



Persons and their Digital Replicas

Jurgis Karpus¹ · Anna Strasser^{1,2}

Received: 30 October 2024 / Accepted: 30 January 2025 / Published online: 17 February 2025
© The Author(s) 2025

Abstract

Creating a legacy is often seen as a way to circumvent mortality. At the very least, it is a way to ensure that someone's ideas live on and their influence on others extends beyond their own lifetime. Common examples of such legacies are cultural products, such as books, music, or art, that one creates and leaves behind. In light of rapid advances in artificial intelligence research, it is conceivable that it will soon become possible – and, to some extent, it already is – to create a new type of legacy in the form of a virtual entity, such as a chatbot or an avatar, that acts as a digital replica of someone. In this paper, we investigate in what sense, if any, one could say that the original person continues to exist in their digital replica. Taking Derek Parfit's work on the nature of personal identity as our inspiration and starting point, we suggest under what circumstances one could (and should) perceive such replicas as genuine extensions of one's personhood.

Keywords Personal identity · Digital replica · Psychological connectedness · Extended personhood · Parfit

1 Introduction

Since the beginning of humanity, the subject of one's finiteness has been a topic of great concern. Therefore, it's not surprising that we can find ideas across all cultures about how to circumvent one's mortality. Many religions have generated ideas about eternal life, for example, through reincarnation or an afterlife in heaven or hell. In addition to such spiritual ideas, people have also found other ways to ensure that

✉ Jurgis Karpus
jurgis.karpus@lmu.de

Anna Strasser
berlinerdenkwerkstatt@gmail.com

¹ Faculty of Philosophy, Philosophy of Science and Religious Studies, LMU Munich, Munich, Germany

² DenkWerkstatt Berlin
<https://www.denkwerkstatt.berlin/>

some form of legacy persists after their death. This may consist in planting a tree, building a house, having children, producing cultural products, for example, books, music, art, or committing actions that are worth mentioning in historical reports.

It is indisputable that people are able to create powerful legacies. For example, influential books can substantially impact subsequent generations. In a certain sense, the authors of these works continue to exist – their ideas live on. Recent developments in artificial intelligence research suggest that there may emerge new ways to create long-lasting legacies based on all kinds of data that contain information about a particular person's life. This could take the form of an interactive virtual entity, such as a chatbot or an avatar, that acts as a digital replica of someone and represents that person to others, possibly long after the original person's biological death. But in what sense, if any, could we say that the original person continues to exist in a digital replica of this kind? Moreover, if enough data about individual people is collected and stored, it may become possible to create digital replicas of already deceased persons. How should we, in such cases, describe the relation that holds between the origin (that is, the replicated person) and the newly created digital entity?

Taking inspiration from recent works by artists, roboticists, novelists, and filmmakers – some of them pure science fiction and others slightly less so – we consider what kinds of replicas could be constructed in our time and in the near future (Sect. 2). We suggest that a replica can only genuinely prolong a person's existence if some relation that constitutes what we will call *personhood* holds between the original person and their digital replica. To this end, we draw on Derek Parfit's influential work on the nature of personal identity to see if it can shed light on how such a relation could be described. Put differently, we investigate whether there is a special kind of relation that holds between us and the digital trace that we leave behind (or purposefully create) about our personality, our likes and dislikes, our worldviews, and our character traits, out of which a digital replica of us could be constructed.

The question concerning the possibility of digitalizing people is not new, and there are various philosophical debates that deal with it. However, given the rapid technological developments of recent years, it seems to us an exciting enterprise to connect the state of research on the potential realization of digital replicas to the philosophical debate on the persistence of personhood. As our starting point, we use Derek Parfit's account of personal identity. We will refer to some other approaches, but it is beyond the scope of this paper to give a comprehensive overview of all work in philosophy that has been recently developed on this topic. In this light, it is best to see our project as a contribution to an already rich and growing body of literature in this domain.

According to Parfit, our continued existence – and, by extension, the relation of personal identity – is best described in terms of psychological connections that hold between us here and now and our future and past selves. Taking some of the core ideas from Parfit's work, we investigate under what circumstances we could think of a digital replica as an extension of someone's personhood even when such a relation would not constitute a relation of personal identity (Sect. 3). Subsequently, we elaborate on the relation that might hold between us and digital replicas and ask whether this relation may serve as a sufficient reason for us to claim privileged rights not only with respect to the creation of digital replicas but also concerning personal data about our psychology, out of which a digital replica of us could be created (Sect. 4).

2 Visions and Reality

Artists, roboticists, novelists, philosophers, and filmmakers have been exploring the idea of replicas for quite some time (for an excellent and rich review, see Stokes, 2021b). In addition to fictional genres that tell us of successful mind uploads, such as living in “The Matrix” or the possibility of teleportation (see, e.g., Kurzweil, 2000, or Bostrom, 2003), some visions are closer to reality than one might initially think. The most referenced vision was imagined in the famous episode ‘Be right back’ of the Black Mirror TV series, in which Martha’s boyfriend, Ash, tragically dies in an accident. Grieving, Martha subscribes to a service that takes all of Ash’s social media posts and online communications before he died to create an interactive Ash chatbot for Martha to converse with. Later in the film, Martha upgrades to have an embodied version of Ash – a robot that looks and moves like Ash as well. Even though she knows that this is not her boyfriend, Martha struggles with the striking resemblance.¹

This science fiction story inspired a real product – the Replika chatbot app.² After Eugenia Kuyda’s best friend unexpectedly died in a traffic accident, she decided to feed all her saved online conversations with her friend into a system powered by artificial intelligence (AI) to create a chatbot version of her friend. The result was so impressive that Kuyda went on to found a company that now offers an AI-powered chatbot-friend to anyone. Unlike in Kuyda’s tragic story, the Replika chatbot app continually learns from its conversations with a user to become ever more like the user (Murphy, 2019). The ultimate goal of the app in the initial versions of the product was to replicate the user in order to become the user’s soulmate.

While the original science fiction story depicted in the Black Mirror episode is far too futuristic in the context of present-day technology, many have started to grapple with the idea of digital replicas of persons. For a philosophical exploration of the idea of using AI-powered replicas of deceased persons in grieving, see Stokes (2021a, b), Öhman and Floridi (2017, 2018), Krueger and Osler (2022), Lindemann (2022), Hollanek and Nowaczyk-Basińska (2024). YOV³ is one among many companies presently developing such services. For numerous others, see excellent recent reviews by Öhman and Floridi (2017, 2018), Danaher and Nyholm (2024a), and Iglesias et al. (2024).

How informative the data that we leave behind on social media, search engines, online stores, and other platforms is about our personalities, our likes and dislikes, our desires, and our character traits was explored by *Laokoon*, a collective of Berlin-based artists who were interested in how far one can get with constructing a doppelgänger of a person using data available online about that person’s life. In their

¹ Black Mirror Season 2, Episode 1 (2013). A related idea was explored in a recent novel by Kazuo Ishiguro (2021), in which a robot is tasked to learn to imitate a person, a young girl who appears to be gravely ill, in order to subsequently be able to represent that girl to her close relatives in case they would outlast her when she dies.

² <https://replika.com>

³ <https://www.mygov.com/index.html>

investigative project ‘Made to measure’, the group conducted an experiment in which they obtained anonymized data of a person’s Google search history from five years of that person’s life.⁴ Based on this data, they retraced the life of that person and re-enacted that life in a film. Using the so-gained knowledge about the person’s life, the group was able to establish contact with this person via a personalized message posted on Instagram and subsequently invited her to meet and discuss the re-enactment of her life with the actress who played the data donor’s doppelgänger in the film. The reconstruction of the person’s life, according to that person herself, was often strikingly accurate, all the way down to the fine details about how she felt at various points in her real life. Finally, the data donor insisted on ending the confrontation when, as a result of listening to her own life story re-enacted by someone else, her own memories began getting mixed up to the point where she could no longer tell the truth and fiction apart.

Art and fiction aside, similar ideas have found their way into scientific research projects. Hiroshi Ishiguro, a famous roboticist, anecdotally explores the idea of one day being able to send a robot replica of himself to give a talk at a conference while he gives a talk elsewhere, this way allowing him and his replica to take part in more events than he currently can do alone (Nishio et al., 2007; for the same idea as one of the purposes or benefits of creating digital replicas of persons, see also Sweeney, 2023, and Danaher & Nyholm, 2024a). Another research group, consisting of Anna Strasser, Matthew Crosby, Eric Schwitzgebel, David Schwitzgebel, and Daniel Dennett, explored the limits and risks of creating a digital replica of a philosopher. They investigated how far one can get with creating a chatbot that interacts with a user based on the corpus of the philosopher Daniel Dennett (Schwitzgebel, 2021; Strasser et al., 2023).⁵ To create a bot that would generate interpretable textual outputs, the researchers trained (fine-tuned) a GPT-3 engine. As it is already quite well known, GPT-3 is a powerful large language model (LLM) developed by OpenAI that is able to generate strikingly interesting outputs. It is a complex artificial neural network that is trained on a vast amount of data to predict what the next likely word in a sequence will be. According to an early review of this technology, GPT-3 is “shockingly good – and completely mindless” (Heaven, 2020). The question of whether an AI-powered chatbot can succeed and be of any use in representing a living philosopher is new, provocative, and, to some, gravely unsettling (see, for example, Dennett, 2023, and Strasser, 2024). So far, the inherent problems of deep neural networks concerning their reliability and interpretability make it questionable whether a machine that is tasked to predict the next likely word in a sequence may

⁴ <https://www.madetomeasure.online/en>

⁵ Strasser and colleagues created this digital replica, DigiDan, with explicit consent from Dennett. Their agreement stated that no third parties would get access to DigiDan and that DigiDan would not survive Dennett himself (in line with the criteria for determining the permissibility of creating digital replicas of persons that were recently proposed by Danaher and Nyholm, 2024a). For more on DigiDan and Dennett’s own take on his digital replica, see the recording of a talk at the Weizenbaum conference in Berlin featuring DigiDan and an interview with Dennett by Strasser: https://youtu.be/xkOCQH5WMbI?si=yHm5PfM76TVu_oGP (accessed on September 9, 2024). Sadly, Dennett died shortly after this event. DigiDan was deleted as agreed.

generate output that could, with some reservations, be considered as “speaking on behalf of an author.” Despite that, Schwitzgebel and colleagues showed that it was difficult even for experts to distinguish machine-generated answers from those given by Dennett himself to a set of questions on various topics in philosophy; naïve participants in the experiment (that is, those who were not philosophers) were no better at that than chance (Schwitzgebel et al., 2023).

With these examples in mind, it seems plausible to envisage that, with further developments of AI methods and ever more data about our lives being recorded, a reasonably accurate digital representation of our character could, in principle, be constructed.⁶ As noted and richly explored by Patrick Stokes, unlike cultural legacies, such as paintings, books, or pieces of music, which can be influential but do not act in any way themselves, the new type of legacy could take the form of interactive chatbots and even robots – embodied versions of our digital replicas (Stokes, 2021a, b). Öhman and Floridi (2017, 2018) refer to such services as belonging to the digital afterlife industry (DAI) and explore ethical concerns with their creation. Danaher and Nyholm (2024a, b) propose criteria concerning consent, value, transparency, risks, and context for determining the permissibility of such services, and discuss how the creation of digital replicas of persons may affect the value that we attach to persons who are replicated. Sweeney, (2023, 2024) discusses the ethics of people’s use of avatars that sometimes act on their behalf and the dangers of creating replicas of persons without the replicated persons’ knowledge or consent, and Iglesias et al. (2024) explore whether digital replicas of persons are in fact able to fulfil the goals for which they are created. We will come back to some of these works later and this is certainly not an exhaustive list of the rapidly expanding literature on the topic. Our primary question in this paper is how one should construe the relation that holds between the original person and their replica. Should such a replica be considered a mere counterfeit of the original person, as recently suggested by Dennett (2023), or is there something more to it? For example, could it be an extension of that replicated person? If it can, what kind of extension would it be and what would legal and ethical ramifications of that conclusion be? In the next two sections, we address these questions in turn.

3 Parfit, Personhood, and Replicas

Our starting point for thinking about these questions is Derek Parfit’s influential work on the nature of personal identity (Parfit, 1984). We focus on Parfit’s theory because it, or the key elements of it, continues to be endorsed by many thinkers on the topic to this day (Olson, 2024) and it presents the possibility that personhoods need not be restricted to biological organisms alone. It is also the key focus of another recent article on digital replicas of persons (Iglesias et al., 2024) and is

⁶ For reviews of attempts to do that, see Stokes (2021b, pp. 125–131) and Nakagawa and Orita (2022). For a review of recent developments in robotics that could aid these developments, see Mazumder et al. (2023).

often mentioned and discussed in related works (Cerullo, 2015; Danaher & Nyholm, 2024a, b; Campbell et al., 2025). However, our goal here is not to scrutinize and explicate Parfit's theory in its entirety, nor is it to arrive at the best guess at what Parfit would have said about extending his work to the context of digital replicas, but to take what we see to be the most relevant elements of Parfit's theory to the present topic and to develop a new proposal for what relations might and might not hold between persons and their digital replicas.

3.1 Personhood without Personal Identity

One element of Parfit's theory that we do uptake is his claim that what ultimately matters for our continued existence is not personal identity understood as an all-or-nothing relation, but psychological connectedness and/or psychological continuity, which he calls the relation R. To see why, Parfit invites us to consider a futuristic thought experiment. Suppose that one day we will be able to travel by teleportation. Upon setting out on a journey, we will clamber into a teleportation device, enter our desired destination, and press the button "go". The device will then scan our body, destroy it, and send the scanned data to another location, where a similar device will recreate our body cell by cell in nearly an instant. There will be nothing to fear, though, for every fact that constitutes a part of what we are before the journey – our appearance, memories, desires, beliefs, intentions, tastes, fears, worldview, and so on – will be recreated in our new body. Parfit wants to convince us that because all facts about our psychology will be preserved and continue to exist in our post-teleportation replicas, our replicas will essentially be us. As Parfit puts it, "being destroyed and replicated [will be] about as good as ordinary survival" (1984, p. 201).⁷

Thus far, our continued existence could still be described in terms of the relation of personal identity, for if our replica is indeed a perfect copy of ourselves, then any fact that constitutes this relation in us before the journey would also be preserved in our replica after we travelled. But suppose that our future teleportation devices are pretty good but not flawless. Some cells of our bodies may be missed during the scanning process, information about them lost while it travels, or a few neurons accidentally rearranged when our replicas are printed. These glitches might be minor and rare, so that most people would still consider teleportation an enticing mode of transport. And yet something about us could change in the process. If this amounted to a change in one's favourite flavour of ice cream, we might not think much of it. But what if, due to some unfortunate rearrangement of one's brain cells, an astrophysicist becomes a flat Earther, a pianist loses her musical skill, or a teleported "Angela Merkel" becomes "Donald Trump"? It is possible to imagine cases where changes to one's appearance and personality are so severe that there would be no

⁷ The reason why Parfit says "about as good" instead of "as good" is that we may sentimentally value our old bodies. But this is just as important (or irrelevant) as feeling sentimental about an old pair of shoes or a spent T-shirt that one can replace with a new one. In his later writing, Parfit actually dropped the qualifier "about" (2012, p. 7).

doubt about the fact that someone we knew before teleportation was no more and someone else – a different person – emerged in their place post-travel.

That raises a tricky question: at which point along the spectrum, from minute to drastic changes to one's personhood, would the relation of personal identity cease to hold between someone who enters a teleportation device and someone who emerges from another? Parfit argues a) that this is impossible to pinpoint and b) that this cannot depend on some trivial fact, for example, the presence or absence of some particular single brain cell. And yet, in cases that fall into the middle of this spectrum, some person emerges from a teleportation device post-travel. The question then is whether this post-travel person is the same as the pre-travel person, whether they are someone else, or whether they are someone in between. That opens the space to think about a graded notion of what ultimately matters for our continued existence through time. Importantly, this can provide us with a way to understand what happens in the undecidable cases – cases in which the question of whether one person is *identical* to another has no categorical answer and is, therefore, as Parfit puts it, an empty question.⁸ We agree with and adopt Parfit's view. However, we also extend it by suggesting that a graded notion of what ultimately *matters* for survival opens the possibility of thinking about a graded notion of what ultimately *constitutes* our continued existence through time.

3.2 Psychological Connectedness and Continuity

Parfit's theory relies on what he calls the reductionist view. This is the view that there is no matter of fact that is separate from our brains and bodies that constitutes our continued existence through time. By that, Parfit, in particular, means that there is no soul, no Cartesian ego, or simply no further fact beyond what is customarily contained within our brains and bodies that constitutes our personhood. This does not mean that only *biological* brains and bodies can constitute personhoods.⁹ Rather, this means that if we know all facts about a person's brain and body in the usual, natural circumstances (the person is not a cyborg, for example), then we know all facts that also constitute that personhood. In and of itself, this view does not preclude the possibility that personhoods could be instantiated synthetically or perhaps even digitally.

Parfit's reductionist view may nevertheless be seen as too restrictive from the point of view of other approaches to theorizing about personhood. For example, according to some theories, our minds are distributed across many features of the world around us,

⁸ In Parfit's theory, the relation of personal identity can be defined in terms of what ultimately matters for our continued existence (the relation R). However, the question of whether some person X at one point in time and some person Y at another point in time are identical is not always answerable. As Parfit puts it, "[i]t is not true that our identity is always determinate" (1984, p. 216). Sometimes questions regarding the relation of personal identity are empty. Parfit presents two such futuristic cases of branching personhood (1984, pp. 199–201, 254–255). According to him, answering an empty question regarding personal identity is more like a decision than a discovery. Or, as Joshua Rust, referring to van Fraassen, clarifies, "empty questions can only be corrected or otherwise responded to, not answered" (Rust, 2019, p.15).

⁹ Although Parfit (2012) discussed a close connection between personhoods and brains, we do not think that one needs to be strictly committed to the view that personhood presupposes a biological brain.

not all of which are contained within our brains and bodies. With that, our personhoods may be distributed across our bodies and our environments too (Heersmink, 2017). It may be the case that this view is still compatible with Parfit's, but we will not explore that possibility further in this paper.¹⁰ In fact, in the proposal that we are developing here inspired by Parfit's view, our personhoods, or, as we will explain later, at least slivers of our personhoods, may indeed extend beyond our brains and bodies. The view of an extended mind is also the prism through which Patrick Stokes explores the possibility of digital replicas of persons (Stokes, 2021a, b). Conclusions that we will reach in our exploration of the topic will be, to a large extent, compatible with Stokes's. We will point out some of those similarities (and differences) later on. For now, we return to Parfit.

According to Parfit, when we think and talk about our continued existence, ordinarily, two criteria seem to be important: the physical criterion – the continued existence of our physical bodies – and the psychological criterion – the continued existence of our psychology (our memories, desires, beliefs, intentions, and so on). In the usual, natural circumstances, the two criteria go hand-in-hand. The continued existence of one's physical body is necessary for the continued existence of one's psychology. But in unusual and unnatural circumstances, as in Parfit's thought experiment with teleportation, the two can come apart.

Parfit argues that what ultimately matters for our continued existence is captured by the psychological criterion alone. In particular, it is the psychological connectedness that we hold towards our past and future selves. If there are many psychological connections, the psychological connectedness is strong. If there are only a few psychological connections, the psychological connectedness is weak. Our continued existence, in this sense, is a graded matter. Sometimes more of whatever makes us *us* persists from one time to another and other times less. Importantly, as time goes by, our personhood – the “stuff” that makes us *us* – can gradually change to the extent that we become truly different persons from those we were before.¹¹ Utterances such as “I am now a different person from the one I was ten years ago” should not necessarily be thought of as mere metaphors – they can be stating facts.¹²

¹⁰ In his later writing, Parfit (2012) suggested that personhoods “reside” or are constituted only in specific parts of our bodies, namely brains. For example, personhoods do not “reside” in fingernails even though fingernails are parts of bodies. However, personhoods can be intricately related to our wider bodies and things beyond. Indeed, as Parfit pointed out, we often use pronouns “I” and “me” to refer to things beyond our bodies, for example, one might say that they were splashed with mud by a passing car when it was only their trousers that were splashed (2012, p. 20).

¹¹ With *personhood* we refer to what is constituted by what Parfit calls the relation R: the “psychological connectedness and/or psychological continuity, with the right kind of cause” (1984, p. 262) where “the right kind of cause” is, in fact, “any cause” (p. 287).

¹² On the one hand, if the psychological connectedness between a person now and her predecessor ten years ago is sufficiently weak, they can be treated as distinct persons. On the other hand, they may be linked by what Parfit calls an overlapping chain of strong psychological connectedness. For example, at every point throughout the ten-year period, a person may be strongly psychologically connected to her one-month younger and older self. This overlapping chain of strong psychological connectedness – psychological continuity – between the person now and her predecessor ten years ago suggests that they can be treated as the same person. Parfit seems to be open to both conclusions, but he might also regard the question of whether these are two instances of the same person to be empty. See, in particular, his discussion of the example involving a nobleman who undergoes a significant change of character throughout his lifetime (1984, pp. 327–329) and his discussion of the different ways in which questions regarding personal identity may be empty (p. 260).

What precisely psychological connectedness consists and does not consist of is not clear cut, neither in Parfit's original work, nor in our further exploration of his theory here. Take it to contain a collection of psychological features that one usually associates with and considers to be a defining feature of someone's personal (though not physical) characteristics. These will include beliefs about the world, hopes and desires, behavioural and emotional dispositions, memories and outlooks, and similar. As we will discuss shortly, a crucial question will be whether psychological connectedness necessarily entails one's ability to consciously experience psychological and mental states. Instead of answering that question, we will consider how possible answers to it change what digital replicas of persons are or can be.

3.3 Slivers of Personhood, Consciousness, and Replicas

Taking these ideas from Parfit's theory on board, brings us to our original question: what connection would we hold to a digital replica of us if it were possible to create one based on data about our psychology? Would such a replica be entirely independent of our personhood and, hence, have nothing to do with our own existence, or would it be an extension of our personhood and thus a part of what we are? Taking the position that the physical criterion is not what matters and that personhood is graded, these questions are no longer silly.

It is important to note here that, in light of the more recent debate about the role of embodiment for a variety of human cognitive abilities, we may need to revisit the importance of the physical criterion in the constitution of one's personhood (for opposing views in this debate, see, for example, Goldinger et al., 2016, and Farina, 2021). We will not go into that debate here. Instead, we will side with Parfit's idea from his later writing, where he suggested that persons are best thought of as (usually) embodied parts of human beings. The point is that while personhood is usually constituted in one's brain, it may nevertheless be closely and intricately related to one's wider body (Parfit, 2012, p. 20). This relation may be so intricate that severe changes to that body may affect one's personhood to a great extent.¹³ We leave the question to what extent that does or can happen to another occasion. For now, we will just add that the importance of embodiment will not render the rest of our analysis irrelevant since a digital replica of someone may also be embodied (it could, for example, be a robot).¹⁴

According to Parfit, our continued existence through time is best understood and described in terms of degrees of psychological connectedness. As we said earlier, he rejects the idea that we always stand in an all-or-nothing identity relation with

¹³ We thank Hong Yu Wong for this suggestion.

¹⁴ Iglesias and colleagues discuss a renewed emphasis on the importance of the physical criterion based on an argument that most people would likely *prefer* continued existence in their "old" physical bodies over continued existence as body-less or robotic digital entities (Iglesias et al., 2024). We believe that Parfit would respond to this by suggesting that intuitions that underlie this argument, similarly to what he argued concerning the intuition that we stand in an all-or-nothing identity relation with our potential future and past selves, are wrong.

our potential future and past selves (for an alternative view that retains the possibility of an all-or-nothing relation of personal identity and what that means for digital replicas of persons, see Campbell et al., 2025). Focusing on the question of what ultimately matters for the continuation of our personhood and taking the view that personhood is graded, we can describe a “personhood spectrum.” At one end of the spectrum, only slivers of one’s personhood are extended through time (only few psychological connections hold). At the other end of the spectrum, one’s personhood extends in near totality (many psychological connections hold). In the former case, we treat the two (weakly psychologically connected) “selves” as separate persons. In the latter, we treat them as instances of the same person. With that summary, we turn to our question.

First, consider a digital replica that retains everything that there is to be retained about our psychology. This is deliberately vague. What we mean is that the replica “carries” all facts about our psychology that, taken together, ultimately constitute our personhood. This may even include the ability to consciously experience mental states, but we will return to this in more detail shortly. If, from the point of its creation, it truly carried all facts about our psychology in the sense we just described – meaning that a strong psychological connectedness holds between us and the replica – such a fully-fledged replica would truly be an extension of our personhood in near totality.

Of course, the prospect of us being able to create such replicas is as far-fetched as the prospect of inventing teleportation.¹⁵ So let us consider less sophisticated replicas instead. For example, let’s suppose that, unlike humans, such digital replicas do not consciously experience any psychological or mental states. Nevertheless, they are able to use vast amounts of information about our psychology to independently perform certain tasks based on that information. Importantly, our psychological states (our likes and dislikes, beliefs, worldviews, and so on) causally determine how such entities “represent” the environment with which they interact and what (speech or other) acts they perform, with their ultimate goal being to perform acts that we would perform ourselves. In a sense, our beliefs and desires, along with other psychological mental states, shape the “beliefs” and “desires” of these replicas, even though, unlike us, they are incapable of consciously experiencing those mental states. For instance, an advanced chatbot powered by artificial intelligence (AI) that is tasked to represent some real human in a conversation is a good candidate for being what may be called a “zombie” replica of this kind.¹⁶

The answer to the question of what connection holds between a person and her zombie replica is more nuanced. If the ability to consciously experience psychological and mental states is a necessary requirement for being psychologically connected to anyone or anything, then a zombie replica would not constitute an extension of

¹⁵ However, see Kurzweil (2000) or Cerullo (2015) for a more optimistic outlook on this prospect.

¹⁶ Personalized AI-powered recommendation engines, which we increasingly use in our day-to-day lives, are a good example of how vast amounts of information about our psychology can causally determine what digital entities do, with their ultimate goal being to do that which we would do ourselves. While these systems do not watch movies, listen to music, read books, or click on websites, they recommend options that they predict we would choose ourselves (Schrage, 2020).

our personhood. It might be able to mimic some things about us that make us *us*, but it would not constitute any form of our (partial or total) continued existence. This is in line with a point recently made also by Iglesias and colleagues, who argued that most digital replicas of persons will not be able to extend subjective personal experiences of their replicated persons (Iglesias et al., 2024). If, on the other hand, the ability to consciously experience psychological and mental states is a part of, but not a necessary requirement for being psychologically connected, then a zombie replica could, in principle, constitute an extension of our personhood. And this is so irrespective of how weak the psychological connectedness between us and such a replica may be. As long as some psychological connection holds between us and it, some of what constitutes us *us* would be present in the replica. And with that, at least a sliver of what we call *personhood* would be extended to a zombie replica as well. By sliver we here mean a tad of what psychological connectedness consists of – a little of what makes us *us* – but not enough for it to constitute a continuation of the person. Of course, a fully-fledged replica would be much more psychologically connected to its origin than a zombie replica, and a basic zombie replica would be less connected than an advanced zombie replica (for most advanced versions, the difference between a fully-fledged and a zombie replica may even be difficult to tell; Sandberg, 2014).

We will not settle the question of whether having the ability to consciously experience psychological and mental states is a necessary requirement for being psychologically connected here (for a review of arguments in both directions in the debate, see Pernu, 2017). Parfit himself did not elaborate on this in great detail, even though he does talk of *subjects of experiences* and *streams of consciousness* throughout his book, and all Parfit's examples concern human beings and, hence, presumably also conscious beings. In his later writing, Parfit is more explicit about this when he suggests that "any conscious being that can think about itself, and its identity, is ... a person" (2012, p. 23). However, without a further argument, by itself this doesn't imply the reverse: that every person is necessarily a conscious being. Be that as it may, Parfit would certainly agree that losing one's ability to consciously experience psychological and mental states would amount to losing something crucially important of what matters to us deeply when we think about our own survival over time.¹⁷

Relatedly, Stokes, in his exploration of the topic, suggests a fruitful distinction between *personhoods* and *selves*. While we might be able to speak of continued existence in our zombie replicas (e.g., in the way that we proposed above) as *persons*, we would not be able to exist in them as *selves*. The latter is the domain of conscious experiences: only *selves* can anticipate future conscious experiences, consciously experience mental states, and subsequently recall past conscious experiences from their lives (Stokes, 2021a, b). According to Stokes, we thereby simultaneously continue to exist in zombie replicas as *persons* and do not continue to exist in them as *selves*. The idea that digital replicas of deceased persons would be constitutive of *personhoods* has also been suggested earlier by Öhman and Floridi

¹⁷ We thank an anonymous reviewer of our paper for this point.

(2018). This is similar to the conclusions that we reached earlier with Parfit's theory as our starting point. The difference is that, following Parfit, we did not distinguish between *personhoods* and *selves*. Everything is rather captured by a single metric: psychological connectedness. If consciousness is a necessary requirement to be psychologically connected to anything or anyone, then not even slivers of personhoods could be extended to zombie replicas. If, on the other hand, consciousness is only a part of what it means to be psychologically connected, then zombie replicas could, in principle, extend personhoods or slivers thereof. And if consciousness itself is graded, that adds an additional impetus for personhood to be graded too. Our reading and extension of Parfit's work, therefore, offers a distinction between a broader and a narrower interpretation of *personhood* depending on what role consciousness plays in its determination. However, as we said, we will not expand on the role of consciousness further here. Instead, we will consider one other way in which a replica may play a role in extending one's personhood.

3.4 Replicas that Mediate Extensions of Personhood to other Persons

Parfit argued that even after his physical death (which happened in 2017), some future psychological experiences would continue to be related to Parfit's own experiences before he died. There would be memories about his life, thoughts that would be influenced by his, and actions taken upon his advice. As Parfit put it, "[m]y death will break the more direct relations between my present experiences and future experiences, but it will not break various other relations" (1984, p. 281). It is not entirely clear whether by "more direct relations" Parfit meant *all* psychological connections that there may be. However, consider his reflection on how his theory of what ultimately matters for our continued existence changed his view of death (1984, p. 281):

When I believed that my existence was such a further fact [i.e., a fact that constitutes one's personal identity], I seemed imprisoned in myself. My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air. There is still a difference between my life and the lives of other people. But the difference is less. Other people are closer. I am less concerned about the rest of my own life, and more concerned about the lives of others.

Inspired by this view, we next consider the possibility that, following one's physical death, many (perhaps most) psychological connections to any experiences of future persons are immediately broken, but some may nevertheless persist. In other words, a tad of psychological connectedness (a sliver of one's personhood) can continue to hold between a person before her physical death and someone else who continues to exist.

In explicating his theory, Parfit talks of *direct* psychological connections as the necessary ingredient of psychological connectedness. For example, between an entity X today and an entity Y twenty years ago, there may be direct memory connections "if X can now remember having some of the experiences that Y had twenty

years ago.” Other such “direct psychological connections are those which hold when a belief, or a desire, or any other psychological feature, continues to be had” (1984, p. 205). The way we see it – and what we suggest here – is that similar psychological connections could hold also between an entity X and some other entity Y, so long as the relevant psychological features of X are causally linked to the psychological features of Y. Instead of cases in which psychological features “continue to be had,” we can consider cases in which those psychological features are “otherwise causally connected,” for example, when one’s beliefs and desires (indirectly, but nevertheless causally) shape the beliefs and desires of their offspring, or when the beliefs and desires of an influencer shape the beliefs and desires of their followers. More, of course, needs to be said about when thereby causally connected psychological features actually constitute an extension of slivers of X’s *personhood*. For example, the desires and beliefs of X may cause the desires and beliefs in Y, but the desires and beliefs of Y may be entirely different from those of X. Perhaps, as a result of being exposed to X’s desires and beliefs, Y comes to the conclusion that they completely disagree with X’s worldview. Or perhaps Y deliberately forms their desires and beliefs in opposition to those held by X. These cases would not constitute an extension of X’s personhood, or slivers thereof, to Y. It would rather be a case where X’s personhood merely influenced the personhood of Y. In order for X’s personhood to be extended to the personhood of Y, Y should genuinely and sincerely adopt the desires and beliefs of X. It may also be required for the desires and beliefs thereby acquired by Y to directly or indirectly come from X. In other words, Y should learn and adopt those mental states from X rather than come to form similar desires and beliefs entirely independently from X. We will not offer a more detailed account of these considerations here but simply suggest that it is possible in this manner for one’s personhood, or slivers thereof, to extend to other beings beyond one’s physical death.

To Parfit, the above realization was a “liberating, and consoling” experience (1984, p. 281). To us, this brings up another interesting possibility regarding the relationship between us and our digital replicas. Even if the ability to consciously experience psychological and mental states is a necessary requirement for being psychologically connected, a zombie replica, while not itself being an extension of our personhood, may act as a medium through which psychological connections are established and maintained between us and other persons. And this is so irrespective of how weak the psychological connectedness between us and other persons thereby established is. As long as our psychological states, mediated by an advanced zombie replica, can shape and influence the psychological states of other persons, a sliver of our personhood may be extended to other conscious beings (Fig. 1).

If this makes any sense at all, you may wonder whether digital replicas would be unique in being able to mediate psychological connections between persons in this way. We think this need not be the case. A book can play a mediating role of this kind too. Of course, this does not mean that those of us who, after reading *Reasons and Persons* (1984), become influenced by Parfit’s worldview truly become Parfit. But this does suggest that a sliver of Parfit’s personhood may indeed be extended to those who study and are affected in psychologically relevant ways by his work.

In two noteworthy ways, a significantly advanced chatbot replica of someone would nevertheless differ from a book as far as their roles as mediums through which personhoods may be extended are concerned. They have to do with how and to what extent a chatbot and a book can mediate one's personhood to others.

The first concerns the type of data about someone's personhood that is collected in and communicated to other persons by these mediums. A book may or may not contain rich information about the author's psychological and mental states before that information can be communicated to and psychologically affect a reader. Differently from a book, a chatbot replica that is purposefully designed to serve as an accurate representation of someone's character (in order to "speak" on behalf of that person) would be deliberately fed as much data about that person's psychology and mental states as can be mustered (as also noted by Öhman & Floridi, 2017). As such, a chatbot replica, by its design, may often be better "equipped" than a book to mediate someone's personhood to other persons.

The second difference concerns the way in which a chatbot replica and a book perform their mediating roles. A book always presents to us its author's message in words chosen by the author. Those words are then interpreted by us. We might disagree with others on what the correct interpretation of the author's message is and our views about that may change through time. Differently from this, in a conversation with an advanced AI-powered chatbot, the chatbot could present to us the origin's (that is, the replicated person's) message in words that the origin *may have* intended to use (or even choose words that it deems most appropriate in the context, e.g., use simple expressions when communicating with children, or switch to the native language of its conversation partner). Importantly, part of the interpretation of what the origin may have intended to communicate to us will be performed by the chatbot replica itself (for a discussion of this point, see also Danaher & Nyholm, 2024b). For this reason, there are grounds to be particularly concerned about how well the chatbot mediates one's personhood to other persons. The strength of psychological connections that it establishes and mediates between the origin and others will depend on how it interprets what the origin's message may have been. Since part of what is directly communicated and what is left to be interpreted by others lies outside of the origin's control, a chatbot replica will play a significantly different mediating role from one that a book can do.¹⁸

All that said, aren't the conclusions that we reached so far bonkers? If writing a book can extend slivers of one's personhood, should we next consider whether parts of Jesus and Muhammad *in fact* continue to exist in those who follow their teachings? We might instead take this to mean that, by considering what can follow from Parfit's theory as our starting point, we have constructed a *reductio ad absurdum* argument: since the conclusions are absurd, our starting point (or some step in our deductive exploration of the core elements of Parfit's view) must have been false. This is a tenable position. However, if we take it, we will have to readdress some of

¹⁸ The fact that spoken conversation allows one to communicate ideas in a much richer way than is possible through written text (that is also of relevance here) was discussed already a long time ago, for example, in Plato's *Phaedrus* (Plato, 1982).

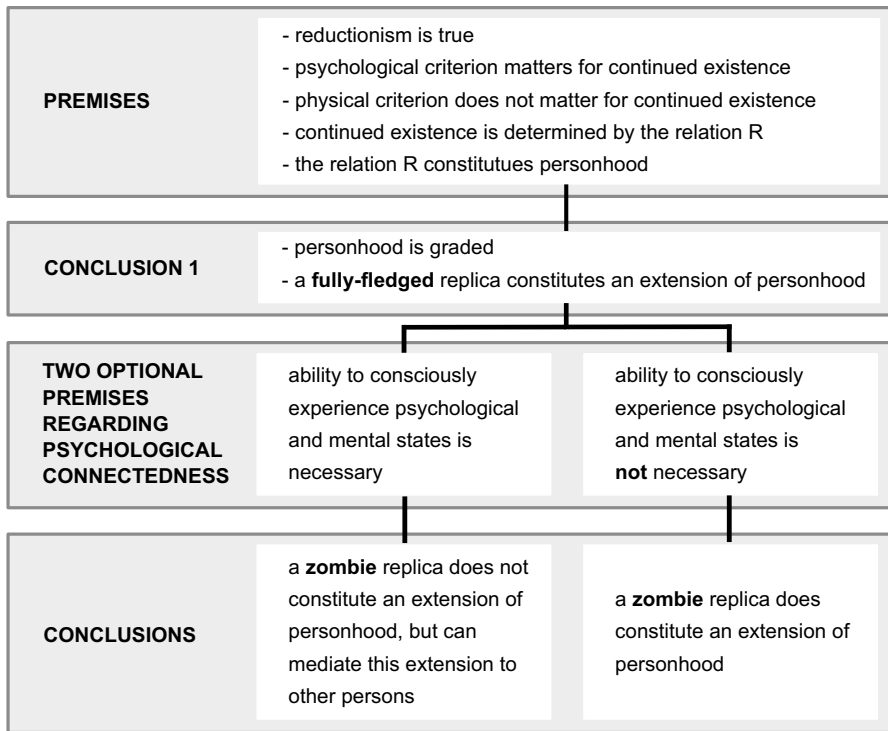


Fig. 1 Summary of the argument so far

Parfit's original challenges. For example, we may need to delineate more rigidly the notion of personhood or reintroduce and, hence, explicate in more detail the relation of personal identity as a defining feature of our continued existence. One possibility is to explore further the idea that the ability to consciously experience mental and psychological states may be a necessary condition to be a person. Another is to introduce the idea of autonomy as a necessary feature of personhood in addition to psychological connectedness. In order for an entity to have personhood, it may be necessary that it can autonomously form its own mental states (beliefs, desires, intentions, worldviews, and so on). If digital replicas of persons are never able to reach this level of autonomy, they will not constitute an extension of our personhood. That said, this would still make it possible for digital replicas to mediate psychological connections between fully autonomous beings.¹⁹

Another option is to bite the bullet and accept the worldview in which slivers of other personhoods continue to exist in us and slivers of our own personhood extend to other persons. In Parfit's metaphor, which we referred to earlier, this is the view of

¹⁹ We thank Kai Spiekermann for bringing our attention to the idea of autonomy. For related arguments, see also Stokes (2021b, pp. 133–134), Schroer and Schroer (2014), and Campbell et al. (2025).

a life in the open air where the walls of the glass tunnel disappeared. That said, we can accept this worldview without drastically changing most of what we are accustomed to in terms of how we think about and treat distinct persons in our daily lives. It may simply be the case that, in order to qualify as the same *person*, the psychological connectedness that holds between temporal instances of that person must be sufficiently strong. And while slivers of one's personhood may indeed extend to other persons, they remain mere slivers and do not constitute a person.

If we choose to bite this bullet, a conclusion that we will have to accept is that grey areas concerning the question of whether an extension of one's personhood qualifies also as a continuation of the person (as opposed to an extension of mere slivers of that personhood) will be bound to remain grey. At one end of the spectrum, psychological connectedness will be sufficiently strong to undoubtedly constitute a continuation of a person. At the other end of the spectrum, it will be weak, constituting a mere sliver of someone's personhood. But in between, there will be cases where the question of whether there is a continuation of a person will, as a matter of fact, not have a categorical answer. In Parfit's words, the question will be empty. We can imagine cases of sufficiently advanced AI-powered digital replicas of persons falling into this "in-between" category (Fig. 2).

4 Ethical and Legal Considerations

Even on the assumption that the ability to consciously experience psychological and mental states is not a necessary requirement to be psychologically connected, psychological connectedness between persons and their digital replicas will most likely be weaker than that between a person at two different points in time (at least that is what we and many others, e.g., Danaher & Nyholm, 2024b, envisage; Fig. 2). Nevertheless – and this includes cases where digital replicas merely mediate slivers of personhood – ethical and legal questions will arise with respect to creating digital replicas of persons. Among many others, there will be those that concern the ownership of data that goes into the creation of a replica, the origin's approval of a replica's creation, and the authorship of whatever replicas produce. There may also be unwelcome consequences of people overestimating or underestimating the degree of psychological connectedness that actually holds between the original person and a replica with which they interact. There is already rich and fast-growing body of literature on the ethics of digital replicas of persons. For many important ideas and arguments, see Sandberg (2014), Öhman and Floridi (2017, 2018), Stokes (2021a, b), Krueger and Osler (2022), Lindemann (2022), Danaher and Nyholm (2024a, b), Hollanek and Nowaczyk-Basińska (2024), Sweeney (2023, 2024), and further references within those works. This list itself is not exhaustive and we will not attempt here to review all ethical matters that arise – and that need to be discussed – when one considers whether to create a digital replica of someone. In our discussion here we will only mention a few points that we believe are novel and ought to be considered in addition to those already discussed in prior works.

4.1 Copyright and Data Privacy

From a legal perspective, the use of copyright-protected data is extensively regulated. However, in the case of new technologies, such as the training of deep neural networks, existing regulations for the most part do not currently prohibit the use of copyrighted content. For example, large language models, such as those used to develop ChatGPT, are often trained on copyrighted data. Motivated by this gap in regulation, a legislative consultation was initiated in the UK to address concerns about the extent to which artificial intelligence (AI) systems may be trained on copyright-protected data.²⁰ In the USA, cases have been brought to court concerning the use of people's copyrighted material as training data (in 2023 the Writers Guild of America went on strike for similar concerns).²¹ There is an urgent need to develop regulatory frameworks for dealing with copyrighted material that is used in the development of AI technologies in general. This is reflected in the European Union's AI Act that puts forward transparency requirements of all General Purpose AI systems, which include compliance with the EU copyright law and the requirement to publish detailed summaries of the content used for training AI systems.

So far, we have imagined replicas that communicate the opinions and views of replicated persons. Their function, however, may extend beyond just that. A digital replica that is capable of composing music in a manner that is similar to the way music was composed by the original (that is, the replicated) person may infringe copyright laws because copyrighted material is used as training data in its creation. This will apply to any type of digital replica that uses the origin's copyrighted creative output to perform its task. Future research and resulting legislative decisions should clarify under what conditions it would be acceptable, for example, to train potential replicas of persons on the works of an author without that author's (or the copyright holder's) prior explicit permission to do so.

The second important legal area that will need to be reviewed is that which governs data privacy. Considering the vast amount of data that people leave about themselves online (the scandals involving Cambridge Analytica in the UK revealed the richness of the data available about our lives on social media platforms), it appears that data privacy is not a major concern to many. Although the investigative artists behind the above-mentioned art project 'Made to measure' (see Sect. 2) had to obtain their anonymous data donor's consent to run the experiment, it is a fact that vast amounts of data about our character traits and psychology are constantly collected and stored in various databases to which we grant these rights by accepting (often mindlessly) their respective terms and conditions.

While this may lie beyond what is (or ought to be) protected by copyright laws, we should nevertheless ask ourselves whether we would want to have a legally

²⁰ <https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/artificial-intelligence-call-for-views-copyright-and-related-rights>

²¹ The comedian Sarah Silverman (and others) recently sued OpenAI for the fact that ChatGPT was trained on copyrighted material extracted from their books (<https://apnews.com/article/sarah-silverman-suing-chatgpt-openai-ai-8927025139a8151e26053249d1aee20>). For coverage of the Writers Guild of America strike, see https://en.wikipedia.org/wiki/2023_Writers_Guild_of_America_strike.

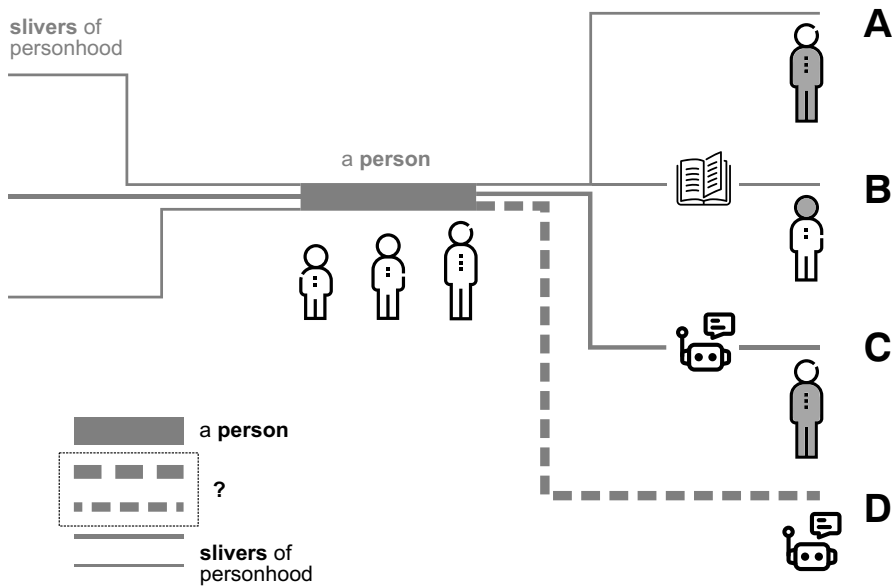


Fig. 2 The “bite the bullet” worldview. The thickness of lines indicates the strength of psychological connectedness. Dashed lines are associated with cases where it is not clear whether psychological connectedness constitutes a mere sliver of personhood or what we would ordinarily call an extension of a person as a whole. In the middle of the diagram is a person at three different points in time. **A–C.** Slivers of this personhood are extended to other persons. The extension is direct (**A**) or mediated by a book (**B**) or a chatbot replica of the person (**C**). **D.** The personhood is extended to a fully-fledged or a “zombie” chatbot replica (the latter if the ability to consciously experience psychological and mental states is not a necessary requirement for being psychologically connected). Whether this constitutes a mere sliver of personhood or an extension of the person is unclear

protected right to decide whether data about our psychology may be used to create a digital replica of us in the future. When creating a replica of a living person, the person’s consent, of course, could be sought beforehand (creating a digital replica of themselves might even be something that many would fancy). That would not be possible, however, with deceased persons. Danaher and Nyholm (2024a) suggest that, mirroring existing copyright protections, deceased people’s or their representatives’ say in how their personhoods are used or developed may weaken over time. This may be a tenable position. However, if one’s digital replica truly constitutes an extension of one’s personhood (and even if it merely mediates the extension of one’s personhood to other persons), we may have stronger grounds to demand the right to have a say concerning the use of data about our psychology, from which a digital replica of us might be created, than the grounds available to us today based on concerns regarding data privacy alone (for related discussion, see also Jurcys et al., 2024). For other works on the idea that people ought to have some meaningful control over what happens with their postmortem reputation has been discussed in the context of chatbots that are created to represent deceased persons to help their living relatives and friends with grieving, see, for example,

Harbinja (2017), Öhman and Floridi (2018), Hollanek and Nowaczyk-Basińska (2024).

4.2 Responsibility

Another issue concerns questions regarding responsibility. As soon as we take our digital replicas to be extensions of our personhood, we have grounds to claim ownership (or at least partial ownership) over whatever they produce. But that comes hand-in-hand with accepting the responsibility for the effects of what they produce too.²² At the very least, this should apply to persons who have explicitly consented to the creation of digital replicas of themselves (Danaher & Nyholm, 2024a, also discuss this point). Whether this should extend to cases of digital replicas created without the consent of the replicated persons is a more difficult question. It may be a stretch to demand the origin to be accountable for whatever their digital replica produce when they did not consent to the replica's creation. This is especially so if the replica was purposefully created not to be the most accurate extension of the origin's personhood, which may be due to someone's commercial considerations (Öhman & Floridi, 2017), defamation, intent to misinform or manipulate others (Danaher & Nyholm, 2024a), or acts of pure hate or revenge (Sweeney, 2024). However, the stronger the psychological connectedness between a digital replica and the replicated person, the more inclined one might be to attribute (at least partial) responsibility to the replicated person, irrespective of whether they gave their consent to the creation of the replica or not. This seems rather repugnant and is particularly worrisome in cases in which replicated persons are still alive. Perhaps, to be held responsible for an event, it is insufficient for one's personhood to be merely causally connected to that event. We will not say more on and will not settle this matter here. For now, we merely raise this question to be addressed in future work.

Additionally, since psychological connectedness and personhood are graded (as, following Parfit, we argued in Sect. 3), responsibility and accountability for whatever replicas produce may either have to be graded too, or might come into play only when psychological connectedness between digital replicas and their origins is sufficiently strong. If the latter, there will be cases in which questions concerning the replicated persons' responsibility and accountability will not have categorical answers (Fig. 2).

Lastly, the persons who produce digital replicas will have to be considered too. After all, most digital replicas will not be produced by the replicated persons themselves and companies that produce them will play a crucial role in determining what those replicas can and cannot do (Öhman & Floridi, 2017). Even if personhoods are primarily or solely constituted by the psychological criterion, they may be intricately tied to our bodies (see Sect. 3.3). Similarly, digital manifestations of our personhoods may be intricately tied to our digital "bodies" – the infrastructure that enables the realization of our digital replicas (Jurcys et al., 2024). As such, the

²² The idea that personhood has forensic implications can be traced back to John Locke (Gordon-Roth, 2020).

co-responsibility of the actual designers and producers of digital replicas will have to be delineated as well.

4.3 Should Replicas be Distinguishable from their Origins?

Imagining that we might in the future be able to create replicas that are easily mistaken for real (replicated) persons, for example, in communications online, would it be acceptable to treat them as representing the replicated persons' views? As mentioned earlier, we envisage that the relation of personhood that holds between origins and their digital replicas will most likely be weaker than that between a person at two points in time. Therefore, we recommend that it should always be made clear that the replica is only a partial extension of a person, or that it constitutes an extension of mere slivers of one's personhood, and that it should be expected to produce statements (or other forms of output) that are sometimes very different from what its origin may have actually stated themselves. Related matters are already considered in the European Union's AI Act. For example, it is required that all artificially created or manipulated images, audio or video content ("deepfakes") must be clearly labelled as such. A similar approach could be taken to regulate people's interactions with digital replicas of persons (Danaher & Nyholm, 2024a, make a similar point).

The concern here is that those who interact with someone's digital replica might attribute too much or too little psychological connectedness to the relation between the replicated person and their replica than actually should be the case (a similar point was made by Hollanek & Nowaczyk-Basińska, 2024). To avoid potential confusions and inaccurate ascriptions of acts or statements that replicas produce to the personhood of their origins, replicas should in some meaningful way disclose to others not only their digital nature, but also the strength of psychological connectedness that holds between them and the origins whose views they represent. It will not be easy to determine that precisely because digital replicas will be meant to interpret what the origins may have intended to communicate or do in specific situations (Sweeney, 2023, discusses such *epistemic gaps* in more detail).

5 Conclusion

AI-powered tools have been used to create "new" songs by Nirvana and Amy Winehouse (Daly, 2021), was consulted to complete Beethoven's unfinished 10th symphony (Elgammal, 2021), and taught to compose chorales in the style of Bach (Hadjeres et al., 2017). In light of the arguments developed in this paper, would these late artists, had they been alive today, have grounds to claim ownership of these AI-enabled creations? Should we care about whether these artists would approve of the new music created in their name?

It seems to us that the answer to both questions might well be "yes." If the AI-powered replicas of the late artists are sufficiently psychologically connected to their origins, then, in principle, a part of what they were able to produce today presumably belongs to their origins as well. If, on the other hand, these replicas

merely act as mediums through which psychological connectedness can be established between the original artists and listeners of the newly generated music, the success of psychological connections thereby established depends on how well the mediums do (or did) their job.

This is not to say that the late artists (or their representatives) would win a lawsuit brought to court on some such grounds. After all, any psychological connection that holds between them and these emergent AI-powered mediums might be so weak that the slivers of personhood that are thereby extended are not worthy of a deeper investigation. But it does seem to us that the late artists' representatives could bring Parfit's book to a courtroom and demand their cases to be heard. In principle, the matters concerning the extent of psychological connections established in these ways and whether they constitute mere slivers of one's personhood or lie within the grey area that is less clear will have to be debated and settled in courts.

Acknowledgements We thank Lucas Battich, Ophelia Deroy, Joerg Fingerhut, Linda Hirsch, Slawa Loev, Mantas Radzvilas, Antonio Scarafone, Kai Spiekermann, and the members of the Cognition, Values, Behaviour, and the Crowd Cognition research groups at LMU Munich for valuable comments and discussions.

Authors' Contributions Both authors contributed equally to the development of this paper.

Funding Open Access funding enabled and organized by Projekt DEAL. Jurgis Karpus was supported by LMUexcellent, funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Exchange Strategy of the Federal Government of the Länder.

Data Availability Not applicable.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors have no competing interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bostrom, N. (2003). Are we living in a computer simulation? *Philosophical Quarterly*, 53(211), 243–255. <https://doi.org/10.1111/1467-9213.00309>

- Campbell, S. M., Liu, P., & Nyholm, S. (2025). Can chatbots preserve our relationships with the dead? *Journal of the American Philosophical Association*. <https://doi.org/10.1017/apa.2025.1>
- Cerullo, M. A. (2015). Uploading and branching identity. *Minds and Machines*, 25, 17–36. <https://doi.org/10.1007/s11023-014-9352-8>
- Daly, R. (2021). AI software writes new Nirvana and Amy Winehouse songs to raise awareness for mental health support. *NME*. <https://www.nme.com/news/music/ai-software-writes-new-nirvana-amy-winehouse-songs-raise-awareness-mental-health-support-2913524>
- Danaher, J., & Nyholm, S. (2024). The ethics of personalised digital duplicates: A minimally viable permissibility principle. *AI and Ethics*. <https://doi.org/10.1007/s43681-024-00513-7>
- Danaher, J., & Nyholm, S. (2024). Digital duplicates and the scarcity problem: Might AI make us less scarce and therefore less valuable? *Philosophy and Technology*, 37, 106. <https://doi.org/10.1007/s13347-024-00795-z>
- Dennett, D. (2023). The problem with counterfeit people. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2023/05/problem-counterfeit-people/674075/>
- Elgammal, A. (2021). How a team of musicologists and computer scientists completed Beethoven's unfinished 10th symphony. *The Conversation*. <https://theconversation.com/how-a-team-of-music-ologists-and-computer-scientists-completed-beethovens-unfinished-10th-symphony-168160>
- Farina, M. (2021). Embodied cognition: Dimensions, domains and applications. *Adaptive Behavior*, 29, 73–88. <https://doi.org/10.1177/2F1059712320912963>
- Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin and Review*, 23, 959–978. <https://doi.org/10.3758/s13423-015-0860-1>
- Gordon-Roth, J. (2020). Locke on personal identity. In: Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy (Spring 2020 Edition)*. <https://plato.stanford.edu/archives/spr2020/entries/locke-personal-identity/>
- Grow, K. (2021). In computero: hear how AI software wrote a 'new' Nirvana song. *RollingStone*. <https://www.rollingstone.com/music/music-features/nirvana-kurt-cobain-ai-song-1146444>
- Hadjeres, G., Pachet, F., & Nielsen, F. (2017). DeepBach: a steerable model for Bach chorales generation. *Proceedings of the 34th International Conference on Machine Learning*, 1362–1371.
- Harbinja, E. (2017). Post-mortem privacy 2.0: Theory, law and technology. *International Review of Law, Computers and Technology*, 31, 26–42. <https://doi.org/10.1080/13600869.2017.1275116>
- Heaven, W. D. (2020). OpenAI's new language generator GPT-3 is shockingly good—and completely mindless. *MIT Technology Review*. <https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/>
- Heersmink, R. (2017). Distributed selves: Personal identity and extended memory systems. *Synthese*, 194, 3135–3151.
- HollanekNowaczyk-Basińska, T. (2024). Griefbots, deadbots, postmortem avatars: On responsible applications of generative AI in the digital afterlife industry. *Philosophy and Technology*, 37, 63. <https://doi.org/10.1007/s13347-024-00744-w>
- Iglesias, S., Earp, B. D., Voinea, C., Mann, S. P., Zahu, A., Jecker, N. S., & Suvaescu, J. (2024). Digital doppelgängers and lifespan extension: What matters? *The American Journal of Bioethics*. <https://doi.org/10.1080/15265161.2024.2416133>
- Ishiguro, K. (2021). *Klara and the Sun*. Faber and Faber.
- Jurcys, P., Greenwald, A., Fenwick, M., Loikkanen, V., Porsdam Mann, S., & Earp, B. (2024). Who owns my AI twin? Lights and shadows of data ownership in a new world of simulated identities. Pre-print available at SSRN: <https://doi.org/10.2139/ssrn.4940663>
- Krueger, J., & Osler, L. (2022). Communing with the dead online: Chatbots, grief, and continuing bonds. *Journal of Consciousness Studies*, 29, 223–252.
- Kurzweil, R. (2000). Live forever, uploading the human brain, closer than you think. *Psychology Today*. Available at: <https://www.thekurzweillibrary.com/live-forever-uploading-the-human-braincloser-than-you-think>
- Lindemann, N. F. (2022). The ethics of 'deathbots.' *Science and Engineering Ethics*, 28, 60.
- Mazumder, A., Sahed, M. F., Tasneem, Z., Das, P., Badal, F. R., Ali, M. F., Ahamed, M. H., Abhi, S. H., Sarker, S. K., Das, S. K., Hasan, M. M., Islam, M. M., & Islam, M. R. (2023). Towards next generation digital twin in robotics: Trends, scopes, challenges, and future. *Heliyon*, 9, E13359. <https://doi.org/10.1016/j.heliyon.2023.e13359>
- Murphy, M. (2019). This app is trying to replicate you. *Quartz*. <https://qz.com/1698337/replika-this-app-is-trying-to-replicate-you/>

- Nakagawa, H., & Orita, A. (2022). Using deceased people's personal data. *AI and Society*. <https://doi.org/10.1007/s00146-022-01549-1>
- Nishio, S., Ishiguro, H., & Hagita, N. (2007). Geminoid: teleoperated android of an existing person. In: de Pina Filho, A. C. (ed.), *Humanoid Robots: New Developments*. IntechOpen.
- Öhman, C., & Floridi, L. (2017). The political economy of death in the age of information: A critical approach to the digital afterlife industry. *Minds and Machines*, 27, 639–662. <https://doi.org/10.1007/s11023-017-9445-2>
- Öhman, C., & Floridi, L. (2018). An ethical framework for the digital afterlife industry. *Nature Human Behaviour*, 2, 318–320. <https://doi.org/10.1038/s41562-018-0335-2>
- Olson, E. T. (2024). Personal identity. In: Zalta, E. N. and Nodelman, U. (eds.) *The Stanford Encyclopedia of Philosophy (Winter 2024 Edition)*. <https://plato.stanford.edu/archives/win2024/entries/identity-personal/>
- Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.
- Parfit, D. (2012). We are not human beings. *Philosophy*, 87, 5–28.
- Pernu, T. K. (2017). The five marks of the mental. *Frontiers in Psychology*, 8, 1084. <https://doi.org/10.3389/fpsyg.2017.01084>
- Plato. (1982). *Plato I: Euthyphro*. Translated by Fowler, H. N. Harvard University Press.
- Rust, J. (2019). Institutional Identity. *Journal of Social Ontology*, 5, 13–34. <https://doi.org/10.1517/jso-2018-0032>
- Sandberg, A. (2014). Ethics of brain emulations. *Journal of Experimental and Theoretical Artificial Intelligence*, 26, 439–457.
- Schrage, M. (2020). *Recommendation Engines*. The MIT Press.
- Schroer, J. W., & Schroer, R. (2014). Getting the story right: A Reductionist narrative account of personal identity. *Philosophical Studies*, 171, 445–469.
- Schwitzgebel, E., Schwitzgebel, D., & Strasser, A. (2023). Creating a Large Language Model of a Philosopher. *Mind and Language*, 1–22. <https://doi.org/10.1111/mila.12466>
- Schwitzgebel, E. (2021). Two robot-generated splintered mind posts. *The Splintered Mind*. <http://schwitzt splinters.blogspot.com/2021/11/two-robot-generated-splintered-mind.html>
- Stokes, P. (2021). Death and survival online. In M. Cholbi & T. Timmerman (Eds.), *Exploring the Philosophy of Death and Dying: Classical and Contemporary Perspectives* (pp. 200–207). Routledge.
- Stokes, P. (2021b). *Digital Souls: A Philosophy of Online Death*. Bloomsbury Publishing.
- Strasser, A., Crosby, M., & Schwitzgebel, E. (2023). How far can we get in creating a digital replica of a philosopher? In: Hakli, R. Mäkelä, P., and Seibt, J. (eds.), *Social Robots in Social Institutions. Proceedings of Robophilosophy 2022*. Series Frontiers of AI and Its Applications, vol. 366, 371–380. IOS Press, Amsterdam. <https://doi.org/10.3233/FAIA220637>
- Strasser, A. (2024). DIGITAL LAB DGPhil 2024: DigiDan installation [video recording]. Available at: <https://youtu.be/rNEHzQ15tJQ?feature=shared>
- Sweeney, P. (2023). Avatars as proxies. *Minds and Machines*, 33, 525–539. <https://doi.org/10.1007/s11023-023-09643-z>
- Sweeney, P. (2024). The ethics of ex-bots. *AI and Society*, 39, 3055–3056. <https://doi.org/10.1007/s00146-023-01754-6>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.