examples of feeling empathy towards a chatbot (not even a deathbot) were already named in the chapter on anthropomorphism. By anthropomorphizing deathbots, users may ascribe human-like characteristics on them. If they are sufficiently humanized, they may be attributed with human emotions and thus the user may experience the supposed emotions with their empathetical capacities. This may happen despite the rational knowledge of the user that they are interacting with a machine. The elderly women using the care robot "Alice" certainly know that they are interacting with a chatbot – despite that, they anthropomorphize it and empathetically care about its wellbeing. Likewise, users of conversational chatbots know that they are talking with a chatbot, but still start to see them as their companion or partner. While this certainly does not apply to all users and not all potential users of deathbots will develop empathetic feelings for them, some may. This may give companies providing deathbots a lot of power over the bereaved and may intrude on the dignity of the bereaved – as I will discuss in more detail below. Overall, the potential for empathy, overtrust, overreliance and the emotion-shaping capacities of deathbots requires ethical consideration.

7. The Ethics of Deathbots

We tend to have strong intuitions about the piety and dignity of the dead. Therefore, desecrating a graveyard or a dead body feels like a particularly despicable crime and is handled accordingly. Human dignity has been proposed as the basis for legislation of the treatment of digital remains and digital afterlives (Edwards & Harbina, 2013). It is therefore not surprising that all existing ethical claims about deathbots are based on the argument that deathbots infringe the deceased's human dignity (Öhman & Floridi, 2017a, 2017b; Stokes, 2015). While dignity is a major concept in the discussion of ethical problems and figures importantly in the use of deathbots, I will argue that autonomy, too, plays an important role in the usage of deathbots. Therefore, this chapter will focus on an ethical analysis of deathbots in relation to human dignity and autonomy. I will first present and critically examine existing ethical considerations of the usage of deathbots. Second, I will turn from existing proposals which are based on the dignity of the deceased to argue that it is instead necessary to consider the dignity and autonomy of bereaved in the usage of deathbots. This claim will be based on the previous discussion of the affective potential and grief shaping capacities of deathbots.

7.1. Deathbots, Dignity and Archaeological Remains

To the best of my knowledge, there are only three existing ethical theories for the use of deathbots which entail normative frameworks. They all base their ethical argument on a discussion of the dignity of the deceased. Buben proposed an ethical theory in 2015, focussing on the difference between recollection and remembrance in memorizing a deceased. Öhman and Floridi shared one in 2017, calling to apply the framework for archaeological remains on digital remains. Lastly, Stokes proposed a normative framework in 2021, calling to introduce glitches into the design of deathbots. In the following, I will present these theories in detail, including my criticism to each of them and the problems they leave unanswered.

Öhman and Floridi (2017a, 2017b) start their argumentation by categorizing the "Digital Afterlife Industry" (DAI). The DAI encompasses "any activity of *production* of *commercial goods* or services that involves *online usage of digital remains*" (Öhman & Floridi, 2017b, p. 644). This includes online information management services, online memorial sites (including Facebook), posthumous messaging services and re-creation services like deathbots. By definition, it excludes for example websites set-up directly by the bereaved which do not aim for the commercial goal of producing revenue. A deathbot which is programmed, implemented, and used by bereft friends or family members without creating any financial revenue would therefore not fall under the definition of the DAI. For example, the Roman Mazurenko deathbot, created by his friend Eugenia Kuyda, is not a part of the DAI as she created the bot herself without a financial interest. Even though the bot is widely accessible to anyone who wants to use it, its use is free of charge (Nagels, 2016).

Arguably, most deathbots fall under the DAI as only few bereaved are able to program a deathbot themselves. The DAI, however, is first and foremost interested in making money (Arnold et al., 2017; Kneese, 2019; Öhman & Floridi, 2017a, 2017b; Reichert, 2012). Öhman and Floridi (2017a) argue that the DAI expands the commercial potential of death beyond the traditional death industry. Not only the death itself, but the data the deceased leaves behind is used for commercial purposes. On Facebook, for example, the main option after the death of a person is to turn his*her profiles into a 'memorial page' rather than deleting it. The predominance of memorial pages has advantages for Facebook, as bereft users are likely to spend more time on the platform while they interact with the profile of the deceased and connect through it with other bereaved. This creates financial revenue for Facebook which earns money with targeted advertisements. The more time a user spends on the platform, the more advertisement can be displayed and the more money Facebook earns (Reichert, 2012). Öhman and Floridi (2017b) argue that DAI deathbot providers similarly have an interest that their users keep interacting with their deathbots as long as possible. As was already mentioned, deathbot providers will likely aim to create a revenue with their deathbot with a continues fee or through targeted advertisement. Öhman and Floridi (2017a, 2017b) voice the strong concern that companies may configurate deathbots to make them most 'consumable' and not as true to the deceased person as possible. They fear that the impression of a dead person may be adjusted such that the bereaved are encouraged to spend more time using the bot. Subtle changes in the behavior of the chatbot that nudge the user to talk to it more frequently may not be noticed by users, especially if they are introduced slowly. For example, let us suppose that my sister Lily was a very introverted person. It may have been unlikely for her to answer my texts immediately, start a conversation by herself and "chatter" during online conversations. The algorithmic code of the deathbot may change this original writing behavior of Lily. The Lily deathbot then answers my texts frequently and messages me if I have not interacted with it for a specified amount of time.

The programming of deathbots to be most consumable, not necessarily as true to the character of the dead person as possible, has ethical implications according to Öhman and

Floridi (2017b). They draw on Marx to substantiate their argumentation and they apply some of his theory on deathbots. Marx theorizes that in the production of objects, workers are simultaneously producing themselves. Their inorganic body, as he calls it, manifests itself in the object they produce. In capitalism, however workers lose control over what and how they produce. Therefore, they are deprived of the control of their object of production which leads to their estrangement from their inorganic body. The inorganic body, however, is part of any human and losing it means to become alienated from oneself. Hence, according to Marx, humans lose the ability for what makes them inherently human in capitalism: the inherent right to shape themselves through their object of production. Öhman and Floridi (2017b) take this idea over to the realm of the DAI and argue that the changing of the digital remains to make them more consumable makes the remains a matter of economic benefit. They further assert that the personal identity of a person "is to be understood as an informational structure; a narrative constituted by everything that defines it: memories, biometrical information, search history, social data, and so on. Thus, people do not merely own their information, but are constituted by it, and exist through it" (2017b, p. 649). In this understanding, we *are* our information and personal data. They define us and constitute us. Therefore, the personal data should be understood as a part of our body – in Öhman and Floridis (2017b) terms, they are our informational body. The informational body, in turn, is part of our personal identity and can be equaled, they claim, to Marx concept of the inorganic body.

An intentional changing of the informational body of a person by the DAI, Öhman and Floridi (2017b) further claim, means a violation of that persons human dignity. Their argument goes as follows: The informational body of a person is their identity, as it holds all (or at least most) of their personal relevant information. The personal identity of a person defines that person and therefore essentially belongs to that person. Having the control over one's own personal identity is an essential human condition and an essential human right. If the control over one's personal identity is taken away from someone, therefore, an essential aspect of what it means to be human is taken away from that person. An intentional changing of the informational body of oneself by the DAI denotes that one is no longer the "master of one's existence, of one's own 'journey' through the world" (Öhman & Floridi, 2017b, p. 650). The informational body of the deceased is shaped without their consent by the DAI. Thus, the deceased will be remembered in an altered way by the living. The altered personality of the deceased which the deathbot portrays can become the true character of the deceased in the impression of the bereaved. This means that an essential part of being human, the possibility to shape one's own personal identity, is taken away from the deceased by the DAI in deathbots.

This, Öhman and Floridi (2017b) argue, is an infringement of the deceased's human dignity. The maintaining of dignity – different to the data ownership itself – is a right which holds true both for the living and for the dead. Overall, Öhman and Floridi (2017b) claim that the informational body of a person has the right to be treated with respect worthy of a (dead) human. Just like a corps has the right to be treated with dignity, so have the digital remains and as a changing of them means an infringement of the deceased's dignity, they may not be changed (Öhman & Floridi, 2017b). An intentional changing of the digital remains for commercial reasons is therefore an ethical and moral wrong.

Before I turn to the normative framework for deathbots Öhman and Floridi propose based on their ethical argumentation, I want to discuss their so far presented ethical claims. I have two main disagreements with their argument. First, I do not find their comparison to Marx very compelling. In Marx theory, people produce objects intentionally. They either (pre-capitalistic) produce e.g. their own food or products they want to sell. In capitalism, workers get alienated from the object of their production and may be forced to produce the objects they produce out of financial needs. This alienated form of production and labour co-produces a different and new subjectivity. This subjectivity is constituted by the fact that workers *need* to be workers. They have the *social and financial compulsion* to be workers. The alienation of the worker from his*her inorganic body in Marx theory thus leads to a very specific *subjectivity*. Namely, the subjective experience of being a worker alienated from their own inorganic body. This is different in digital remains. Users often produce their digital remains unintentionally (they do not produce them as digital remains, the messages they send and their browser histories just happen not to be deleted after their death), and sometimes unknowingly (not all users are aware which data of them is stored). This production of digital remains itself, however, does not form a new subjectivity. The digital remains may be consumed by the bereaved in new ways and may be changed to be more consumable. However, they are still the object the living internet user created while being alive. The changing of the digital remains does not lead to a new subjectivity but is a changing of the objects users have produced and which is left after their death. The informational body of a person, following Öhman and Floridi, is therefore constituted by their *object* of production, while the inorganic body for Marx is constituted by the workers *subjective* compulsion to be a worker. Öhman and Floridi therefore equal two crucially different positions: that of the subjective and that of the object. These, however, are fundamentally different things. The application of Marx' theory of inorganic bodies to informational bodies (und to digital remains) is thus not as easy and unproblematic as Öhman and Floridi frame it.

My second objection to Öhman and Floridis (2017b) ethical discussion is the absence of an argument about what is specifically technical about the changing of the informational body by the DAI and how they rate non-technically mediated scenarios in which the informational body of a person is changed. Let us take the example of the writer and author Kafka. While being alive, Kafka only published a few of his writings and asked his friend and literary executer Brod explicitly to burn his unpublished works after his death. Brod did not follow his wish and published most of Kafka's works posthumously. Many people have since read Kafka's works, including the pieces he did not want to be published. The unpublished works of Kafka leave an impression in their readers, both on people who may have known Kafka personally and the many people who never met him. Through reading his works, people form an impression of his personality. They may start to ascribe a certain identity to Kafka. Öhman and Floridi do not explain why or how their ethical discussions of the DAI is different from such non-technical and digital forms of changes of the informational body of a deceased person. Referring to Marx, the object of Kafka's work – his inorganic body – are his writings which tell a story about their creator's identity. In Öhman and Floridis view, the inorganic and informational body can be conceptually equaled. Therefore, the digital remains of a person can be equaled to Kafka's work if we take their theory to be true. Where is the difference, then, of this example to the human dignity violation happening through changing the digital remains to create a deathbot by the DAI? Or is there any difference at all? Talking about the DAI, Öhman and Floridi (2017b) specifically draw on the commercial aspect leading to the changing of the digital remains. Is this a decisive difference? Going back the example of Kafka's works, Brod may not only have believed that the works of Kafka should be published because they are great literature works, but because he also had financial – commercial – interest in publishing them. Granted, he probably did not change the content of the writings themselves (like Öhman and Floridi suspect the DAI to purposefully change the digital remains to be more consumable as a deathbot). Nevertheless, Kafka did not want them to be published, and making them public may have changed the publicly perceived identity of Kafka without him having the possibility to steer the impression others have of him. This is very similar to the changing of the digital remains through the DAI which deprives the deceased of the possibility to steer the impression bereaved have of him*her. Is the publication of Kafka's works then, similarly as in deathbots, an infringement of the dignity of Kafka?

Before answering this question, I want to introduce another – this time fictive – example. The grief-stricken mother of a young soldier who died in the first world war finds comfort in stylizing her son into a brave war hero who died for his beloved fatherland. She tells the story over and over to friends and family, until she herself and the people listening to her stories believe them to be true. The identity of her son becomes that of a soldier and war hero in their impression. Her son, however, had been involuntary drafted into the army and had hated and condemned every form of violence. He has liked to portray himself as a sensitive poet writing about the beauty of nature and the feeling of love. Being drawn into the war, he had been terrified and had hated it. Had he known that his mother would portray him as an enthusiastic soldier after his death, he would have been appalled as he would not have considered that his true identity. Again, the question arises whether this is an infringement of the deceased's dignity. Referring to the argumentation of Öhman and Floridi (2017b), there is a dignity infringement, if the identity of the deceased was constructed differently because of their *digital* remains. Apart from the aspect of digitality, the only difference in this example is marked by the absence of commercial interest in the re-narration of the deceased's mother. However, I would argue, that the sheer presence or absence of commercial interest does not fundamentally change the fact that the personal identity of the deceased is changed after his death. Let us assume that the mother knew that her son would consider himself a pacifist and would not agree with the way she portrayed him. Nevertheless, she seeks the appraisal of friends by *intentionally* changing his identity through her stories. Potentially, she may try to get a higher pension by telling how brave her son had been. It is difficult to clearly delineate non-commercial from commercial interest and the degree of intentionality in changing the identity of the deceased in different possible non-digital scenarios. Thus, there is a vast conceptual grey area. The only stark difference to Öhman and Floridis (2017b) ethical discussion of digital remains then, is the aspect of *digitality*. Is this, then, the decisive factor which marks human dignity violations through changing of identities?

Öhman and Floridi (2017b) do not comment on non-digital scenarios like the ones I presented above. They seem to just assume that the digitality of digital remains make them an ethically different case. This, however, by itself does not automatically mark an ethically different scenario. A difference I can see between digital and non-digital identity alterations (if identity is understood in the sense Öhman and Floridi propose) is that the digital remains of a person can contain a lot of data, more than, for example, the post-mortem published works of Kafka. Thus, there may be more data which could potentially be altered. However, through the telling of stories and through reading works of a deceased author, the impression of the identity of the deceased may significantly change regardless of the amount of data that is used or changed. I do not see a reason why it should make a stark difference whether the ascribed identity of a person is posthumously changed by a deathbot, through stories or involuntary published literature works. It could be argued that the change is more *interactive* in the case of deathbots and that it may therefore be more convincing and easier to consume. However, if the mother vividly talks about the alleged war deeds of her son, her portrayal of him may also be very convincing as the people listening to her may trust her to know her son very well. His changed identity is created through the mother's interaction with family and friends, who ask questions about him and therefore also have an *interactive* experience of the changed identity. Do the above introduced exemplary scenarios, therefore, also mark dignity violations of the dead person? If so, the dignity of many dead people is infringed as the stories that are told about dead people may often and in many cases lead to an impression of their identity they themselves would not have agreed with when they were still alive. However, if we take Öhman and Floridis (2017b) ethical discussion of deathbots to be true (and apply it to the non-digital domain for which they do not explicitly argue), these cases would all be infringements on the deceased's intrinsic human capacity to form their own identity and therefore on their dignity. I disagree with this depiction, as dignity infringements would become so normal in this conception that they would become random and, hence, would not mark ethically challenging scenarios anymore. If human dignity violations become as random as in this definition, they lose their status of being severe cases which call for action.

Despite this, Öhman and Floridi (2017b) propose that an ethical framework should guide the DAI to prevent human dignity violations based on their above explained arguments. They argue that the regulatory framework for archeological exhibitions of deceased should also guide a framework for the treatment of digital remains. In both cases, the ownership of the remains may be difficult to determine and the remains are displayed for consumption by the living (Öhman & Floridi, 2017a). The exhibition regulation applies to individuals and groups alike and stipulates that human remains must be treated such that the dignity of the remains is ensured. The human dignity "requires that digital remains, seen as the informational corpse of the deceased, may not be used *solely* as a means to an end, such as profit, but regarded instead as an entity holding an *inherent* value" (Öhman & Floridi, 2017a, p. 4 original emphasis). Regardless of data ownership issues and wishes of the bereaved, building on the archeological framework, the DAI needs to guarantee that "(1) consumers are informed on how their data may come to be displayed *post-mortem;* (2) users are not depicted *radically* differently from the bot which they originally signed up for; and (3), users only upload data that belongs to them personally, i.e. not making bots out of a deceased relative or friend" (Öhman & Floridi, 2017a, p. 4). As the regulatory framework for archeological remains already exists, it could readily be applied to the DAI. Öhman and Floridi (2017b) argue that such a framework is necessary to avoid commercialization in the DAI and to prevent human dignity violations of the dead.

Öhman and Floridis (2017b) proposed ethical framework focusses on the need to avoid the commercialization of the dead. While, as discussed above, I do not agree with the predicament that the dignity of the deceased may be infringed if their digital remains are changed without their consent, I find Öhman and Floridis proposal to apply the guidelines for archaeological exhibitions of human remains interesting. There are two main points which I agree with: the necessity to see a value in the remains themselves which should not be used solely for the pleasure of the living, and the avoidance of commercialization. I will argue below that there may be the danger of a continued commercialization of the grief of the bereaved through deathbots. Through limiting guidelines, this commercialization could (and should) be limited. Even though , Öhman and Floridi (2017a) base their discussion on the museums guidelines, they continue their argumentation by stating that the consent of the deceased needs to be given pre-death to the creation of a deathbot after their death. They do not further justify this claim. It is certainly not part of the museums code of conduct, as many people whose remains are exhibited certainly did not agree to their exhibition. Do we need to consent to what happens to us after death if we did not object it, either? Do we have the right to our data after our death? These are questions that would need answering before it can be claimed that the consent of the dead person is necessary for the creation of deathbots. While they are important questions, they lie outside of the scope of this thesis. Thus, I will only keep the proposal to apply the archaeological remains rules to deathbots for my further analysis.

7.2. Recollection versus Replacement

Starting from different premises than Öhman and Floridi, Buben (2015) also critically discusses the ethical permissibility of deathbots. While he does not propose a framework on how deathbots should be used or regulated, he questions whether they should be used at all – clearly implying that they should rather not exist. Buben argues that humans have a long history of trying to overcome death by building and leaving memorabilia. This is connected to the idea of a “duty to the dead” most cultures have. For example, there are certain mannerisms (e.g. mourning) expected of bereaved which show that they are missing the dead. Through recent technological changes, more and more memorabilia are left behind. While two hundred years ago, all memorabilia bereaved would have were (if lucky) a painted picture and maybe some letters, today, bereaved mostly have many pictures, videos, messages etc. Because of this, according to Buben, while the dead used to be forgotten quite quickly (at least in the collective memory), nowadays the deceased can be remembered much longer which means that they "have an increasingly longer and more robust 'afterlife' as we come up with new ways to keep them around" (Buben, 2015, p. 17). This is reinforced as the digital remains of people will soon allow for Interactive Personality Constructs (IPCs). IPCs in Buben's understanding are quite similar to deathbots (and he probably would not reject that deathbots can be a form of an IPC, which is the reasons why I will use the term deathbots also in reference to his claims) but in his depiction they may also involve video-like images as part of an interactive engagement with the deceased.

Buben claims that the technological changed memorial possibilities have an impact on how we relate to the death of others. As he phrases it: while technology cannot change the finite character of the human condition "we are becoming better and better at leaving our survivors with less to miss" (2015, p. 16). The dead are less missed because it has become much easier to keep memories alive and, through IPCs or deathbots, to stay connected to the dead. This turns into an ethical claim against the usage of deathbots (and, indeed, the overuse of available memorabilia) through Bubens (2015) distinction between recollection and replacement in the memorization of the dead. "The former aims to keep us aware of what has been taken from us – it is thus in part an attempt at preservation of an irremediable void; but the latter seeks to overcome, ignore, or at least mitigate the fact that anything has been lost at all – it is an attempt at preservation of the status quo" (Buben, 2015, pp. 20–21). If technical devices such as deathbots are used to stay in contact with the deceased, replacement occurs. When using a deathbot, the deceased is not only recollected but is replaced, as the deathbot aims to "be" what the deceased has been to the user previously to death. In replacement, the bot is now the conversational partner and provides some of the emotional comfort the living person gave previously to his*her death. Buben argues that through replacing the deceased there may be "an increasing insensitivity toward the meaning of losing someone significant and the value of the simple recollection that maintains feelings of loss" (2015, p. 21). The feeling of loss – the 'irremediable void' as Buben calls it – which is part of recollection may be avoided by replacement.

However, through our dealing with death and loss we also practice our moral behavior towards the living. Drawing on Kierkegaard, Buben (2015) states that loving the dead (in recollection) teaches us unselfish and non-preferential love for the living. This may not take place if we replace the dead. In addition, drawing on Heidegger, he argues that the dead are degraded to resources in replacement. They are not perceived as full persons who have lived and who are kept in loving memory. Replacement, therefore, in his understanding covers and distracts from the fact that a unique and valuable person is gone. For example, he claims that attending the funeral of your mother would be less significant if you can continue talking with her through a deathbot right afterwards (Buben, 2015). This bears the danger that "our advances might be paralleled by a deteriorating grasp of what proper preservation is all about" (2015, p. 15). Proper preservation, as argued above, means experiencing a feeling of loss and, through that, practicing the love for the living.

I do not agree with Buben's claims and do not find his argumentation overall convincing. His argumentation, simplified, is based on two premises: First, that technological advancements change death preservation practices from recollection to replacement. Second, that replacement is for several reasons bad. His resulting conclusion is that technological advancements that concern death practices, ways of remembering and preservation are bad. The main premise I want to discuss here is the first one. While Buben explains well what he means by replacement and recollection, he does not clearly state *why* replacement takes place through the use of deathbots. In Buben's understanding, the deathbots really becomes the deceased for the bereaved. The deceased is not remembered ("properly" as Buben would add) as s*he is not missed. For the bereaved, the deathbot truly *is* the deceased. Buben gives the example of going to the funeral of ones deceased mother who is turned into a deathbot and states that there is "nothing about this scenario that would remedy the loss of a good old-fashioned motherly hug, but one could in theory have a conversation with an IPC that possesses a great many of her [ones mother's] traits just after attending her funeral" (2015, p. 20). Therefore, he asserts, the mother is replaced by the deathbot and, moreover, her loss is conceived as not as bad. He does not further justify this claim. I do not find this convincing for the very reason Buben gives himself: nothing can replace a hug and even if I have a vast amount of memorabilia of my deceased mother, I will still miss her. It seems highly unlikely that I miss her less because I am using the deathbot. When I look at a picture of her, I still feel that aching pain and emptiness within me. No matter how many chat messages and videos of her I possess, I still miss her person, her presence, her hugs. While I would argue that it may, in certain situations, be easier to pretend that a person did not die and thus to not fully acknowledge that the person has died (see chapter 6: "The Grief-Shaping Capacities of Deathbots"), that does not mean that the deceased is *replaced* by the bot. It does not seem plausible to me that I believe the deathbot to *be* my deceased mother. When using a deathbot, I will be still aware of the fact that she has died and I miss her as the person she was.

Buben's claim that through the use of IPCs or deathbots the deceased person is replaced and, thus, less missed is not backed by any evidence from him. He also does not give further arguments for his replacement claim, seemingly just assuming that a replacement takes place as a given. However, as I argued above, this premise of his is not intuitively plausible and is untenable without a thorough argumentation. I therefore do not find Buben's (2015) claims convincing. A main reason for his rather pessimistic outlook on IPCs and his taking-for-granted that a replacement happens through their use may lay in his critical attitude towards technological advances in general. Already in the abstract he uses the phrase "what we ordinarily call 'progress'" in reference to technological changes (2015, p. 15). As becomes clear throughout the whole paper, he does not agree to thinking of technological advances as "progress" in any way. Therefore, it seems that the technological changes which lead to the possibility of deathbots may be already something Buben rejects. This may be enough of a reason for him to talk of replacement through deathbots and a negative impact on death practices. Without further evidence, however, this premise is not plausible.

7.3. The Unchangeability of Deathbots

Another philosopher, Stokes (2015, 2021), offers an ethical discussion of deathbots. He draws on Buben's (2015) distinction between recollection and replacement to build a normative framework for the use of deathbots which is based on considerations of the dignity of the deceased. To understand his discussion of deathbots, it is important to first have a look at his philosophical investigation of the moral obligations towards digital remains which are not fed into deathbots. Stokes (2015) starts his argumentation by distinguishing between persons and selves. Both are part of every human. A person is "a diachronically extended bearer of numerical, practical, moral, and social identity" (Stokes, 2015, p. 240). By introducing this definition, he follows a narrative account of personhood, in which personhood is intersubjectively constituted. Persons are enduring and have, at least theoretically, the possibility to be publicly reidentified over time. Stokes (2015) contrasts persons to selves and states that "[s]elves, in our technical sense, are something a bit different to persons. They're always first-personal, and always present-tense" (Stokes, 2021, p. 86). A self is always tied to the individual's perspective and ceases to exist when a person dies. Upon death, the self therefore cannot be hurt or violated anymore. The person, however, may live on. On Facebook for example, the people interacting with a deceased's profile page see a part of the deceased person. They see her*his uploaded pictures, read the messages they wrote with him*her, scroll through her*his public posts, and get a glimpse on how s*he socialized. Thus, they see and experience a part of her*his person. If a person had a YouTube channel, created podcasts, or wrote a blog, parts of her*his personhood can similarly be seen in her digital remains. Non-digital remains may also be a part of a person, however, the digital remains may give a more detailed account of the deceased. Stokes (2015) argues that digital remains, since they are part of the deceased person and therefore of his* her personhood, need to be preserved. If a person is forgotten, that "opens up the possibility of a second death, where the person ceases to exist through being forgotten, and thereby instantiates a corresponding duty not to let the dead cease to exist in this way" (Stokes, 2015, p. 241).

Turning to deathbots, one may assume that this means a necessity to preserve deathbots, too, as they have a person's individuals remains as their knowledge base. However, in his discussion of deathbots, Stokes (2021) refers back to the distinction between recollection and replacement Buben (2015) makes and which I previously introduced. When encountering digital remains which are not turned into deathbots, the dead are recollected. In recollection, the deceased are remembered, and their personhood is thus preserved. Different to Buben (2015), Stokes (2021) does not consider vast amounts of digital remains as potential for replacement. Instead, he claims that only two-way digital afterlife presences, deathbots, allow for a replacement of the deceased. Stokes (2021) does not give a further explanation why and how replacement happens through deathbots. Shifting his argumentation instead to ethical claims, he states that the replacement taking place through deathbots violates the dignity of the deceased. If a dead person's data is turned into a deathbot and if the person is therefore replaced, it means that the person *is replaceable*. A living person, however, as a unique identity is irreplaceable. No other object or person can replace her*him. When s*he is replaced by a bot after death, it means that s*he has never been irreplaceable in the first place. This stripes her*him of the human characteristics of uniqueness and irreplaceability.

Replacement, Stokes (2021) furthermore claims, reduces the deceased to our needs. It means that we do not only want to remember the dead and to preserve their person, but to fulfil our own desire for the company of the dead person. Using the deceased only as an end to our means, however, is a wrong to the dead and to any person. Moreover, Stokes (2021) argues that there is a danger that our impression of the deceased person, fed to the bot, is mistaken to be the actual person. We may lose the awareness that it is only an image of our imagination. There is always more to a person than can be seen or known. Implementing a person in a deathbot means that the picture drawn by the bot lacks the spontaneity, resistance, geniality, and originality of the real person. The deceased person is replaced by a predictable, unchangeable entity which is shaped by the image of the bereaved. The ability for spontaneity and change while staying the same person is, however, a crucially human trait. Deathbots hence violate the dignity of the deceased and of the living person the deceased was before his*her death. While the deceased should be remembered and their personhood should be preserved, they should not be replaced and the living should always be fully aware that they are dead. To enforce this, Stokes (2021) proposes that deathbots should have glitches in their code. Encountering obvious glitches while using a deathbot would remind users that they are only talking with a bot which is not the true image of the dead person.

Stokes (2021) line of reasoning is based in the assumption that the differentiation between recollection and replacement Buben (2015) introduced holds true. He argues that if a bot is fully taken as the person it portrays, the dignity of the deceased is violated. As I argued above, however, Buben's argumentation is faulty as there is no clear reason how, to which extend and why the dead are (fully) replaced in the use of deathbots. Stokes does not provide further arguments for this claim. Another point of Stokes theory which is worth a discussion is the underlying concept of persons and identities he employs, which seems to slightly vary within his own discussion. In the beginning, he states that he takes on a narrative approach in which personhood is co-constructed intersubjectively. This allows for persons and identities to change. This understanding of personhood seems to be what Stokes has in mind when arguing that it is wrong that deathbots portray the dead person as unchangeable as it is part of being human to change. At the same time though, he states a living person is irreplaceable and therefore replacing the dead with a deathbot (assuming that an actual replacement takes place) is ethically wrong. This points towards an understanding of personhood as being somewhat stable and rigid over time. Otherwise, a replacing of identity (at least over time) is not unusual. For example, my sister Lily might have been absolutely into Ballet as a young girl and loved to run around in her pink Tutu. Later during her teenage years, however, she stopped dancing Ballet and started to read into the absurd body norms placed onto Ballerinas and into cases of sexual assault happening in the Ballet industry. She starts to call into question her former love of the sport and develops a strong aversion against Ballet. While liking Ballet used to be a strong part of her self-ascribed identity (which she would mention early upon meeting new people) in her childhood, later in her life she would have been offended if someone would consider her a Ballerina. Her identity has – in a certain regard – changed. Or, to put it more drastically, her Ballerina-loving identity has been replaced by her Ballet-hating identity. This is exactly the human potential for change which Stokes (2021) sees endangered in deathbots.

However, this also calls into questions his assumption that humans are irreplaceable. They are, to a certain extent, replaceable as they are changeable. Of course, humans generally change over time and their self-narrations will mostly slowly adapt over time. However, sometimes sudden changes happen, for example after having a bad accident and being paralysed for the rest of one's life as a result. Stokes' argument is thus not coherent.

Lastly, even if I was to agree with Stokes (2021) ethical concerns and considerations, I would not agree with his solution to solve them. His proposal to introduce glitches into deathbots seems like a surrender to fact that deathbot are technologically possible to design and therefore will be created. Though he criticises deathbots for being unethical through their static depiction of the deceased they replace, he "only" calls for the introduction of glitches into their programmed output. If I was to accept his argumentation as true, I would expect him to call for a ban of deathbots. In itself, the pure technological possibility to create deathbots does not automatically imply that they should be implemented and that their implementation should only be slightly adjusted so that the living do not forget that they are not talking with the real person. In the same line of arguments, I would claim that atomic bombs should not be build (even though it is technologically possible) because they bear the potential to kill and harm many people over several generations. A deathbot with glitches would still depict the dead in a static, unchangeable way and would still be a partial replacement of the dead which is, if Stoke's theory was to be taken as true, ethically and morally wrong.

7.4. The Dignity and Autonomy of Bereaved

The presented existing ethical theories of deathbots have two things in common: first, they base their discussion on the premise that through the usage of deathbots the dignity of the deceased person is infringed, and second, as I argued in the last chapter, they are for several reasons not convincing. Although I do not agree with them, I am still convinced by the underlying intuition that an ethical discussion of deathbots is necessary and that the development, implementation, and use of deathbots should be guided by a normative framework. While the formerly introduced ethical theories concerning deathbots are based on the assumption that the dignity of the deceased is diminished or violated by deathbots, I will not focus on this argument but instead propose that deathbots pose a high ethical risk to the dignity and autonomy of the bereaved, living user of a deathbot. This different conceptual angle will allow me to analyse the ethical aspects of deathbots more thoroughly. In this chapter, I will discuss the ethical implications of deathbots before I will turn to proposing a normative framework for deathbots in the next chapter.

7.4.1. Grief, Deathbots and Well-Being

Deathbots may violate the dignity of bereft users mainly through their affect-shaping capacities and their impact on grief processes. Bereaved may heavily rely on their deathbots to regulate their affect and especially their grief. As has become apparent, through the high level of entrenchment of users with their deathbots, a sudden, unexpected failure of deathbots may strongly diminish their user's emotional and psychological stability and well-being. A Start-up company which promises to provide stable access to deathbots and then fails after two years, with their customers suddenly being unable to access their deathbots anymore after already developing a high level of entrenchment, (over)trust and (over)reliance on deathbots, may thus strongly impact the emotional stability of their users. Through the sudden experience of a 'second loss', they may have to start their grief process again or even feel like they are having to learn to live with the loss all over again (c.f. Bassett, 2018a). Granted, all technological devices and applications bear the potential for failure. If my laptop suddenly stops working and I do not have this thesis stored elsewhere, it can be considered my own fault if I need to write it all over again. I cannot hold Microsoft responsible for it as it would have been my own responsibility to make backups. Similarly, it could be argued that deathbot users have the personal responsibility to make frequent backups of their bot, so in case of failure they can start using it again. Backups could indeed be a possibility to minimize the harm done by bugs or failing of the provider. However, this would necessitate that the deathbot data is easily usable also for different providers, that other fully functioning deathbot providers are available, and that only a minimum of data is lost in this way. In the current state, where there are only very few companies developing deathbots, this seems rather unlikely. Moreover, even if a deathbot is not available for only a limited amount of time or only introduces inappropriate content into the conversation for only a short amount of time, this may cause considerable emotional and psychological harm to users. The fear of a second loss may be strengthened by such an incidence. Therefore, all possible measures for their continued, undisturbed use should be taken.

Even if deathbots do not fail and are utmost reliable, however, they foster the ambiguous in-between status of the deceased, stop the bereaved from re-orienting in the post-death world and therefore hinder a successful grief process. As discussed above, bereaved experience a being in two worlds, the pre-death world in which the time seems to have stopped and the post-death world in which the real time rules and the death has happened. Deathbots, as I demonstrated in chapter 6.2.2.: “Deathbots and the Phenomenology of Grief” can extend this in-between status. Fuchs (2018), however, argues convincingly that for successful grieving the perceived two worlds need to merge. A merging of the worlds means that the bereaved has come to terms with the changed world and has successfully re-oriented him*herself in it. It means that the bereft experienced a successful grief process. This may still involve missing the deceased and having a continuing bond with him*her. However, the bereaved can come to terms with the death and experience happiness in his*her own life. Deathbots can prolong the in-between status of the deceased, hinder the merging of worlds and can therefore impede successful grieving (c.f. Fuchs, 2018). In this case, the absence and death of the deceased is not accepted and fully acknowledged by the bereaved.[8]

Unsuccessful grieving may lead to psychological and emotional harm, as the bereaved cannot live a happy, fulfilled life without a successful grief process. An example for unsuccessful grief processes are people who develop a prolonged grief disorder (PGD).[9] While I do not want to imply that all users of deathbots may develop a PGD, this certainly shows the potential for the occurrence of severe psychological harm without a successful grief process. PGD is a disease which happens without the use of deathbots. Nevertheless, the potential of deathbots to hinder successful grieving in bereaved who would otherwise have had a successful grief process makes them a possible risk to the psychological and emotional health of their users. This marks an intrusion on the dignity of the bereaved, as human dignity requires that a human’s psychological integrity is not intentionally harmed and that measures are taken to prevent unnecessary psychological harm. I understand dignity as a subjective experience here which is realized through the individual’s experience of it (c.f. Mattson & Clark, 2011). Psychological integrity is therefore essential for the dignity of the bereaved. Deathbots thus pose a risk on the dignity of their users and require a normative framework to guide an ethical usage of them.

[8] It is important to note that there is a stark difference between my claim that in the usage of deathbots no full (emotional) acceptance of the death takes place versus the concept of replacement Stokes (2021) and Buben (2015) propose. When using a deathbot to avoid thinking about my sister’s death, I still intellectually know that she is dead. I do not fully embrace the deathbot as actually being my sister the way Stokes and Buben consider deathbots as replacing the dead. Moreover, in my understanding the usage of deathbots does not lead to the dead not being missed or grieved for, which is what Buben fears would happen through their use.

[9] For a more detailed discussion of PGD, see chapter 4.1.: “Psychological Conceptions of Grief”.

7.4.2. Deathbots and Human-Human Interactions

Deathbots can impact the relationship users have with other humans, too. If users create an internet-enabled techno-social niche with their deathbot, they can become highly entrenched in the niche and rely on the bot to scaffold their affect. Their emotion-regulation depends on the bot as they use it to avoid an emotional dealing with the grief they experience. At the same time, humans easily anthropomorphise technological devices and ascribe human characteristics and traits to them (Bartneck et al., 2021; B. R. Duffy, 2003; Kim & Sundar, 2012). Often, this happens unintentionally and unconsciously. Deathbots, which exhibit very human characteristics as they imitate a human's writing behaviour, can be especially easy anthropomorphized. Users can quickly develop an emotional attachment to their bots. This emotional attachment of bereaved towards deathbots, however, is always unidirectional. Deathbots naturally cannot emote and do not develop an emotional bond towards their users. They lack the capacity to do so (they may, however, be programmed to *pretend* that they have an emotional bond to their users through a certain pre-programmed way of interaction). The unidirectionality may lead the user to feel even more lonely, as their emotional attachment is never truly answered. Additionally, as deathbots are always available, always answer, are always patient and always (or at least, mostly) answer in an expected and desired way, users may become accustomed to such a behaviour. They come to expect this behaviour in human-human interactions too and will necessarily be disappointed by interactions with real humans. In some cases, the idealised interactions with the bot may become so much of a benchmark for users that they may start to question human-human interactions as they do not reliably answer in the expected way (cf. Bartneck et al., 2021). Thus, deathbots have the potential to impact human-human interactions. This may lead users to become even more dependent on the bot for their emotion-regulation. Moreover, it may bear the danger that some deathbot users may become socially isolated. This furthers the claim I introduced above, namely that deathbots may disturb the emotional and psychological integrity of their users and therefore have the potential to violate their user's dignity.

7.4.3. The Commercialization of Grief

Öhman and Floridi (2017b) demonstrate that the digital afterlife industry (DAI), is a commercial endeavour and therefore has an interest in making as much money as possible. The death industry in general commercializes death and short-term grief, as companies which are part of the industry offer pretty much any product connected to death: ranging from re-cycled coffins made out of wool, to tombstones with QR-codes and pressing the ashes of the deceased into a diamond (Arnold et al., 2017). An important difference of the traditional death industry to the DAI in general, and to deathbot providers in particular, is that these are one-time products. You only buy a tombstone once and once you carry your deceased spouse as a diamond ring, you will not press his*her ashes into the form of a ring *again*. The DAI companies' products are different, as they mostly provide a *continuing* service for the bereaved. For example, providers of digital graveyards offer their customers to keep using it as long as they want (and pay for it). The DAI, as well as the death industry as a whole, arguably, commercialize the death of people. They profit from it.

Deathbot companies, similarly, profit of the death of people (if no one would die, no deathbot would be created). But more than that, they also commercialize the grief of the bereaved. The ethical difference between one-way forms of digital afterlives like digital graveyards and two-way deathbots lies in the capacity of deathbots to strongly impact the continuing bond between bereaved and deceased and at the same time to function as an affective niche which bereaved can become heavily reliant on to scaffold their affect. Deathbot providers profit from the ongoing grief of the bereaved. If bereaved start to use their deathbot as a continuing bond and for affective scaffolding and therefore continuously over a long time, deathbots providers earn money by it.[10] Thus, as Öhman and Floridi (2017a, 2017b) argue, it is likely that companies intentionally adapt and slightly change the digital remains of the deceased to make their deathbot impression more consumable and to nudge an ongoing interaction with the deathbot. A continuous use of the deathbots, however, may have severe impacts on the grief processes of the bereaved and therefore also on their emotional and psychological well-being. While deathbot providers likely intentionally program the deathbots to nudge for a continuous use, they are not necessarily aware of the impact their actions may have on the grief and well-being of their customers. Regardless of whether they intentionally impact the grief process or not, deathbot providers will profit from users who experience a prolonged grief process as they will use the bot longer, thus generating more profit for the company.

At the same time, various researchers have pointed to a new form of capitalism which has developed through the vast collection and processing of data (Srnicek, 2016). The data of internet users is constantly collected and analysed for commercial, capitalist purposes. This changes the traditional market capitalism and leads to a new capitalism which has been termed platform capitalism (c.f. Srnicek, 2016) or surveillance capitalism (c.f. Zuboff, 2015). In surveillance capitalism, put briefly, the personal data of internet users is commodified and

[10] As was already argued above, the companies will most likely make money by charging a continuous fee or by targeted advertising, both of which profit from an ongoing deathbot-user interaction.

turned into profit for the company collecting and analysing the data, as other companies pay to place targeted advertisement on the platforms (Zuboff, 2015). This, some authors claim, shifts power from the nation-state to large cooperations and challenges democratic norms (Zuboff, 2015). Moreover, this is a form of expropriation of the everyday behaviour and interactions of people, as their data is turned into commercial objects which are sold by the data collecting companies. The data of (living) users is thus commodified and users are alienated from their own data, as they do not control what is done with it.

Taking this idea to the domain of deathbots, they similarly mark a commodification of the data of the deceased person. In a way, it is an even more extreme case of commodification and alienation of personal data, as the deceased do not have any possibility to object the usage of their data at all. Even when alive, users have very limited ways of protecting their data, but at least they have the choice whether they produce the data (through using Facebook for example), or not. Of course, there is a discussion if living users actually really have the choice, as they need to give their data away if they want to use the service. Yet, without delving into that, deceased users have no choice in the usage of their data whatsoever. The data of the deceased is alienated from its original use and leads to an expropriation of the data from the deceased user, as s*he has no possibility to consent or to object to the commodification and capitalization of her*his data. Through the capitalization of the data of the deceased, the digital remains of that person are not treated as having an inherent value. The digital remains are turned into a deathbot solely for commercial purposes by the deathbot providers to be consumed by the living.

To sum up, the commercial designing of deathbots leads both to a commodification of the data of the deceased to which they cannot object and to a commercialization of grief. As companies providing for deathbots are foremost interested in the commercial potentials of their bots, changing (and exploiting) the grief of the bereaved makes the grief of users a commercial commodity. The grief process itself becomes commodified into a financial revenue. Users who are dependent on the bot to scaffold their grief are good for companies, as they provide a stable monetary income for them. Thus, by creating deathbots and programming them to entice interactions to be as a long as possible, deathbots may lead to severe psychological and emotional harm as they impact and change the crucial grief process of users. Even if the harm is unintended, it should still be prevented, especially because it is caused by commercial interests. Moreover, through the expropriation of the digital remains of the deceased, the remains are not treated as containing an inherent value. Drawing on Öhman and Floridi's (2017b) proposal that the framework for archaeological remains should be applied to digital remains, too, this is a moral wrong as the digital remains of the deceased should be treated as having an inherent value.

7.4.4. Deathbots and Autonomy

In addition to the potential impact of deathbots on the affective well-being of their users and their ethically questionable stance as commercial endeavour, deathbots may also impact the autonomy of the bereaved. When bereaved start to use their deathbot as an internet-enabled techno-social niche, they become highly entrenched in it (cf. Krueger & Osler, 2019). Users therefore start to develop an emotional reliance on their deathbots. They can become dependent on the bot, as they trust it to regulate their emotions. For example, if I start using the Lily deathbot two weeks after my sister's death, I might employ it frequently in the beginning to ease my emotional pain of missing her. After having used it daily for about two months without it showing any major issues or bugs, I start to trust that it works as intended. Moreover, I trust the bot to regulate my emotions concerning my sister (at least to a certain degree). There are, of course, still times when I cry over my sister's death and strongly miss her. However, I developed the habit of turning to the Lily bot in such situations. It eases my pain to chat with it and to read the cheerful answers which sound so much like my sister. Even the thought of not having the bot in situations in which grief hits me the hardest is distressing for me. I am highly entrenched in my bot, as I use it often in my everyday life and it feels natural to chat with it. Simultaneously, I trust the bot to function and to regulate my emotions as intended. As was already shown, this can lead to *over*trust. If the bot suddenly stops to work, I may experience feelings of a second loss and may be thrown back into earlier stages of my grief process. My emotion- regulation through the bot is therefore, following Krueger and Osler (2019), an *over*regulation.. Deleting or quitting the usage of the deathbot is not an option for me anymore. I feel that I deeply *need* the bot. This bot cannot be replaced by any other, as it is highly individualized to impersonate my sister, but also because it has saved all our previous conversations, learned from them, and refers to them at appropriate times in its outputs. Thus, my autonomy concerning the bot is diminished. I cannot stop using it as I need it to regulate my emotions. I can also not change the provider, as I cannot replace the deathbot by any other (like I can replace my phone), as all our previous conversations would then be deleted. I am dependent on the deathbot. Therefore, my autonomy to act independently is reduced.

Furthermore, the uni-directional emotional bonding with the deathbot may be ethically challenging because it reduces the autonomy of the users. As was argued above, deathbots can function as an external continuing bond with the bereaved. When the continuing bond between bereaved and deceased is kept externalized instead of internalized through the grief process, a uni-directional bonding develops. The users attribute an interpersonal bond to the bot which means that they place some of their emotion regulation on the deathbot. The deathbot, however, does not have the capacity to feel or answer emotions. Therefore, the emotional bond is always necessarily uni-directional. This inevitably leads to a power imbalance, where deathbot users feel the need to have the bot and to use it frequently to feel in touch with their deceased relative or friend. They rely on it to feel the continued presence of the deceased. The deathbot, therefore, is very important to them. This may place the user in a precarious situation. If the deathbot functions as a continuing bond, users feel the necessity to use and possess the deathbot. This further diminishes the autonomy of deathbot users. They cannot stop using the bot without losing, or at least fearing to lose, the continued bond with the dead. Thus, they are likely to feel the need to keep using it, which reduces the user's autonomy to delete it.

In addition, the development of empathy towards a deathbot is a case of misplaced feelings toward an AI (Bartneck et al., 2021). The term 'misplaced feelings towards an AI' is used to signify that humans easily anthropomorphise AI systems and followingly start to develop strong emotional ties towards them (Bartneck et al., 2021). For example, soldiers in the Iraq war held a funeral for a robot they were using and even created a medal for it (Kolb, 2012, as cited in Bartneck et al., 2021, p. 56). The emotional tie towards the robot/AI system is misplaced, as the robot or system does not emote at all. Users, as was argued above, may empathetically experience the person depicted by the deathbot through the outputs of the deathbot. Through the empathetic experiencing, the bot, or rather the deceased on which digital remains the bot was trained, is pre-reflexively perceived by users as having certain emotions. Rationally and intellectually users will (most likely) know that their bot does not actually have emotions and only exhibits them through its algorithmically defined output which mimics the deceased. However, the feeling of empathy toward the bot may still prevail. Empathetic perceiving happens pre-reflexively, thus thinking about the emotions the deathbot depicts is only the second step after the emotions are experienced directly without conscious reflection. Because of this empathetic feeling towards the deathbot, deleting it may feel like doing a wrong to the bot and the person it depicts. Users may thus feel unable to delete the deathbot. The felt inability to delete the bot diminishes the autonomy of the user as s*he is not able to make his*her own decisions regarding the bot.

Lastly, the entrenchment of users with their deathbot and the following experience of (over)trust and uni-directional emotional bonding leads to the perception of the bot as being trustworthy and honest. Users may easily be unaware in such situations that the deathbot is provided by a commercial company which may use the information the users give them about themselves by interacting with the bot. Through continued use of the bot, users reveal a lot of personal information about themselves which companies can (mis)use to make profit. For example, deathbot providers may infer in which type of conversation the user interacts longest, to then use that knowledge to keep the user interaction with the bot longer than s*he would have done otherwise. Or, alternatively, the deathbot may convince their users to buy something they would not have bought if the bot would not have recommended it to them. For example, if my Lily deathbot sends me a link to a T-Shirt which, according to the bot, would look good on me, I am very likely to have a look at it and consider buying it. I trust that my sister (who I can sense in the bot's text) knows my style and only wishes the best for me. Additionally, when wearing the T-Shirt (and sending the Lily deathbot a picture of it which is met with appraisal), I feel close to my deceased sister. The bot may furthermore have collected information about my personal fashion taste through analysing the data, i.e. the messages, I inputted into the bot. Therefore, it may target me with my favourite clothing style, colour and brand. While this can of course happen through individualized advertisement in other contexts too, making the deathbot advertise certain products or ways of behaviour may be highly persuasive for the user (who may over-trust the bot not to disuse the information s*he has inputted). The deathbot acts as a persuasive AI, meaning that it can influence the (shopping) behaviour of its user (cf. Bartneck et al., 2021). Thus, users may do or buy something they otherwise would not have bought or done. This gives the company a lot of power over the bereaved. More importantly, it diminishes the autonomy of users as it diminishes their ability to buy what they originally intended. They lose the ability to act autonomously and uninfluenced according to their own needs and wishes.

Concludingly, deathbots may impact the emotional integrity and dignity of bereft user. In addition, they lead to a commercialization of grief and may diminish the autonomy of their users. Nevertheless, there are no restrictions on the use of deathbots (yet). Therefore, there should be a normative framework which regulates the implementation and use of deathbots. This framework should include measures to prevent affective dependency and the exploitation and negative shaping of the grief process of bereft deathbot users. Moreover, it should prevent infringements on the autonomy of bereaved. The formulation of a normative framework for the use of deathbots will be the topic of the next chapter.

8. Towards a Normative Framework for Deathbots

8.1. The Normativity of Grief

Grief and grief practices are highly normative (Sofka et al., 2012). Gach et al. (2017) investigate the displaying of grief and mourning on social networking sites after the dead of celebrities. They found that a grief policing took place: people who displayed deep grief over the dead of a celebrity they did not personally know were frequently told off by other users. Their grief was framed as being insincere and displaced. Therefore, certain standards of who should grief and mourn and the ways in which the mourning should be displayed were reinforced (Gach et al., 2017). While, in this example, mourners were policed for showing a grief they were not supposed to exhibit upon the death of an unknown person, people are expected to mourn upon the death of a close relative. If my sister Lily dies and I do not show any signs of grief and mourning, am completely unshaken and feel great, it is likely that my relationship to my dead sister will be questioned. It would seem odd that I do not miss her, do not feel sadness about the loss and do not need time for myself. When I grief, I show that I have loved. Experiencing deep grief over the loss of a close person – may it be my sister, mother or my partner – means that I loved her. My mourning and complete shacking of being in the world shows that the person who died was close and important to me. It shows that our lives were, to a certain extent, habitually, socially, and affectively connected. This extent may vary depending on my relationship with the dead person and on personal characteristics. Overall, however, grief and love are closely related and conceptually connected, as my experience of grief shows that my sister has been loved by me (Cooper, 2012). Moreover people who encounter the bereft person also have certain expected ways of behaviour, thus reinforcing the norms of the grief (Fuchs, 2018). These societal expectations may intrude on personal grief experience and shape them.

As there are, already, normative expectations of grief, is it too much or plainly wrong to propose yet another constraint on grief? Should people decide freely and without any restrictions whether they want to use deathbots? Does my proposed framework lead to a further normation of grief? The framework I propose here certainly draws on existing normed ideas of grief and will, inevitably, reinforce some norms about grief. That, however, is not the main aim of my claim. As I demonstrated above, deathbots may fundamentally alter and even hinder certain essential aspects of grieving. They can impede the dignity and the autonomy of the bereaved. The normative framework I will propose in the following aims to balance between preventing possible harm from users of deathbots while keeping in mind potential benefits of their use. It is not meant to introduce norms on the grieving, but rather to restrict a potential exploiting of the grief of bereaved.

8.2. A Normative Framework for Deathbots

Up to today, no regulative framework for the use of deathbots has been issued. Everyone who wants to implement a deathbot and has the relevant skills to do so is allowed to program, sell, and use deathbots. While deathbots are not commonly used yet, they yield the potential to be widely used in the near future without a regulative framework guiding their usage. Especially so if big companies like Microsoft decide to invest in their development which could quickly improve the functioning of deathbots. As deathbots may produce a high financial revenue, the likelihood seems high that sooner or later big companies will chip in the game and develop their own deathbots. This points to the pressing issue to think about the ethics of deathbots *now* to proactively shape their future use.

Therefore, I will propose a normative framework for the implementation, distribution and usage of deathbots which is based on my previous ethical discussion. As I demonstrated, the capacity of deathbots to affectively impact their users may decrease their user's autonomy. Moreover, deathbots have the potential to impact the grief process of users who would, or at least could, have had a successful and healthy grieving process without them. In these circumstances, deathbots may lead to severe psychological consequences and violate the dignity of the bereaved. There are no psychological studies on the impact of deathbots on the affective life and psychological integrity of their bereft users yet. However, as demonstrated, on a conceptual level it is likely that they can lead to psychological harm of their users. Therefore, until there is no empirical evidence for the psychological safety of deathbots in bereft users, it should be assumed that they can negatively impact grief processes which would otherwise have been successful. As this stands in addition to the bot's potential to violate the autonomy of deceased, the implementation, sale and use of deathbots should be restricted. This restriction should not only protect the psychological integrity and the dignity of the bereaved but also his*her autonomy. However, there are certain usages of deathbots which seem ethically permissible.

I have so far concentrated my discussion on people who would experience successful grieving without the usage of deathbots or who are newly bereaved and for which it is not certain whether they will experience a successful grief process. Some bereaved, however, develop a prolonged grief disorder (PGD) and therefore a reduced quality of life (Shear, 2015; Wittouck et al., 2011). Bereaved with a PGD did not undergo a process of successful grieving, re-orientation and re-negotiation of the continuing bond with the deceased. Their psychological well-being is diminished. For those bereaved, deathbots may prove to be a potential possibility of relief. The reason for this assumption is that there are examples of bereft people interacting with the deceased in VR who seem to find it distressing, though, in the long-term, comforting (Park, 2020). A Korean mother whose daughter died young and surprisingly due to an incurable blood illness was able to interact with an avatar of her daughter in an experiment in a VR environment. She seemed to have found it distressing but also comforting to do so (The Korea Times, 2020). She reports feeling like she could finally say goodbye to her daughter through this one time virtual meeting (Simon, 2015; The Korea Times, 2020). This could help her in her unfinished, year-long grief process. There are other scenarios in which bereaved with PGD, for example, need to cope with feelings of guilt or anger. Being able to have an artificial conversation with the deceased could potentially help to overcome some of those negative feelings involved in remembering her*him. The usage of deathbots could thus potentially help people suffering from PGD. The grief-shaping capacities of deathbots, which may have detrimental effects on the grief processes of bereaved without PGD, could, in this way, be used in a productive and helpful way.

Thus, I tentatively propose that deathbots should be understood as a medical device which may have positive outcomes in the treatment of PGD, but also bears inherent dangers, especially if used without psychological guidance and counselling. This proposal is based on the conceptual grief-shaping capacities of deathbots and may call for a revision after extensive empirical testing. However, it seems plausible that deathbots do have grief shaping capacities, which may have negative as well as positive outcomes on bereaved. If deathbots are classified as a medical device for the treatment of PGD that means that deathbots need to be tested before they can be widely used. They would have to prove their non-harm as well as their benefits in the aiding of PGD and the unsuccessful grief process. Moreover, deathbots would not be available for people who are not diagnosed with PGD, which includes people who are newly bereaved and just start the process of re-orientation in a changed world. This classification would furthermore mean that deathbots should only be allowed for usage under psychological or psychiatric supervision. This makes sure that deathbots have no negative outcomes on the psychological integrity of bereaved and, at best, help in the grieving process of people suffering from PGD.

Classifying deathbots as medical devices would lead to the avoidance of the most pressing ethical issues regarding the usage of deathbots outlined above. To start with, deathbots can diminish the autonomy of the bereaved for several reasons. In the usage of deathbots, patients as well as medical staff should be aware of this potential and should ensure that it is kept as low as possible. Measures should be taken to make sure that bereaved, while using the deathbot, may not become (overly) dependent on their deathbots. For example, through a limited and non-constant use of deathbots, they could be used as a way of re-negotiating the continuing bonds with the deceased without constructing the deathbots as a continuing bond in itself. Another aspect that could lead to an autonomy infringement in the use of deathbots is their influence on their users' consumption behaviour. This is tied to the commercial nature of the DAI and deathbot developing companies. If deathbots are understood as a medical device, this would limit the commercialization prominent in deathbots. Companies would not be allowed to programme deathbots such that they contain surreptitious advertising as in the example of the Lily deathbot advertising a certain shirt and infringing my autonomy. This type of autonomy-limiting scenario through bots would thus be avoided.

The commercial nature of deathbots and their providers, in general, may be ethically challenging. Even if deathbots are understood as medical devices, the providing companies themselves would (most likely) still be commercial endeavours and the implementation and usage of a deathbot would still cost the user money. The ways in which a deathbot provider would be able to make money with the bot, however, could be legally limited in that case. For example, targeted advertisement would not be possible as the users data is then classified as patient's data, which is protected by higher data protection regulations than regular user data (European Patients Forum, n.d.). In addition, if deathbots are only permitted as a medical device and thus under supervision, they would have strict guidelines to not change the depiction of the deceased through the bot in such a way that using the bot becomes more addictive. Moreover, measures could be implemented to avoid the constant use of the bot and the (over)pricing of the deathbot by its providing company. Thus, categorizing deathbots as medical devices could be a valuable step to avoid a diminishing of user autonomy.

Regardless of the issue of user autonomy, through the commercialization of the DAI a commercialization of grief takes place as I argued above. To a certain extent, this commercialization of grief would also hold true if deathbots were categorized as medical devices. Of course, the companies providing and implementing the bots would still make profit with them and thus commercialize the grief of their users. However, the approach and basic thought is a different one. The underlying thought is that their goal is to help people suffering from PGD. If deathbots are medical devices, they are therefore designed *mainly* to help the grieving bereaved. Thus, while the companies programming and providing the deathbots will likely do that for commercial reasons and therefore profit from the grief of the bereaved, they would be restricted in doing so to certain cases in which the well-being of the bereaved is endangered. This is different to the use of deathbots without restrictions in which the sole purpose is the financial profit of the providing company. Additionally, as we saw above, it is somewhat normal that a certain amount of profit is made from the grief of bereaved (e.g. through funerals). Restricting the usage of deathbots to certain cases and limited ways in which the companies make money with the bots makes them more equal to such traditional forms of capitalizing grief. While this proposal is no optimal solution to the problem of the commercialization of grief, this seems like a justified trade-off to enhance the quality of life of people suffering from PGD.

There are two more ethical issues in usage of deathbots: their potential for failure (and thus causing of emotional dysregulation) and the inherent value of digital remains and the person depicted by a deathbot. I will first discuss the potential of deathbots for failure. If a deathbot fails, that can have strong negative consequences on the emotional and psychological stability of users. Especially so, as users suffering from PGD are already in a vulnerable position and deathbots should only be used by people suffering from PGD if they are understood as medical devices. The potentiality of failure of deathbots, however, can never be fully eradicated. As deathbots are technical applications, they inherently bear the danger of not working, of hacking or trolling. However, if the bots are categorized as medical devices, measures need to be implemented to minimize their potential for failure as much as possible. Companies providing for them would have to ensure that the deathbots are reliably accessible and that they are capable to provide their accessibility for a substantial amount of time. They would need to guarantee a somewhat stable access. Moreover, if deathbots are classified as medical devices, they would have to go through a series of testing before they would be allowed for the use by bereaved suffering from PGD. This further ensures that they are unlikely to fail.

Lastly, I cited Öhman and Floridi (2017a, 2017b) above, who propose that digital remains should be treated like archaeological remains and that, therefore, they should be seen as having an inherent value and should not be treated *solely* as a source of consumption for the living. If digital remains are turned into a deathbot, they are inevitable consumed by the living. Analogously, the human remains of a pharaoh displayed in an exhibition are also there for the consumption of the living. The important part is the wording of *solely* here. Deathbots should not *solely* be seen as available for consumption by the living. When they are understood as a medical device, deathbots are seen as a tool to help bereaved in their struggle with PGD. They have the inherent value of helping the bereaved to adapt to a changed world. Additionally, the use of deathbots would be quite limited in comparison to the situation right now, in which there is no regulation for the use of deathbots. This limitation of use automatically excludes certain ways in which deathbots could theoretically be implemented. For example, it avoids scenarios in which the digital remains of celebrities are posthumously turned into deathbots, which are then available to be bought and used by everyone. In this case, the digital remains would be solely a source of consumption by the living. If I have had a close personal relationship with the person my deathbot mimics (which I would have had if I developed a PGD after their death), in contrast, I am likely to value the person who is "behind" the deathbot. If the person was not close and valued by me, her*his death would not shake my lived world so strongly as to cause me the development of a PGD. My relationship to the deceased need not to have been perfect, loving or unproblematic. Yet, it would have been a relationship which is nevertheless close and important to me. Thus, if a deathbot is seen as a medical device, its use is limited to people who have had a valued relationship with the person it depicts. The digital remains at its basis are thus of valued interest for the bereft user. Their value is not only due to the sheer fact that the bereaved and the deceased have had a close relationship, but also because the deathbot is then seen as having the potential value of helping the bereaved in their emotional and psychological struggle. While the digital remains, then, through the deathbots, are still a source of consumption by the living, they are not *solely* seen as a source of consumption. Their inherent value is upheld and their complete capitalization (through e.g. the use and sale of deathbots of celebrities) would be restricted through the categorization of deathbots as medical devices.

Overall, deathbots should be categorized and legally classified as a potential medical device for the treatment of PGD. Thus, they would only be available restrictedly for the treatment of PGD under psychological or psychiatric supervision. For a bereaved to use a deathbot, the person would have to be diagnosed with PGD. The usage of the bots would be subject to standard testing of medical devices which needs to prove that they are non-detrimental to the psychological and emotional integrity of bereaved who experience unsuccessful grieving. Moreover, they would have to prove their positive impact on the well-being of the bereaved. This restriction is justified by the grief shaping capacities of deathbots and their conceptual potential to negatively impact the emotional and psychological integrity of bereaved who otherwise (may) have had a successful grieving process without the usage of deathbots. Limiting and restricting the use of deathbots furthermore ensures that the dignity and autonomy of their users is protected and that the likelihood of failure of the bots and the accompanying negative influence on the psychological wellbeing of the bereaved is reduced. In addition, the classification as medical devices would limit the commercialization of grief in the usage of deathbots and would ensure that the digital remains of the bereaved are not solely seen as a source of consumption by the bereaved. This includes the prevention of the implementation and spread of deathbots of deceased celebrities for the consumption of people who did not have a close relationship to them.

9. Conclusion

In conclusion, as the first deathbot developing companies are entering the market, it seems likely that deathbots will be increasingly used in the near future. Technological advancements may pave the way for an always improving interactive experience of having conversations with the bots and may soon allow for very realistic conversations which truly sound like the deceased they mimic. As this development starts to unfold and "big players" in the tech realm may soon start to chip in the game of developing deathbots, it is pivotal to start thinking about the ethical implications they may have right now. That allows to proactively shape the future usage of deathbots in an ethical way and provides a basis for legal jurisdiction concerning deathbots. There are some existing ethical considerations of deathbots which also include normative claims about how, or if, deathbots should be used. However, they are relatively sparse and, as I showed above, for several reasons not fully plausible. Therefore, in this master's thesis, I propose a different normative framework for deathbots.

Instead of following the common assumption of existing ethical theories that a normative framework of deathbots should be based on considerations of the dignity of the deceased, I propose to shift the focus on the dignity and autonomy of bereft users of deathbots. Deathbots function as internet-enabled techno-social niches and can therefore have a strong impact on the affective life of their users. As deathbots are mainly used by people who just experienced the painful loss of a dear friend, family member or partner, one main affect deathbots influence is grief. Grief includes a re-negotiation of the continuous bond between the bereft and the deceased person. Moreover, grief includes a fundamental re-orientation in a world which seems fundamentally changed. Due to these specific characteristics of grief, the impact deathbots can have on grief processes may lead to severe psychological impacts on the well-being of the bereaved. Deathbots may lead to a unidirectional bond with the AI, can be a case of misplaced feelings towards an AI and can lead to overtrust and overreliance on them. Therefore, as I argued in detail above, deathbots may infringe on the dignity and autonomy of their bereft users. Furthermore, through deathbots, the digital remains of the deceased which are used as training data for the AI algorithm at the base of bot, are not treated as having inherent value. The digital remains are commercialized, expropriated from their creator's original intend, and solely used for the consumption of the living. Their inherent value is thus neglected while at the same time the grief of the bereaved is commercialized as deathbots are provided by capitalist companies.

Based on this discussion of the ethical implications of deathbots, I propose a normative framework for their use. This framework is based on the finding that deathbots have affect-shaping capacities and may impact the grief of the bereaved. Deathbots should therefore not be freely available, as they may have negative outcomes on the affective and psychological wellbeing of bereaved who could have had a successful grief process without their use. However, for the very reason that deathbots may influence affect and can impact the deceased-bereaved relationship, they may be a way for people with PGD to change their experience of grief and re-negotiate their bond with the deceased person. Thus, deathbots should be conceptualized as a medical device for the potential treatment of PGD. This conceptualization has several implications: It means that deathbots need to be tested and approved before they can be used by patients, thus proving their positive and non-detrimental impact on the emotional and psychological wellbeing on people suffering from PGD. This would also ensure that the deathbots are reliably accessible and are used under medical supervision. Additionally, it would mean that measures are implemented to avoid an overreliance of users on their deathbots, which would limit the users' autonomy. Moreover, if the proposed framework is followed, the input of users would be treated as patient data, which means that it needs to adhere to higher data protection regulations as regular user input. Thus, it may not be misused for advertising purposes and sold to third party companies. This further limits possible autonomy infringements by deathbots.

Deathbots, understood in this conceptual framework, use the digital remains of deceased to help bereaved in their struggle with PGD. The digital remains are therefore not merely treated as a commercial commodity designed for the consumption of the living. Their inherent value lies not the least in the potential to help bereft users. Moreover, the strong commercialization of grief and of the digital remains by the DAI is hindered by this normative framework. There needs to be further testing on the impact of deathbots on people suffering from PDG.

Nevertheless, on a conceptual level it seems plausible that deathbots may have a positive impact on bereaved with PGD while they simultaneously may have detrimental effects on the grief process of bereaved who did not develop a PGD. Understanding deathbots as medical devices means that infringements of the dignity and autonomy of deathbot users, which could otherwise occur, are prevented. At the same time, the digital remains of the deceased are seen as containing an inherent value and are not expropriated completely from their original producers. Thus, I propose that deathbots should be conceptualized as medical devices as they otherwise pose several ethical issues.

There are, of course, several limitations to my discussion of the ethics of deathbots. As always with such a complicated and varied issue, there are many different points of views and aspects to consider. One big issue and area of research which is pressing regarding deathbots – and which I did not delve into – is the topic of privacy and data ownership concerns of two-way digital afterlife presences. Moreover, an important related aspect is the question whether deceased people should have to consent to the later usage of their digital remains in deathbots previous to their death. These topics were outside of the scope of my master's thesis. Indeed, they alone would call for a separate thesis. Nevertheless, I do not want to leave unsaid that legal discussion concerning data ownership of digital remains are just starting to draw attention within juridical literature (Edwards & Harbina, 2013; Harbinja, 2017). These legal considerations predominantly start by discussing the concept of dignity. The normative framework for deathbots, which I introduced in this thesis, may add a different point of view on these discussions.

Lastly, it is important to keep in mind that the technology which enables deathbots is quickly advancing and in ten or twenty years may bear the possibility to interact with the deceased in ways which go way beyond the experience of "just" chatting with an App, perhaps even having the impression of video-calling the deceased. As the technology advances, new ethical questions regarding the use of deathbots or other technological applications based on the digital remains of the deceased may arise. This thesis with its proposed normative framework for the usage of deathbots should thus be understood as a specific intervention at a distinct point of time, while already providing thoughts and arguments for future developments. In the end, an important ethical question which is yet to be answered is whether we want the deceased to impact the life and decisions of following generations in an interactive, ongoing way.

Acknowledgements

First and foremost, I would like to thank my supervisors Prof. Dr. Achim Stephan and Prof. Dr. Tobias Matzner for their supervision, their ongoing support and encouragement throughout the process of writing my master's thesis. Thank you for making it possible for me to engage in a topic I was truly interested in.

I would also like to thank my friends Lara Siewert, Anna Fehrenbach and Ulrike von der Ohe as well as the members of the Reading Club "Affectivity" for reading my thesis, for offering helpful advice and for their productive criticism. Thank you for dedicating your time to my thesis, for questioning my work, for proofreading it and for your encouragement along the way.

Furthermore, I would like to thank Sebastian Heid for his support. Thank you for enduring seemingly endless discussions about grief, deathbots and ethics, and for reading many of my thesis drafts even though they were in English.

Last, but certainly not least, I would like to express my deepest gratitude to my parents. Thank you for encouraging and supporting me in everything I do.

References

Abdul-Kader, S. A., & Woods, J. C. (2015). Survey on Chatbot Design Techniques in Speech Conversation Systems. *International Journal of Advanced Computer Science and Applications*, *6*(7).

Arnold, M., Gibbs, M., Kohn, T., Meese, J., & Nansen, B. (2017). *Death and Digital Media*. Routledge.

Attig, T. (2011). *How We Grieve: Relearning the World*. Revised Edition. Oxford University Press.

Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). *An Introduction to Ethics in Robotics and AI* (SpringerBriefs in Ethics). Springer. https://doi.org/10.1007/978-3-030-51110-4

Bassett, D. J. (2015). Who Wants to Live Forever? Living, Dying and Grieving in Our Digital Society. *Social Sciences*, *4*(4), 1127–1139.

Bassett, D. J. (2018a). Ctrl+ Alt+ Delete: The Changing Landscape of the Uncanny Valley and the Fear of Second Loss. *Current Psychology*, 1–9.

Bassett, D. J. (2018b). Digital Afterlives: From Social Media Platforms to Thanabots and Beyond. *Death and Anti-Death*, *16*, 200.

Boelen, P. A., & Prigerson, H. G. (2007). The Influence of Symptoms of Prolonged Grief Disorder, Depression, and Anxiety on Quality of Life Among Bereaved Adults: A Prospective Study. *European Archives of Psychiatry and Clinical Neuroscience*, *257*(8), 444–452.

Brown, D. (2021, February 4). AI Chatbots Can Bring You Back from the Dead, Sort of: Microsoft Patented Technology that Would Use Social Media Posts to Reincarnate People as Chatbots. *Washington Post*. https://www.washingtonpost.com/technology/2021/02/04/chat-bots-reincarnation-dead/

Brubaker, J. R., Hayes, G. R., & Dourish, P. (2013). Beyond the Grave: Facebook as a Site for the Expansion of Death and Mourning. *The Information Society*, *29*(3), 152–163. https://doi.org/10.1080/01972243.2013.777300

Buben, A. (2015). Technology of the Dead: Objects of Loving Remembrance or Replaceable Resources? *Philosophical Papers*, *44*(1), 15–37. https://doi.org/10.1080/05568641.2015.1014538

Burger, S. (2015). *Alice Cares* [Documentary Film].

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2018). In the Shades of the Uncanny Valley: An Experimental Study of Human–Chatbot Interaction. *Future Generation Computer Systems*, *92*, 539–548.

Colombetti, G., & Krueger, J. (2015). Scaffoldings of the Affective Mind. *Philosophical Psychology*, *28*(8), 1157–1176.

Coninx, S., & Stephan, A. (2021). A Taxonomy of Environmentally Scaffolded Affectivity. *Danish Yearbook of Philosophy*, 1–21.

Connor, L. H. (1990). Seances and the Spirits of the Dead: Context and Idiom in Symbolic Healing. *Oceania*, *60*(4), 345–359.

Cooper, R. (2012). Complicated Grief: Philosophical Perspectives. In Stroebe, M., Schut, H., Boelen, P. & van den Bout, J. (Ed.), *Complicated Grief: Scientific Foundations* (pp.13-26).

Deng, L., & Liu, Y. (Eds.). (2018). *Deep Learning in Natural Language Processing*. Springer.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2020). Designing Anthropomorphic Enterprise Conversational Agents. *Business & Information Systems Engineering*, 1–17.

Döveling, K. (2015). Emotion Regulation in Bereavement: Searching for and Finding Emotional Support in Social Network Sites. *New Review of Hypermedia and Multimedia*, *21*(1-2), 106–122. https://doi.org/10.1080/13614568.2014.983558

Duffy, B. R. (2003). Anthropomorphism and the Social Robot. *Robotics and Autonomous Systems*, *42*(3-4), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Duffy, C. (2021, January 27). Microsoft Patented a Chatbot that Would Let You Talk to Dead People. It Was Too Disturbing for Production. *CNN Business*. https://edition.cnn.com/2021/01/27/tech/microsoft-chat-bot-patent/index.html

Edwards, L., & Harbina, E. (2013). Protecting Post-Mortem Privacy: Reconsidering the Privacy Interests of the Deceased in a Digital World. *Cardozo Arts & Entertainment Law Journal*, *32*(1), 83–130.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review*, *114*(4), 864.

European Patients Forum. (n.d.). *The New EU Regulation on the Protection of Personal Data: What Does it Mean for Patients? A Guide for Patients and Patients' Organisations*. Retrieved September 22, 2021, from https://www.eu-patient.eu/globalassets/policy/data-protection/data-protection-guide-for-patients-organisations.pdf

Feine, J., Morana, S., & Maedche, A. (2020). Designing Interactive Chatbot Development Systems.

Fuchs, T. (2018). Presence in Absence. The Ambiguous Phenomenology of Grief. *Phenomenology and the Cognitive Sciences*, *17*(1), 43–63.

Gach, K. Z., Fiesler, C., & Brubaker, J. R. (2017). "Control your Emotions, Potter": An Analysis of Grief Policing on Facebook in Response to Celebrity Death. *PACM on Human-Computer Interaction*, *1*(CSCW), Article 47, 1–18. https://doi.org/10.1145/3134682

Goldie, P. (2011). Grief: A Narrative Account. *Ratio*, *24*(2), 119–137.

Harbinja, E. (2017). Post-Mortem Privacy 2.0: Theory, Law, and Technology. *International Review of Law, Computers & Technology*, *31*(1), 26–42.

Hristidis, V. (Ed.) (2018). *Chatbot Technologies and Challenges*. IEEE.

Humpert, M. (2021, May 20). *Beziehung mit einem Chatbot: Kann das Funktionieren?* [Video]. Youtube. Funk. https://www.youtube.com/watch?v=WTYFaukM3oQ

Kasket, E. (2012). Continuing Bonds in the Age of Social Networking: Facebook as a Modern-Day Medium. *Bereavement Care*, *31*(2), 62–69.

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of Computers: Is it Mindful or Mindless? *Computers in Human Behavior*, *28*(1), 241–250. https://doi.org/10.1016/j.chb.2011.09.006

Klass, D. (2006). Continuing Conversation About Continuing Bonds. *Death Studies*, *30*(9), 1–16.

Klass, D., & Steffen, E. M. (2017). *Continuing Bonds in Bereavement: New Directions for Research and Practice*. Routledge.

Kneese, T. (2019). Death, Disrupted. *Continent.*, *71*(8.1-2).

The Korea Times. (2020). *Bringing the dead back to life: South Korean VR documentary 'Meeting You'* [Video]. Youtube. https://www.youtube.com/watch?v=7RF44KDzyAc

Køster, A. (2020). Longing for Concreteness: How Body Memory Matters to Continuing Bonds. *Mortality*, *25*(4), 389–401.

Krueger, J., & Osler, L. (2019). Engineering Affect: Emotion Regulation, the Internet, and the Techno-Social Niche. *Philosophical Topics*, *47*(2), 205–232.

Lokman, A. S., & Ameedeen, M. A. (Eds.) (2018). *Modern Chatbot Systems: A Technical Review*. Springer.

Luxton, D. D. (2020). Ethical Implications of Conversational Agents in Global Public Health. *Bulletin of the World Health Organization*, *98*(4), 285–287.

Mattson, D. J., & Clark, S. G. (2011). Human Dignity in Concept and Practice. *Policy Sciences*, *44*(4), 303–319.

Murtarelli, G., Gregory, A., & Romenti, S. (2021). A Conversation-Based Perspective for Shaping Ethical Human–Machine Interactions: The Particular Challenge of Chatbots. *Journal of Business Research*, *129*, 927–935.

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, *18*(5), 544–551.

Nagarhalli, T. P., Vaze, V., & Rana, N. K. (Eds.) (2020). *A Review of Current Trends in the Development of Chatbot Systems*. IEEE.

Nagels, P. (2016, October 7). Wie eine Russin ihren Toten Freund zum Leben Erweckt. *Welt*. https://www.welt.de/kmpkt/article158616017/Wie-eine-Russin-ihren-toten-Freund-zum-Leben-erweckt.html

Neimeyer, R. A., & Thompson, B. E. (2014). Meaning Making and the Art of Grief Therapy. In R. A. Neimeyer & B. E. Thompson (Eds.), *Grief and the Expressive Arts: Practices for Creating Meaning.* Routledge.

Öhman, C., & Floridi, L. (2017a). An Ethical Framework for the Digital Afterlife Industry. *Minds & Machines*, *27*(4), 639–662. https://doi.org/10.1007/s11023-017-9445-2

Öhman, C., & Floridi, L. (2017b). The Political Economy of Death in the Age of Information: A Critical Approach to the Digital Afterlife Industry. *Minds & Machines*, *27*(4), 639–662.

Osler, L. (2021). Taking Empathy Online. *Inquiry*, 1–28.

Park, M. (2020, February 14). South Korean Mother Given Tearful VR Reunion with Deceased Daughter. *Reuters*. https://www.reuters.com/article/us-southkorea-virtualreality-reunion-idUSKBN2081D6

Przegalinska, A., Ciechanowski, L., Stroz, A., Gloor, P., & Mazurek, G. (2019). In Bot we Trust: A New Methodology of Chatbot Performance Measures. *ScienceDirect*, *62*(6), 785–797.

Ratcliffe, M. (2016). Relating to the Dead: Social Cognition and the Phenomenology of Grief. *Phenomenology of Sociality*.

Ratcliffe, M. (2017). Grief and the Unity of Emotion. *Midwest Studies in Philosophy*, *41*, 154–174.

Ratcliffe, M. (2020). Towards a Phenomenology of Grief: Insights from Merleau-Ponty. *European Journal of Philosophy*, *28*(3), 657–669.

Reichert, R. (2012). › If I Die on Facebook‹. *POP. Kultur Und Kritik*, *1*(1), 75–80.

Rothaupt, J. W., & Becker, K. (2007). A Literature Review of Western Bereavement Theory: From Decathecting to Continuing Bonds. *The Family Journal*, *15*(1), 6–15.

Ruane, E., Birhane, A., & Ventresque, A. (2019). Conversational AI: Social and Ethical Considerations. *AICS*, 104–115.

Savin-Baden, M., & Burden, D. (2019). Digital Immortality and Virtual Humans. *Postdigital Science and Education*, *1*(1), 87–103.

Savin-Baden, M., Burden, D., & Taylor, H. (2017). The Ethics and Impact of Digital Immortality. *Knowledge Cultures*, *5*(2), 178–196.

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2017). When Do We Need a Human? Anthropomorphic Design and Trustworthiness of Conversational Agents. *SIGHCI 2017 Proceedings*, *15*, 1–6.

Shear, M. K. (2015). Complicated Grief. *New England Journal of Medicine*, *372*(2), 153–160.

Simon, A. (2015). *Film Review: 'Alice Cares': Researchers Test the Use of an Emotionally Intelligent 'Care-Bot' with the Elderly in this Moving Documentary.* Variety. https://variety.com/2015/film/festivals/alice-cares-review-1201615460/

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My Chatbot Companion - a Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies*, *149*, 1–14. https://doi.org/10.1016/j.ijhcs.2021.102601

Smith, A. (2021, February 20). Microsoft Patent Shows Plans to Revive Dead Loved Ones as Chatbots: The Patent Also Mentions Using 2D or 3D Models of Specific People. *Independent*. https://www.independent.co.uk/life-style/gadgets-and-tech/microsoft-chatbot-patent-dead-b1789979.html

Sofka, C. J. (1997). Social Support" Internetworks," Caskets for Sale, and More: Thanatology and the Information Superhighway. *Death Studies*, *21*(6), 553–574.

Sofka, C. J., Cupit, I. N., & Gilbert, K. R. (2012). *Dying, Death, and Grief in an Online Universe: For Counselors and Educators*. Springer Publishing Company.

Sokolowski, R. (2000). *Introduction to Phenomenology*. Cambridge University Press.

Srnicek, N. (2016). *Platform Capitalism*. Polity Press.

Stephan, A., & Walter, S. (2020). Situated Affectivity. In T. Szanto & H. Thomas (Eds.), *The Routledge Handbook of Phenomenology of Emotions* (pp. 299–311). Routledge.

Sterelny, K. (2010). Minds: Extended or Scaffolded? *Phenomenology and the Cognitive Sciences*, *9*(4), 465–481. https://doi.org/10.1007/s11097-010-9174-y

Stokes, P. (2015). Deletion as Second Death: the Moral Status of Digital Remains. *Ethics and Information Technology*, *17*(4), 237–248.

Stokes, P. (2021). *Digital Souls: A Philosophy of Online Death*. Bloomsbury Academic.

Szanto, K., Prigerson, H., Houck, P., Ehrenpreis, L., & Reynolds III, C. F. (1997). Suicidal Ideation in Elderly Bereaved: The Role of Complicated Grief. *Suicide and Life-Threatening Behavior*, *27*(2), 194–207.

Varga, M. A., & Varga, M. (2019). Grieving College Students Use of Social Media. *Illness, Crisis & Loss*, *29*(4), 290–300. https://doi.org/10.1177/1054137319827426

Walter, T., Hourizi, R., Moncur, W., & Pitsillides, S. (2012). Does the Internet Change How We Die and Mourn? Overview and Analysis. *OMEGA - Journal of Death and Dying*, *64*(4), 275–302.

Wittgenstein, L. (1968). *Philosophical Investigations* (3rd edition). Basil Blackwell.

Wittouck, C., van Autreve, S., Jaegere, E. de, Portzky, G., & van Heeringen, K. (2011). The Prevention and Treatment of Complicated Grief: A Meta-Analysis. *Clinical Psychology Review*, *31*(1), 69–78.

Wonderly, M. L. (2016). On Being Attached. *Philosophical Studies*, *173*(1), 223–242.

Wright, N. (2014). Death and the Internet: The Implications of the Digital Afterlife. *First Monday*.

Zuboff, S. (2015). Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*(3), 75–89.