

Harm as part of existence: No escape in a technology-driven alternative

WN Reynecke

 **orcid.org/0000-0003-2883-9046**

Mini-dissertation accepted in partial fulfilment of the requirements for the degree *Master of Arts in Philosophy* at the North-West University

Supervisor: Dr J du Toit

Graduation: August 2021

Student number: 13032585

Contents

1	Study Introduction	6
1.1	Research question	7
1.2	Thesis statement	7
1.3	Study aim	7
1.4	Methodological background	8
1.5	Study overview	8
1.6	Scope of study	9
2	Chapter 2: The preponderance of harm in existence	10
2.1	Benatar, anti-natalism, and the concept of harm	11
2.2	The worth of life	12
2.3	The Relational Asymmetric Principle (RAP)	15
2.3.1	RAP supporting arguments	17
2.4	An attempt at the quantification of harm	19
2.4.1	Four metrics	20
2.5	Psychological aspects of Benatar's argument	22
2.5.1	Pollyannism	23
2.5.2	Habituation	23
2.5.3	Comparison	24
2.5.4	Summary	24
2.6	Three theories on the quality of human life	25
2.6.1	Hedonistic theories	26
2.6.2	Desire-fulfilment theories	27
2.6.3	Objective-list theories	28
2.7	Chapter Conclusion	29
3	Chapter 3: A critical evaluation of Benatar's concept of harm	31
3.1	A voice of dissent: Harman's argument	31
3.2	Alternative methods for assessing a life's worth	33
3.2.1	Belshaw and alternative methods for assessing life's worth	34
3.2.2	Boonin's view on alternative methods	35
3.3	Chapter Conclusion	37

4	Chapter 4: Technology-driven alternatives to existence	38
4.1	The governance of human life by algorithms	41
4.1.1	What are algorithms?	41
4.1.2	Algorithms and existential harm.....	43
4.1.3	The shortcomings of algorithmic technologies for overcoming harm	46
4.2	Hive minds and cyborgs	47
4.2.1	What are hive minds and cyborgs?.....	48
4.2.2	Hive minds, cyborgs and existential harm	50
4.2.3	The shortcomings of hive minds and cyborgs for overcoming harm	54
4.3	Artificial Bodies and Robots	55
4.3.1	What are artificial bodies and robots?	55
4.3.2	Artificial bodies, robots, and existential harm.....	56
4.3.3	The problem of Artificial Intelligence as the other	61
4.3.4	The shortcomings of artificial bodies and robots for overcoming harm	63
4.4	Technological alternatives and the negation of harm	63
4.5	Chapter Conclusion	70
5	Chapter 5: Can harm be negated?.....	72
5.1	Return to the RAP.....	73
5.2	Why does harm remain?	76
5.2.1	Algorithmic control: Algorithmic technologies can, and will, fail us	76
5.2.2	Hive minds and cyborgs: What remains (of us) in the Collective?	80
5.2.3	Artificial Bodies and Robots: Over my dead body	82
5.3	Technological alternatives to existence present no harm-free utopias	85
5.4	Experiencing harm is human, and that's (not) okay	86
5.5	Chapter Conclusion	87
6	Study Conclusion	89
7	References	92

Abstract

This study serves as a response to, and remediation of, utopian conceptualizations of technology-driven alternatives to human existence. Such conceptualizations often present harm-mitigation as a central criterion for the validity of their approaches, an assumption that this study will question.

I firstly investigate the nature of harm as identified in Benatar's concept of anti-natalism, particularly the notion that we have different measurement criteria for assessing a life's worth depending on whether it is a life to be started (before birth – future-life) or a life to be continued (when it is already underway – present-life). I also reflect on humanity's positive view of life and its pleasures due to overly optimistic psychological states and their underlying biasedness.

The study then traces a line from such harm, as experienced existentially in everyday life, to the hypothetical alternatives to everyday existence made possible by technology. The measurement tool of the Relative Asymmetric Principle (RAP), introduced by Benatar, is analyzed to show its applicability in assessing alternative forms of existence. Consequently, it is used to assess an existence's worth when it takes on alternative forms (i.e. technology-driven forms). I then illustrate that an alternative existence through technological alternatives may overcome (or reduce) some of the aspects of harm, whilst other aspects of harm remain inherent to existence (whether biological or technological). Reviewing the shortcomings of technology with the RAP, I conclude that we should be implored not to seek salvation in technology. We should instead abandon our quasi-teleological relationship with technology and instead see it for what it is: a powerful and often dangerous tool.

In this study I employ a philosophical critical method based on close reading and critical analysis. I consider the incidences of harm when allowing for an anti-natalist policy of existence as postulated by Benatar and evaluate its possibilities in accordance with technology-driven alternative existences through the application of the standards set out by Benatar to determine a life's worth.

Key words

Benatar, anti-natalism, harm, philosophy of technology, relational asymmetric principle

1 Study Introduction

“Life is pain – anyone who says differently is selling something” (Reiner, 1987).

The experience of existential badness during a human life results in much harm for both oneself and for other people. This study attempts to conceptualize, as a starting point, what this harm that the biological human individual experiences may entail and how individuals attempt to overcome such harm by means of technology.

From the outset, we may argue that such harm entails not just crying about a long lost love, or feeling the desolation of an open grave. Existential harm as concept is much more nuanced and subtle, and the identification and explication of this concept is made all the more difficult through the capacity of the human mind (this beautiful brilliant gooey mass with billions of neurons firing) to pull us silently through the harms of life and to make us unaware of the importance of such harms (particularly through psychological adaptation).

It makes sense in the context of the omnipresence of such existential harms that humans strive to escape from such harms, often through the postulation of some form of messianic thinking. It also makes sense; therefore, that humanity is constantly looking towards something new, a new Messiah to set us free. Particularly in the Modern era, we find that the lure of technology as Messiah has attracted our human aspirations – even though the technological character of such aspirations has shifted since the industrial revolution. Being crafty with a bow drill, for example, will not assist me in getting a 5-year old toddler fed before bedtime; a refrigerator, microwave and Disney film reruns, on the other hand, certainly will.

Human-machine relations have become an encompassing background phenomenon in contemporary societies and existence (Ihde, 1975:277-278) – we often do not recognize how broadly it affects our experience of the world, and how necessary a part of modern living it has become. Furthermore, many individuals have blindly accepted technology as a positive element in the overcoming of existential harms (Zuboff, 2015:82). It is perhaps this dependence on technology which has led to us turning a blind eye to the open graves that technology may yield, think for example of teenagers absorbed in online lives who are absent

to a great extent from their families' social lives. Surely, these teenagers, absorbed in their online existences, have not succeeded in overcoming the harm that is inherent to existence (perhaps, they experience it most keenly), such as difficult family lives, or pressing school studies.

1.1 Research question

Does technology provide a means to overcome existential harm?

1.2 Thesis statement

Harm forms an inherent part of existence, whether such an existence is based on mere biological existence or on a technology-driven alternative.

1.3 Study aim

In this study, I will investigate the nature of harm as ubiquitous in our existence. I will then investigate the postulation of technological 'refuges' as potential avenues to overcome such harms. I argue that if we understand harm as inherent in human life, we are better equipped to understand the shortcomings that our technology obsession may carry along with it.

This study serves as a response to, and remediation of, utopian conceptualizations of technology-driven alternatives to human existence. Such conceptualizations often present harm-mitigation as a central criterion for the validity of their approaches, an assumption that this study will question. My study will seek to trace a line from harm as experienced existentially in everyday life to hypothetical alternatives to said existence through technology.

In this study I will illustrate that an alternative existence through technological expressions may overcome (or reduce) some of the aspects of harm, whilst other aspects of harm remain inherent to existence (whether biological or technological). I conclude that harm is present wherever existence figures – whatever form that existence consists of – and therefore even

arguments for alternative forms of existence should be avoided due to the inescapability of harm inherent in existence.

1.4 Methodological background

In this study, I employ a philosophical critical method, based on close reading and critical analysis. I consider the incidences of harm when allowing for an anti-natalist policy of existence as feasible through technology-driven alternative existences, with the reduction of harm as deriving from Benatar's anti-natalism as a point of initial departure.

1.5 Study overview

In Chapter 2 of this study, I examine anti-natalism (as school of thought) as a proposed solution to the everyday harm experienced by individuals in modern society. I present a condensed definition of the concept of harm in relation to the notion of anti-natalism.

In Chapter 3 I investigate whether Benatar's anti-natalism adequately addresses the harm that individuals experience in their everyday lives. I approach this question by engaging with the major criticisms to his argument, as well as reflecting on the touch-points and departures that respective critics present.

In Chapter 4 the possible character of an existence based on technological alternatives is evaluated – particularly in terms of contemporary and hypothetical future developments in this regard. In addition, I investigate whether an existence based on technology-driven alternatives – considered in light of Benatar's anti-natalism argument – could serve as a sensible avenue to overcome the harm associated with existence.

In Chapter 5 of this study, I evaluate whether such alternative existences based on technological developments ultimately result in the introduction of new harms, or at least, a net reduction of harm.

1.6 Scope of study

This study will focus on the concept of harm as described in the works of Benatar, particularly in terms of harm as inexorably leading to the anti-natalist hypothesis that he proposes. The study touches on elements of phenomenology as it relates to human existence and harm. Philosophy of technology also forms a significant part of the study where technology-driven alternatives to existence are investigated.

2 Chapter 2: The preponderance of harm in existence

Anti-natalism (as school of thought) is argued to serve as a response to the everyday harm experienced by individuals in modern society. What does this concept of harm, as basis of these arguments, entail? Furthermore, does Benatar's anti-natalism adequately address the harm that individuals experience in their everyday lives?

The aim of this chapter is to define the concept of harm (as discussed by David Benatar) in relation to the notion of anti-natalism. I will explore the concept of anti-natalism as put forth by Benatar in *Better Never to Have Been: The Harm of Coming into Existence* (2006).¹ The anti-natalist position assigns a negative value to birth, due to the persistence of harm in existence. It is a school of thought, which is a response to the harm that arises from coming into existence, and also, an attempt to reduce it. In this chapter, I review two questions posed and addressed by Benatar: i) why coming into existence is *always* a harm, and ii) *how bad* is coming into existence.

From the outset one must note that Benatar does not deny that there is good² in life, and he does not argue that there is nothing beneficial in our existence (Benatar, 2006:1). Thus, Benatar shows an awareness of a balance of both existential good and bad, though in his line of argumentation the bad outweighs the good and it is the bad that should be avoided. Benatar suggests that it should be our aim to reduce this badness – the harm – that forms part of life, to a minimum. This is, however, an impossible feat (Benatar, 2006:2). Even if we acknowledge the good in life, he argues that the bad will *always* outweigh the good. In mathematical (or rather, transactional) terms, this may be stated through the metaphor of a

¹ Anti-natalism is not a new concept; it is only the specific name that appears to be a new invention (Morioka, 2021:3). As Morioka (2021:3-4) indicates, the concept had been known in Eastern cultures / religions well before its entry into Western thought in ancient Greek times. It was Western thinkers such as Schopenhauer, Nietzsche and Cioran who worked with the topic before Benatar crystalized it in the form, which is engaged with in this study.

² The terminology in this study around 'good', 'bad' and 'not bad' always refer to the context from Benatar's core text, *Better Never to Have Been* (2006). I also retain Benatar's formatting of capitalizing the word when referring to his assessment of a life's worth as discussed in Section 2.3: thus *Good*, *Bad* and *Not Bad*. I explicitly mention that whenever the formatting deviates (e.g. GOOD, BAD or *Good*, *Bad*) the reference to Benatar remains. Therefore, I do not make any indirect reference to the Ethics context of 'Good'.

scale, whereby the 'scale' of existence consists of benefits on one end and harms on the other with the scales overwhelmingly shifted in the direction of harm to reveal a net harm tally.

The argument for an anti-natalist approach to existence and this conclusion of net harm will be important later in this study when I evaluate the possibility of harm reduction in relation to technological alternatives to our biological existence.

2.1 Benatar, anti-natalism, and the concept of harm

Before Benatar (2006:18) presents his argument that coming into existence is *always* a harm, he attempts to prove its opposite by asking, whether coming into existence is *ever* a harm. This is a central building block for his argument, and provides a simpler target of attack, since if there is even a single instance of harm (*ever* harmful) in an individual's life then the claim that coming into existence is *ever* a harm is shown to be false. This line of inquiry by Benatar reveals that, at least some of the time, coming into existence is harmful. Benatar suggests that such harm (any harm) should be avoided at all costs, even to the extent of not allowing that life to be lived – to 'not exist', per the anti-natalist position that he is developing.³

As Benatar (2006:18) notes, this condition of identifying something that is *ever* a harm indicates a weak requirement, however, and it is not challenging to come up with examples where an individual's existence will *necessarily* lead to harm. In fact, it is therefore easy to point to cases where harm is inseparable from the individual's existence and even one instance of harm will support this argument. Benatar (2006:19) presents two examples to illustrate his point. Firstly, the instance of a child born with a genetically inherited disease and, secondly, the instance of a child born to a teenage mother incapable of sufficient child rearing due to financial or social reasons (Benatar, 2006:19). In each of these examples, the weak requirement for the validity of the anti-natalist argument is the inherentness of the harm in the postulated existences; that is, harm is tied to the specific individual and their

³ Schopenhauer can be seen as reasoning from a similar, if not the same, point of departure when he expresses 'to live is to suffer' (Jacquette, 2000:43). However, as can be concluded from Benatar's reasoning, the latter aims to prevent the onset of this suffering / harm, whereas Schopenhauer appears to accept its manifestation as a given. Herewith I mean that the onset of a life is inevitable, since it is argued that death cannot be avoided and 'existence is a constant dying' (Jacquette, 2000:43).

specific existence in each case.⁴ For Benatar, it follows that if the parents in such instances choose to terminate the pregnancies, they have indeed reduced the pending harm of the child in question. A subsequent pregnancy wherein a new child's genetic testing is deemed 'healthy' (first example), or wherein the parents are more mature and financially established (second example), does not imply a reduction in harm for this specific child simply because it appears at face-value that the parents have reduced the potential harms. This is because the subsequent pregnancy is obviously a completely new being and therefore cannot be considered as a reduction of harm of the initial specific existence.

From the above-mentioned discussion we may therefore conclude that the requirement of existence *ever* being harmful is a weak one, which suggests the need (as Benatar states) to work towards an argument in the stricter sense – that existence is *always* harmful. The next section presents Benatar's formulation of this stricter requirement.

2.2 The worth of life

Benatar (2006:21) counters the view that 'coming into existence cannot be a harm' in the opening statements of his anti-natalist argument, suggesting that such an argument is simplistic and incomplete. The argument, of which Benatar is critical, may be stated as follows:

- i. Harm suggests that certain elements of the hypothesized state in which the individual is, are worse than those of some alternative state;
- ii. If existence is then taken to be the hypothesized state, then the alternative state can only be non-existence by mutual exclusion;
- iii. However, the comparison is invalid because no individual can non-exist;

⁴ Benatar (2006:7,19) is very explicit in terms of the underlying concept here. The specific individual is a product of two specific gametes from each of the parents. Therefore, to abort a foetus which tested positive for a genetic disease, or an untimely teenage pregnancy, means that that specific individual's existence has been denied indefinitely even though the parents may birth another child (a healthy child, or a child later in life), since it is a different child.

- iv. Thus, coming into existence cannot be a harm.

Benatar argues that the argument made above is invalid, however. The concept of 'worse' is troublesome in his opinion, because if the hypothesized state is merely 'bad' and the alternative state is 'not bad', then the hypothesized state of 'worse' is already presupposing harm to the individual. For Benatar, the argument may be stated as follows:

- i. Harm suggests that certain elements of the hypothesized state in which the individual is, are worse than those of some alternative state;
- ii. However, if the hypothesized state is merely 'bad' and the alternative state is 'not bad';
- iii. Then the hypothesized state is already introducing harm to the individual;
- iv. Thus, coming into existence can be a harm.

Thus, in Benatar's view, given the introduction of harm, we should conclude that the life is indeed not worth living because coming into existence necessarily entails the presence of harm. Yet there is a further tension that he notes: while not of a comparative nature, there is nevertheless a clear 'binarity' being presupposed. This binarity is anchored in a further measure of the assessment of a life's worth, and more specifically; the measure of the life's *timing*. With this, he suggests that we need to be cognisant of the *timing* of the assessment of a life's worth. I expound this binarity of the timing in the remainder of the section below.

The terminology utilized by Benatar (2006:22) shows a much clearer link between the preceding notions of 'existence' vs. 'non-existence'. In Benatar's view, the underlying timing aspect associated with the life's worth should be explicitly distinguished at a more granular level as: i) a life worth continuing; and ii) a life worth starting. It is in this respect where the aforementioned binarity is found, where the assessment is timed either to be while the life is underway, or it is timed to be before the life commences. The link between, respectively, a life worth continuing and the concept of 'existence', and a life worth starting and 'non-existence', follows naturally. This distinction is significant since Benatar argues that there exist

different thresholds in determining the worth of a life since the assessment is in fact based on whether the life is worth *starting* or worth *continuing*.

Benatar turns to an example of an individual with a missing limb (which is illustrative of how the timing of the worth assessment of the life is essential for setting the aforementioned thresholds). Firstly, he considers the case of a living person who loses a limb through some medically necessary amputation (Benatar, 2006:23). In his view, one can surely motivate an adult amputee to find courage to continue his or her life after the unexpected need for amputation. Secondly, he considers a case where medical screening is used to detect that a developing foetus may be missing a limb due to some abnormal development. Benatar points out that one will probably question whether the continuation of the pregnancy is preferable, and thus whether it is preferable to start a new life without a limb.⁵

The key to the assessment of the worth of living is what he refers to as ‘present-life sense’ and ‘future-life sense’, and how ultimately such considerations are incompatible in assessing a life’s worth (Benatar, 2006:24). This is because we assess a future-life instance (thus in the state of non-existence: starting the life) with present-life principles (relating the instance to a state of existence: continuing the life), which presents a situation of utilizing incompatible criteria of evaluation in terms of life’s worth.⁶

The central aspect to keep in mind here is the awareness and appreciation, in one’s argument, that present lives and future lives require different forms of assessment because our presumed thresholds for such assessments cannot be the same. These initial considerations point the way to the stricter interpretation of the anti-natalist argument which states that coming into existence is *always* a harm; based on the idea that the harm inherent in respectively future-life and present-life sense, should be assessed in different ways. This brief

⁵ Benatar (2006:23) presents a further so-called ‘trivial’ example wherein the ‘continuing’ argument is applied to watching a bad movie. The viewer can certainly feel that he should never have gone to see the film, but likely will not be dissatisfied enough to walk out of the film before the ending. I can imagine a similar situation should one be dissatisfied with a meal being consumed. Depending on the cost of the meal and alternatives available, one will probably finish the meal and then avoid the same choice in future, rather than straight-out immediately disposing of the food.

⁶ Ibid. Deeper argumentation is pre-empted by Benatar (2006:7) to my own objection, which I am sure the reader may pose as well: when does existence start? In a truly biological sense: is the moment of a zygote forming not the moment of existence’s start? In his view it is not: existence, in what he terms the ‘morally relevant sense’, is progressive, not instantaneous. I note this anomaly and his argumentation to address it here, but consider it beyond the scope of this study.

overview of the worth of life according to Benatar's anti-natalism has provided both a tool and a starting point for assessment of harm in relation to existence, which Benatar develops through the introduction of the Relational Asymmetric Principle.⁷

2.3 The Relational Asymmetric Principle (RAP)

Benatar (2006:29) wholeheartedly rejects the notion that a life can ever only contain good; that is, that it can be beneficial in its entirety. One may generally concur with this, as it is rather easy to agree that 'all lives contain some bad'. It should be noted that, although I will consider an attempt at 'quantification' in terms of the extent of this 'badness' (see Section 2.4), the mere presence of harm is deemed sufficient to assess coming into existence as a harm from the outset according to Benatar. Specifically, in this section it will be shown that coming into existence will be assessed as *always* a harm according to the anti-natalist position. The reason for the omnipresent '*always a harm*' in the assessment by Benatar is due to the fact that when we consider the balancing scales of all the benefits (the good in life) in their entirety and all the harms (the bad in life) in their entirety, a stable equilibrium between the opposing sides is impossible – both existential good and existential bad dynamically shape the experience of one's life. Furthermore, this impossibility of equilibrium in one's experience, in turn, results in the inherent asymmetry that is postulated to exist between the good and the bad. This is described in the Relational Asymmetric Principle (RAP) that Benatar develops.⁸ The RAP is based on the asymmetry that Benatar identifies in relation to pleasure and pain, stated logically as follows (Benatar, 1997:345):

- i. The presence of pain is bad.
- ii. The presence of pleasure is good.

⁷ Keep in mind that the starting point for this assessment is based on the argument that coming into existence can *ever* be a harm – as developed by Benatar.

⁸ This terminology does not originate with Benatar, but instead with one of his correspondents, Boonin (2012:13). I maintain this terminology here and in later sections for uniformity. Section 3.2.2 provides more details on Boonin.

- iii. The absence of pain is good, even if that good is not enjoyed by anyone.
- iv. The absence of pleasure is not bad unless there is somebody for whom this absence is a deprivation.

In assessing whether existence should be permitted, it is crucial to determine whether there is harm or benefit *present* in said existence. Benatar (2006:30) explains that there is an underlying dimension here; namely that of the element – whether good or bad – of presence (that is, of being ‘present’). Paying attention to this further dimension is significant, because instances that state the reverse (permitting existence, when we are assessing existence in terms of harm or benefit being *absent* in that existence) may naturally arise as well. It is with this reversed state of affairs that the asymmetry of the current assessment becomes apparent, as proposed by Benatar in the following diagram:

Table 1: Diagrammatic illustration of the RAP

Existence	Non-existence
(1) Presence of pain (BAD)	(3) Absence of pain (GOOD)
(2) Presence of pleasure (GOOD)	(4) Absence of pleasure (NOT BAD)

Source: Benatar (2006:38)

From the above diagram it follows that it is good (or even beneficial) when pleasure is *present* (quadrant (2)), just as it is bad (or harmful) when pain is *present* (quadrant (1)). However, we can presume that it is good (or beneficial) when pain is absent (quadrant (3)). Similarly, we will also presume that it is bad (or harmful) when pleasure is absent (quadrant (4)). I argue, in concurrence with Benatar, that the latter two assumptions present reasonable claims. Nevertheless, when we return to the overarching theme of existence vs. non-existence and set the dimensions of *present* and *absent* as respective proxies, the asymmetry is revealed. Benatar argues that it is still good (or beneficial) when pain is absent despite non-existence (quadrant (3)); thus even when there is no existing individual to benefit from the absence of

pain. Though, one may ask whether it is still bad (or harmful) when pleasure is absent in the case of non-existence, when there is no existing individual to be harmed by the absence of pleasure (quadrant (4)). We cannot reasonably conclude that this would be the case, because non-existence presupposes that neither good nor bad experience is possible.⁹

As further support for the RAP, I now discuss four additional supporting arguments for the inherent asymmetry intuited from the foregoing discussion as presented by Benatar. These arguments are supporting in the sense that they serve to illustrate the nuance of the point that Benatar is developing in the RAP.

2.3.1 RAP supporting arguments

A first supporting argument that Benatar suggests is presented from a type of parental point of view of procreation where it is stated there is obviously an obligation, at least in the moral sense, not to bring people into the world that will experience harm. Benatar (2006:32) expands on this point of view by adding that there is, however, no obligation to bring people into the world that will be benefitted (so-called 'happy people').¹⁰ It follows from the RAP that even though their happiness or benefits experienced would be good for them, the instance in which they do not exist suggests that the concomitant absence of the benefits is not a harm to them.

A second supporting argument follows from the belief that one would naturally wish to avoid harm befalling any child, for that child's sake (Benatar, 2006:34). Postulating a counter-

⁹ According to Bolea (2019:79), Cioran provides a different take on the matter, which I nevertheless consider somewhat analogous to Benatar's approach. Cioran postulates a distinction between humankind and animals on the basis that only humankind can be something other than itself: Cioran defines this as being the 'not-man'. The not-man has the ability to transcend its biology, which the animal cannot. The analogy I draw with Benatar's argumentation is that of the transcendence to which Cioran refers. The ability to transcend from the biologic to 'being (not unlike divinity)' in Cioran relates to the transcendence of the transcendence from the existential to the non-existential state in Benatar (Bolea, 2021:80).

¹⁰ An interesting shortcoming in this support argument may arise where Benatar (2006:32) explains the individual certainly has the 'negative' duty to avoid harm (such as by bringing people into the world who will suffer in some way). Whereas the case of a positive duty to create good (such as by bringing people into the world who will experience pleasure in some way), is not positive in its entirety, not even neutral. This is the case since there will be some negative elements in the duty – 'sacrifice' to use his term – where the pregnant individual will have to forego some positive aspects of their normal existence in order to complete the supposedly positive duty: e.g. the discomfort of pregnancy and giving birth.

argument here is more doubtful (if not subconsciously egotistical): Is it possible to wish for the benefit to be experienced by some specific child purely for the sake of that specific child? Reformulating this point in such a manner seems less plausible: for if we start to argue about the child being in a loving family, or being the child of some illustrious individual, or being able to enjoy a good life in a rich and safe country, we only reference external elements and not the child itself.

Benatar presents a third supporting argument for the RAP that is concordant with a topic introduced in Section 2.2, namely that of the timing of the assessment of a life's worth. Continuing from the preceding supporting argument and its discussion regarding for whose sake existence is questioned; Benatar (2006:34) builds this third supporting argument on the basis of how the act of either allowing or disallowing existence to occur (creating people) can be deemed regrettable for different reasons. When we bring people into a world where their existence leads to harm befalling them, we may regret allowing their existence in the first place because we have caused harm. Yet we cannot regret not bringing people into the world where their existence leads to pleasure for them simply for the reason that such people never existed in the timing sense, and as a matter of course we cannot regret that the disallowing of good will never occur.¹¹

A fourth supporting argument is somewhat abstract in how it relates the RAP to a type of geographical sense around assessing a life's worth. The argument that Benatar (2006:35) gives here deals with how one would surely despair at the thought that people in a some distant location may be suffering. Though, when we encounter a counter-example of an uninhabited geographical region we will logically not despair at the thought of non-existent people who could have existed there (assuming such existence would be happy, with good in their lives).

The preceding four supporting arguments of the RAP all function from a single basis - when given that existence is negated, it is logical to conclude that (potential) harm or pain is absent.

¹¹ Naturally there is also the idea that such regret cannot arise since such people (whose existence was disallowed) are indeterminate, but this approach follows from the second support argument and is distinct from this argument of the timing of regret.

On the other hand, when given that existence is negated, it is not forthcoming or immediately logical to conclude that (potential) benefit or pleasure is absent.

2.4 An attempt at the quantification of harm

As first point of anchor for his argument, Benatar (2006:60) indicates that even a single, minute instance of existential badness, harm, or pain, is sufficient to support the anti-natalist position. To drive the argument further, he provides extended reasoning regarding the amount of (the given) bad in human life. This argument, which in my estimation is an attempt at quantifying harm, is the second point of anchor for Benatar's anti-natalist approach. This aspect will be investigated in this section.

The quantification argument can also be seen as a further extension to the conclusion of Section 2.2 (that coming into existence is *always* a harm) to an even stricter form, i.e. that coming into existence is always a *considerable* harm. In reviewing this second argument I will firstly provide an overview of how an attempt at quantification of harm cannot merely consist of a simplistic summation of the positives (good, benefits, pleasure) and negatives (bad, harm, pain). Secondly, I will discuss Benatar's suggestion that our self-assessment of a life's worth is biased and often erratic.

Benatar (2006:61) refers to the diagrammatic illustration of the RAP (see Table 1 in Section 2.3) when explaining that a crude approach to assessing a life's worth would consist of merely summing the positives and negatives as represented by quadrants (1) and (2) of said diagram. In other words, Benatar is critical of the idea that adding the positives and subtracting the negatives to reach a final quantification regarding these factors allows for clarity in estimating a life's worth. There is an undesired crudeness in this evaluation, he suggests, because such an evaluation is based on a type of assessment of a life's worth as a 'once-off' within the temporality of the life as a whole. Once this evaluation has been conducted, Benatar argues, it is presumed that such an evaluation will not take place again (and again) throughout the course of one's life. He argues that this shortcoming of the assessment could be addressed by considering how pains and pleasures are distributed across

a life.¹² In order to facilitate such a ‘distributional’ analysis he proposes four metrics that I discuss in turn below.

2.4.1 Four metrics

The first metric considered is that of the *ordinal* metric of the distribution of good and bad. Benatar (2006:62) suggests several conceivable variations of the ordinal metric whilst juxtaposing their assessment of being either better or worse than the other. In one example, he proposes that one considers a life wherein the distribution of pleasures is concentrated at the start of the life and then followed by a solid succession of the pains later in life. Such a life would seem to be worse than a reversed version, he argues, which starts with a large degree of pains and ends with a large degree of pleasures, even if the actual net amount of good and bad remains the same. A life where the good progressively increases towards the end of said life is presumed to be better than one where the bad steadily increases (because one can contrast these states against each other). These examples thus demonstrate that regardless of assigning the same net values of benefit and harm, the *ordinal* aspect of their distribution will affect our valuation of the worth of a life.

The second metric concerns the *intensity* with which the pains and pleasures are distributed across a life. This metric can have multiple variations according to Benatar. Consider a life where the benefits which an individual experiences are highly concentrated in terms of intensity, but distributed very infrequently across the life as a whole. Such a life is deemed worse by Benatar in comparison to a life where the distribution of benefits is more modest (in magnitude), but at least more frequent. An interesting (secondary) variation he considers is that the value of the benefits can actually be so minute as to be insignificant; in which case

¹² Especially in his usage of the concept of distribution, I see a suitable analogy of this analysis with the same concept of data distribution in statistical sciences. Consider for example how an insurance company will seek to prepare payouts for expected but unknown loss events. The main input to their distributional analysis will begin by analyzing the frequency of the loss events, as well as their severity (magnitude). I thus suggest that the topics are treated analogously in Benatar.

the former would indeed be better (Benatar, 2006:62).¹³ Consideration should thus be given to the intensity of pains and pleasures.

The third metric refers to the *temporality* (my own terminology) of the distribution of pain and pleasure; in other words, over what time span pain and pleasure (of constant amounts between varying instances) are distributed. Firstly, consider two lives where the total amount of good and bad is set to be equal but the lives are of slightly different lengths. Benatar (2006:63) is of the opinion that the longer life would be better if that life possesses sufficient quality (pleasures or benefits) to continue. I am not fully convinced of the argumentation that this assessment holds. I contend that the distribution of the pains and pleasures would change since the distributional area (the lifespan) over which it is allocated is larger (a longer life being lived), but the overall totals of pains and pleasures remain the same. However, a variant example, which Benatar provides, is more compelling: namely the instance of an enduring life where the benefits are somewhat diminished, thus suggesting the presence of a greater amount of harm. In this example, the temporal distribution of pains and pleasure is more agreeable to the individual existing in said life.¹⁴

The final metric is that of the *thresholds* associated with a good life in contrast to a bad life. Benatar's argument suggests that, depending on both the amount of good and the amount of bad in a life and its placement within a life, there may be certain thresholds – a kind of 'point of no return' – beyond which the bad leads to an overwhelming negative assessment of the worth of a life. I argue that such a 'badness breakpoint' and its ultimate negative assessment may be due to an *intensity* factor (a horrifically bad event which cannot ever be recovered from) or an *ordinal* factor (a mildly bad event which occurred at a very young age

¹³ I wish to provide the reader with a concrete illustration of the intensity of pain sensation. Ariely (2008:xiv) relates his personal experience of medical treatment after suffering severe burns over much of his body. He explains that the nurses believed in a specific approach to removing bandages. Their approach was to complete the painful removal as quickly as possible which they believed resulted in less pain overall. Thus bandages were ripped off forcefully and quickly, one after the other. The nurses also held the conviction that it is irrelevant whether the removal starts or ends with the most painful of the wounded sections. A course on brain physiology later in his life led him to further experimentation around this scenario, which he personally experienced. His results concluded that the patients suffer much less when there is lower intensity pain over longer periods; thus removing bandages by slowly pulling on them over many hours is better.

¹⁴ Again, there is linkage with Schopenhauer here. As Jacquette (2000:45) explains, for Schopenhauer sustained suffering may reduce the individual's overall interest in life. Therefore, both the ordering and intensity of harm may pale in comparison to the temporality of it.

whereof recovery is possible in principle, but unlikely, given the immature age of the victim) (Benatar, 2006:63).^{15,16}

In the preceding paragraphs I described the four metrics that Benatar suggests according to which we can attempt to improve on the inadequate argument that pleasure and pain (adding pleasures and subtracting pains) should merely be quantified to obtain supposed final assessment of a life's worth. In my view, these metrics are a useful guideline to a more granular analysis of a life's worth. Benatar, however, subsequently continues to argue against said guidelines in an extended discussion explaining that our assessments of life's worth are in fact unreliable, and that such a quantified account of life's worth is in principle untenable (Benatar, 2006:64) due to nuances of our psychological makeup.

2.5 Psychological aspects of Benatar's argument

Benatar emphasizes the inefficacy of our (self-) assessments in terms of quantifying pleasure and pain. He argues that, due to humanity's somewhat oblivious approach to life, but especially the specific characteristics of the nature of the bad, the pains, the harm – such as its distribution and concentration – we constantly fail at evaluating the worth of life. Benatar suggests three reasons for such failure:

¹⁵ Benatar (2006:63) purports that the fourth metric is not a true 'distributional' metric, a point that I disagree with. I refer the reader to Footnote 12 on the linkage between his distributional analysis and that found in statistical sciences. Also in this instance, the consideration of the fourth metric (thresholds) fits perfectly since the analysis may focus on the distribution of the data after a certain point only, e.g. the aptly named technique of peak-over-thresholds analysis. The point itself, the threshold, can also be the focus of analysis with analogous statistical distributional elements such as 'Value-at-Risk'; a point at which the insurance company will not have sufficient funds to cover all possible payouts in a single loss event given a specific probability of such an occurrence.

¹⁶ Benatar (2006:63) provides the more detailed example of the victim of a gas explosion who suffered severe burns to his body but made a significant recovery. Yet this patient firmly held the belief of an overall negative assessment of his life; even upon this recovery and an apparently successful life. This was since he deemed the treatments necessary for his recovery to have been too excruciating. We should specifically note that his medical personnel had ignored his wishes not to be treated. This example actually highlights a possible shortcoming in Benatar's overall argument on the assessment of a life's worth: who is giving this assessment, and what right does this assessor have to give the assessment? Much of Benatar's views around this second anchor point of the quantification of harm, functions from the angle of *self*-assessment which one would expect to be best source of such an assessment. However, this angle may create problems in the decision-making with respect to potential / future lives where self-assessment is not possible. In Section 2.5 I discuss Benatar's views on how unreliable our self-assessments in fact are.

1. A tendency towards optimism.
2. Adaptation.
3. Comparison.

Our *perception* of the worth of life clouds a sober, distanced, objective assessment, and Benatar identifies the characteristics of human psychology as the culprit here; in other words, our minds are playing tricks on us. In the next section, I provide a summary of three psychological phenomena that cast a shadow on our assessment of a life's worth.

2.5.1 Pollyannism

The obliviousness in terms of humanity's estimation of life's worth is especially represented in the psychological concept of the Pollyanna Principle. The principle states that humans have a tendency towards a more optimistic perspective of life, resulting in human beings viewing both the past and future in a more positive light (Benatar, 2006:65). The psychological studies from which Benatar draws (Okuda and Okuda, 1999), indicate that more tangible or expected factors such as physical health, or social and financial standing, can in fact be insignificant indicators in the individual's perception of their well-being. Interestingly, more education leads to increased dissatisfaction – a greater sense (understanding) of the harms of life. This is perhaps the clearest psychological confirmation of all that ignorance is bliss. It thus follows that our tendency towards optimism clouds our capacity to evaluate clearly one's own life.

2.5.2 Habituation

The second psychological phenomenon to which Benatar refers is the idea that *aller Anfang ist schwer* – all beginnings are difficult. We often counter harmful or negative experiences with encouragements such as 'practice makes perfect' – a catchy summary of the phenomenon of habituation. As Benatar clarifies: humanity – as with many other forms of fauna – remains incredibly resilient. By "what might be called adaption, accommodation, or

habituation” humans push through the pains and reach the other side – often unknowing of what harms or hardships may still await them on the other side (Benatar, 2006:67). This dynamic adjustment allows for a varying perception of well-being where such perception is also relative to one’s encounter with such pains and hardships. Unsurprisingly, a constant and absolute assessment of life’s worth is meaningless, if at all possible, due to the phenomenon of adaptation. As Benatar points out, this phenomenon is further compounded by the Pollyanna Principle whereby the ever-varying perception is directed towards optimism.

2.5.3 Comparison

This third psychological phenomenon of constantly drawing comparisons with the perceived benefits and harms in others’ lives in relation to our own lives, unavoidably informs and ultimately clouds, our existential assessment of our lives. The tendency to compare makes it difficult to discern when a self-assessment is truly independent, and whether such an assessment is even possible. Benatar (2006:68) presents an interesting observation here: once we detect comparable harms, the effect of these harms in our self-assessments is somewhat diminished if not completely nullified. If someone has it worse off than us, we will necessarily evaluate our state as better off. In unison with the previously mentioned psychological phenomena, a consistent evaluation of our life’s worth seems to be a doomed prospect.

2.5.4 Summary

Benatar compares the three psychological phenomena by explaining habituation and comparison as having an inherent ‘negative’ tendency in our estimations, whilst Pollyannaism has an inherent ‘positive’ tendency. In his view, habituation and comparison would then often “operate ... from an optimistic baseline” (Benatar, 2006:68). He illustrates this effect by drawing on a psychological study stating that people are actually more likely to compare themselves to others whose lives they deem worse (those lives that are perceived to lead to an experience of more pain and harm) (Benatar, 2006:68). I question this conclusion by Benatar for the following reason: When considering my own life, I would be inclined to say that my comparisons are geared towards those whom I perceive as experiencing more

benefits and pleasures. Also from a socio-psychological point of view, I would imagine that comparisons such as the latter are the driving force for establishing ambition and goals and working towards them, which is a significant part of daily human endeavour. This one point of divergence allows one to question this particular estimation by Benatar, while providing that even this disagreement in part re-affirms the overall argument that our self-assessments are tainted by mental states and that we cannot accurately reach an independent, absolute self-assessment of the worth of a life.

In the closing remarks on these phenomena, I already see Benatar (2006:69) alluding to an underlying theme of the anti-natalist policy: that it should be a deliberate and conscious decision to follow such a policy even though certain organic facts contradict the claims of anti-natalism (such as natural selection which has as its core tenet the continuation and reproduction of life). Benatar's argument is suggestive of the fact that certain phenomena are likely evolution-driven. However, should our ancestors have realized and understood the badness of life in earlier epochs, the species may have ceased to exist much earlier.

Continuing on from this illustration of how thwarted our utmost attempts at assessing a life's worth are (I retain the underlying psychological thrust of these claims), I next continue to follow Benatar's zooming out from the details of an individual's mental life to more overarching theories relating to the quality assessment of a life's worth.

2.6 Three theories on the quality of human life

Benatar (2006:69) mentions that he follows Parfit's approach from *Reasons and Persons* (1984) in defining three theories on the quality of life. He then relates these theories to his argument that human beings are inept at assessing a life's worth. To reiterate, his goal is ultimately to show that regardless of which theory is applied in the self-assessment of life, the bad or harm or pain is inescapable – it cannot be theorized or imagined away.

In the current section, mental states remain a key theme in terms of the existential theories of life-worth evaluation being investigated. The three theories, through which I describe also Benatar's views, are based on established concepts where mental states form causal determinants. The first theory set – hedonistic theories – follows the principle that a life's

worth will be evaluated in terms of the either positive or negative mental states associated with that life in terms of pleasure. The second theory set – desire-fulfilment theories – relates to life’s assessment as good or bad depending on whether those desires have been fulfilled in that life.¹⁷ The third theory set – objective-list¹⁸ theories – centres on a principle of assessing the life’s worth with respect to the amount of *objective* good and bad contained in that life. In the following sections, I will discuss each theory in more detail.

2.6.1 Hedonistic theories

Hedonistic theories propose that three mental states may be distinguished, these being positive, negative or neutral (Benatar, 2006:70). The negative mental states encompass all conceivable emotionally bad states such as, for example, loneliness, boredom, anger, anxiety, stress and fear. Under positive mental states, a more granular approach is necessary, because there are positive mental states which are positive by definition, in their own right, but also reactive or resulting positive mental states which follow on the alleviation of negative mental states. The example of such a two-sided identification is demonstrated in eating: there may be relief from the negative state of hunger, but also pure enjoyment in eating a well-prepared meal with fine ingredients. Neutral mental states form the third component of these theories: there would be no positive or negative experience in relation to these mental states.

Recall the unreliability of our self-assessments as discussed in Section 2.5 and more specifically how we become cognitively numb to negative assessments due to the various psychological phenomena. It is from this standpoint that the realization of the generally bad nature of life seems compellingly damning, for when we consider the most minimally vexing negative mental states such as discomfort from temperature imbalances or thirst and hunger, we can further recognize that negative mental states fundamentally permeate our existence to a high degree. As clarified by Benatar (2006:72), Pollyannaism makes us disregard the negative; habituation allows us to get used to the hardships however insignificant they may

¹⁷ A specific mental state can also be a desire to be fulfilled.

¹⁸ I retain Benatar’s naming convention throughout of a hyphenated noun, i.e. ‘objective-list’.

be; and comparison creates the illusion that our negative states are actually (more) tolerable relative to the suffering of others.

The case for the near omnipresence of negative mental states is thus convincing in my view. Consider, for example, also the case of the reactive-relief positive mental states – e.g. managing to reach the bathroom in time! By definition, such states had to be preceded by negative ones, which further highlight the intrinsic harm in a life. If we then also add the dull possibilities of neutral mental states, positive mental states would appear to be few and ephemeral.

2.6.2 Desire-fulfilment theories

Desire-fulfilment theories are related to hedonistic theories in the sense that mental states may function as *possible* desires to be fulfilled. We would assume that the natural desire would be firstly, to experience positive mental states, and secondly, to be free of negative mental states. However, Benatar (2006:74) cautions that we should not overlook the distinction in this regard, since one can have a positive mental state when no desire is fulfilled and vice versa. This theory therefore may be quite a harsh criterion in the self-assessment of a life's worth: in the instance where the required desire is fulfilled, the assessment would be that of a good life (benefits or pleasures), regardless of whether a positive mental state is associated therewith. This may become problematic in cases where the desire was misconceived, or perhaps where the resulting mental state may not have been accurately predicted.

As with the ever-pervasive occurrence of negative mental states under hedonistic theories, the ever-pervasive occurrence of unfulfilled desires follows analogously. Desires cannot be satisfied immediately, or even quickly, most of the time. Such a situation creates periods marked by annoyance at the possibility of an unfulfilled desire, and even anxiety or fear. As with negative mental states, Benatar (2006:75) indicates that human beings will most likely spend an extended part of their lives waiting for the fulfilment of desires, and thus be dissatisfied for a large portion of their lives.

The dualistic nature of desires is also forthcoming when considering a longer-term view of how a life unfolds. Whereas desires are often only analysed in the active acquisitive form,

there is also the more passive form: to *not-lose* something. Once the desired positive mental state sets in, or the coveted object or person is confirmed, the desire of *non-loss* of said item is initiated. Although such desires may indeed remain fulfilled for some time, they are in turn associated with the fear or anxiety of suffering the loss of what has been fulfilled; again we see a life overflowing with harm (in this case, specifically the fear of loss that may become overwhelming).

Benatar (2006:76) points to Inglehart when concluding that we do not abandon life's goals if the possibility exists that satisfaction and desire-fulfilment are fleeting; i.e. some dissatisfaction may be good for you, at least in terms of your productivity and continued ambition. Benatar's argument is here linked to my reasoning regarding the psychological phenomenon of comparison (see Section 2.5.3), for when we compare ourselves with someone or something which we perceive as having more good in life, this aforementioned dissatisfaction could assist one in being driven forward to more (perceived) good in one's life (given that it is actually contained in that life onwards).

2.6.3 Objective-list theories

Objective-list theories exist within an even more fluid state than the interplay described between hedonistic and desire-fulfilment theories due to the nature of mental states. It is easy to imagine – when compiling a list of good, beneficial and pleasurable *objects* – that we can surely also include positive mental states or fulfilled desires. However, as with the intangibility of perceived good, the objective-list also comes to us in a deformed manner. Benatar (2006:81) emphasizes that the compilation of such a list of benefits cannot possibly be executed from a generalized or universal viewpoint (*sub specie aeternitatis*), but will always be subject to a human origin or anchor (*sub specie humanitatis*). Despite this apparent shortcoming, this theory does retain some generalization in that, whereas desires will be unique to an individual, an objective-list is meant to be applicable across all individuals, or at least to the general individual.

Nevertheless, the sub specie *humanitatis*¹⁹ aspect remains problematic in this theory. Benatar (2006:82) explicates that said aspect could only have efficacy if the assessment was anchored in the psychological phenomenon of comparison. It would then follow that a human perspective is even indispensable since human lives' worth are assessed relative to each other. Since the endeavour is actually the contrary – to move away from the unreliable comparative assessment – the sub specie *humanitatis* aspect is problematic because our absolute or base assessment is impossible from the outset. As it stands, there is no untainted, independent starting point. Benatar also qualifies this 'censored' starting point by stating that our ability to assess a life's worth is "very much determined by the limits of what we can expect" (Benatar, 2006:82)

The mere word 'expect' in Benatar's statement already foreshadows inaccuracy. As with the persistent problem of 'perception' and how it affects our assessment capability, 'expectation' creates clear limitations, in the sense that the assessment of a life's worth will have to fall within a narrower range, which may be due to our lack of imagination, our inexperience, or our preconditioning. In relation to this statement, one may argue that we do not weep at individuals who do not reach their 240th birthday – given that our experience dictates this to be an impossibility – and thus expectations can restrict our capacity for self-assessment (and resultantly, how we could compile an objective-list as guide for such self-assessment).

2.7 Chapter Conclusion

In this chapter, I have provided an overview of the main arguments that underlie Benatar's anti-natalist position. Moving from the general view that there is *ever* harm in life and that we should actively avoid it, we proceeded to a stricter argument that there is *always* harm in life. This conclusion was reached by investigating the asymmetry that exists between the absence and presence of pain and pleasure, and especially how we evaluate such instances

¹⁹ Benatar (2006:83) takes an interesting angle with regard to sub specie *humanitatis* and sub specie *aeternitatis*. Again, attempting to provide an assessment guide to evaluating life's worth, he explains that it would be much more meaningful if an objective-list could be compiled from the universal perspective instead of the purely human one. This in turn could then also imply that the life's worth would be assessed even higher if the good / benefits / pleasures in it are determined from a universal viewpoint, thus encompassing and eclipsing the human viewpoint.

as good or bad. The asymmetry showed that there are finer nuances in the evaluation, for while it is logical that the presence of pleasure is good whilst the presence of pain is bad, the specific characteristics of our humanness makes an evaluation in such a linear sense problematic. Following on from this logic it seems reasonable that the absence of pain is good. However, it is definitely not logical that the absence of pleasure is good. Furthermore, it is also not straightforward that the absence of pleasure is bad. Benatar argues that the absence of pleasure is 'not bad', and therein is anchored the Relational Asymmetric Principle which states that good and bad cannot merely be 'added up' and 'compared'.

Benatar continues on to suggest that there is *always* harm in life, and that said harm is indeed *very bad*. Much of this argument concerning how bad the 'bads' in life actually are, centres on humanity's inevitable anthropocentric approach to living. We cannot escape our very being, and at the same time, this being is linked to shortcomings in our assessments of our surroundings and experience of them (in terms of existential harm). Psychological phenomena are the greatest problem factor, significantly clouding our ability to present evaluations of absolute, generalized truth of a life's worth. Pertinently, such psychological factors hide from view how lacking human life actually is in relation to existential harm due to our tendency towards optimism, our adaptation to bad situations, and our tendency towards comparison with others.

The central tool that Benatar presents for assessing our existence in relation to harm, the RAP, will be utilized in my analysis of technological alternatives to existence which purport to reduce or avoid harm, as presented in Chapters 4 and 5. However, first, I will investigate Benatar's critics and the opposition they express for the tools utilized in this study in terms of assessing a life's worth and for allowing it to occur in the first place, while also investigating the RAP's robustness for evaluating a life's worth.

3 Chapter 3: A critical evaluation of Benatar's concept of harm

Benatar's influential ideas regarding anti-natalism have been engaged with from various fronts, as will be discussed in the current chapter. Benatar's *Better Never to Have Been*, which engages with the central questions of this study, was published in 2006 – since that date this work has received both criticism (the longer of which has been responded to by Benatar), as well as support. Benatar's replies to many of his critics are available in published form, and specifically to those critics reviewed in this chapter. In the current chapter I will attempt to provide my own counter-arguments against these critiques from a first principles-basis using Benatar's core text. I also wish to focus here mainly on the criticism against the RAP, which forms the cornerstone of this study (to be revisited in Section 5.1).

3.1 A voice of dissent: Harman's argument

Harman (2009:778) objects to anti-natalism, countering Benatar's normative claim that harm is caused when creating a new life. Harman argues that Benatar's argument is based solely on admitting that there will be negative experiences in the individual's existence – any good experiences (and thus benefits) that may occur in an existence are therefore not considered to weigh heavily in Benatar's evaluation. Harman suggests that such benefits deserve greater consideration.

In terms of her first objection, Harman (2009:780-782) is not convinced of the asymmetry in the RAP (see Section 2.3). Much of the argumentation by Harman here centres on whether the goodness or the badness in question is relative to the specific individual who was created, or is evaluated in an impersonal way: i.e. causing harm is bad in general, no matter the person or the situation. This is an important question to raise regarding the asymmetry aspect of the RAP, since it may direct us towards a possible point of weakness in Benatar's argument. The main idea behind Harman's remark centres on the case of a specific person – she states that harm was indeed caused by not bringing the (now non-existing) person into existence, since said person 'suffers' due to not experiencing the pleasures that were destined to occur in their life.

I argue that Benatar (2006) had already addressed this point in his argument. He indicates that there cannot be any introduction of harm through depriving said person of possible pleasures, since such an argument is presumed upon the idea that the person is existent (however, the person does not in fact exist (Section 2.2)).²⁰ Therefore, argues Benatar, any evaluation of goodness and badness can only be conducted in an impersonal way – a point that Benatar addresses through his reference to the specificity of the individual in evaluating the worth of life (Benatar, 2006:173). Since introducing harm has been avoided in the first place, the assessment of a life's worth has shifted towards *Good* in terms of Benatar's estimation. In my view, Harman misunderstands that the basis of Benatar's postulation takes its starting point from avoiding harm in the first place. The next step of concluding that denying pleasure is *Not Bad* may appear as an overextension of Benatar's quadrant system. In addressing Harman's criticism, we can at most conclude that for quadrant (4) (as per Table 1) the counter-assessments of *Good* and *Bad* are fully balanced and that we are therefore left without any possibility of evaluation, save for the guiding principle I propose in Section 5.1 which states that it is best not to introduce harm at all.

Furthermore, Benatar's argument for the distinction of assessing a life's worth in terms of the present-life and future-life sense serves to enlighten further this point (as introduced in the discussion in Section 2.2). This aspect of Benatar's argument does not appear to be addressed in Harman's objection to the asymmetry of the RAP, though as argued earlier, I find that the distinction (in terms of present-life and future-life) is a key feature of the RAP. Indeed, if a life is already underway, it is naturally easier to agree that denying pleasure is causing harm (in terms of evaluating a life's worth as a net balance of good and bad). However, if a life is not yet in existence this argument does not appear convincing, and we are forced to contend that assessing a life's worth can only happen in an impersonal way. Thus, not introducing harm, in general, is assessed as *Good*.

Harman's second objection relates to the psychological aspects of Benatar's argument (as discussed in Section 2.5). Per Benatar's argument, the psychological aspects of Pollyannism, habituation and comparison lead human individuals to assess their lives' worth as better than they actually are. Harman (2009:782-783), on the other hand, contends that there should be

²⁰ There is a comparable objection raised in Boonin (2012:12), which I investigate in the subsequent section.

provision for an optimistic view on situations that Benatar describes as harms (such as experiencing thirst or hunger). She does not consider such situations *Bad*, pointing to Benatar's explanation that the *Good* assessment of said mental states is anchored in relieving those states. In this way she is appealing to "the Millian notion of 'higher quality pleasures' where said pleasures supposedly impart such extensive goodness as to make the badness of the life pale in comparison" (Harman, 2009:783).

This objection to Benatar's argument may be countered by focusing on the actual negative states (see also Section 2.6) that Harman designates as *insufficiently negative* in a more encompassing manner. In this regard I turn to the point of alleviation which Benatar highlights – i.e. that it is the alleviation of the negative mental state that provide a form of pleasure and thus there exists no connection with the negative state which can be assessed as good. We may therefore ask, in terms of alleviation, why the body would seek balance by initiation of the alleviation? The very act of seeking relief points to the fact that something (the mental state, the human individual's neutrality) is out of balance, which in itself points to inherent harm because a neutral state cannot be enduring in this regard. Thus, the appeal to the negative mental state indirectly (or directly in the case of its alleviation) imposes harm to the individual nevertheless.

3.2 Alternative methods for assessing a life's worth

In the subsequent sections, I reflect on two other critics' ideas regarding Benatar's anti-natalism. Firstly, I consider the proposal put forward by Belshaw (2012) that the assessment of a life's worth should be undertaken as a type of instantaneous overarching assessment; all-in-the-moment. Secondly, I review the argument by Boonin (2012) that there may exist an alternative to quadrant (4) of Table 1. From the basis of this alternative suggestion, Boonin develops the approach he entitles the *Relational Symmetric Principle* that serves to challenge the RAP suggested by Benatar ultimately.

3.2.1 Belshaw and alternative methods for assessing life's worth

Belshaw (2012:126) appears to support Benatar's anti-natalist argument, for Belshaw concludes, as does Benatar, that existence in the future-life sense – to prevent births (see Section 2.2) – can be good as harm is avoided. Whilst acknowledging Benatar's notion that we normally set significantly higher thresholds when assessing the goodness of a life which is already underway, Belshaw postulates that the assessment of a life's worth should not be assessed as a time-independent whole (overarching), but at some specific time in the life (instantaneous) – my own formulation of this idea is an *all-in-the-moment* assessment of a life's worth. Benatar does show sensitivity for this netting approach²¹ in assessing a life's worth (Benatar, 2006:1), but as I will show in Section 5.1, the best result may be achieved when no harm is introduced at all, thus rendering the netting approach towards the assessment of life's worth unnecessary. The introduction of no harm is, naturally, only possible in a future-life sense, i.e. never to allow the existence to occur. Whereas in the present-life sense, appealing to the netting result is inevitable since the life is already underway (and pains and pleasures are already occurring). However, Belshaw appears to ignore the bulk of Benatar's reasoning regarding humans' overly optimistic assessment of a life's worth (in the context of psychological aspects of the assessment) in favour of asserting exactly such an optimistic perspective; he argues that human beings are living "in most cases, a worthwhile life, with a sense of, and desires for, a future" (Belshaw, 2012:125).

I am critical of Belshaw's argument for two reasons. Firstly, the aforementioned prospective of a worthy future does not seem to support the approach of netting pleasures and pains, for if the individual has expectations for a worthwhile future and these do not crystallize, there is the added badness of disappointment and unfulfilled expectations that adds a further negative element on the netting calculation. The forms of badness (disappointment and unfulfilled expectation) could not be determined or predicted beforehand through an overarching analysis. Secondly, the statement by Belshaw that humans will mostly live worthwhile lives provides no commentary on Benatar's warnings that ignorance is bliss

²¹ This terminology of 'netting approach' refers to a type of accounting approach to assessing a life's worth; thus summing the pleasures and subtracting the pains to obtain a net result. I related comparable instances in Chapter 2 when I was setting up Benatar's central argument.

(pollyannism), dynamic adjustment to harm (habituation), and the acceptance of the relativity of harm (comparison) as discussed in Section 2.5. Belshaw does not provide a convincing argument to counter the claims made by Benatar.

3.2.2 Boonin's view on alternative methods

Apart from the argument that Benatar amend his anti-natalism with an all-in-the-moment assessment in the present-life sense as presented by Belshaw, there is a further consideration for assessing a life's worth by means of the RAP provided by Boonin. Boonin suggests a reworking of the RAP, which removes its eponymous *asymmetry* and substitutes it with *symmetry* – a Relational Symmetric Principle (RSP). By proposing an alternative to quadrant (4) of Benatar's RAP, Boonin demonstrates how symmetry may indeed be possible when assessing a life's worth in terms of the presence of pleasure (quadrant (2)'s target) and the absence of pleasure (quadrant (4)'s target). A comparison of quadrant (4) of the RAP and RSP is provided below:

Table 2: Comparison of quadrant (4) of the RAP and the RSP

	Non-existence
Relational Asymmetric Principle (Benatar)	(4) Absence of pleasure (NOT BAD)
Relational Symmetric Principle (Boonin)	(iv) Absence of pleasure is (WORSE) than presence of pleasure if: a) there exists a person whose interests are better served by the presence of pleasure; OR b) the absence of pleasure requires the absence of the person whose interests are better served by the presence of pleasure.²²

Source: Benatar (2006:38), Boonin (2012:15); formatted by author.

²² This formulation is my own summary of Boonin (2012). The original wording is as follows: (4) the absence of pleasure is worse than the presence of pleasure if either (a) there is an actual person whose interests are better served by the presence of the pleasure or (b) the absence of the pleasure would require the absence of a person who would otherwise exist and whose potential interests are better served by the presence of the pleasure (Boonin, 2012:15).

In the comparison of Benatar's RAP and Boonin's RSP in Table 2, I designate Boonin's alternative to quadrant (4) as quadrant (iv). Boonin (2012:15) argues that if we take Benatar's quadrant (3) to hold true – that the presence of pleasure is *Good* – then his suggestion for completing the Relational Principle in a symmetric version is indeed possible using the formulation provided above (as quadrant (iv)). However, if one disregards the other quadrants, Boonin's quadrant (iv) does indeed appear to be logically coherent, relative to itself at least. Herewith, I mean that the main statement holds true for both case a) there exists a person whose interests are better served by the presence of pleasure, and b) the absence of pleasure requires the absence of the person whose interests are better served by the presence of pleasure.

However, as was the case with Harman (2009:782), Boonin once more presents an appeal to the argument that there *could* exist a person who supposedly would be deprived (and thus harmed) due to the absence of pleasures. This appeal is seen in both case a) and b) of quadrant (iv), where the following statement is made: "...a person whose interests...". This statement clearly requires a reference to a *specific* individual who will possess particular interests to be served; there is no *overall* benefit diminished in this appeal, only the supposed benefit of that specific non-existing person. Benatar outright rejects this approach, as I have mentioned before, with regards to the seemingly 'non-exist' person (see Section 2.2). As I have also pointed out in the preceding analysis on Harman, it is clear that Benatar's argumentation already allows for an impersonal assessment of a life's worth.²³

An interesting additional nuance in Boonin's quadrant (iv) is the requirement that the person be absent for the absence of pleasure to occur. This necessarily absent person having the causality with respect to the absent pleasure merely reinforces that Boonin's approach to

²³ Benatar does indeed seem to focus on the theme of impersonal assessment very acutely in his reply to Boonin (Benatar, 2012a:137). In his response, he even formulates a first and second modification of Boonin's Actual Persons Principles (respectively the *Reformulated Actual Persons Principle*, and the *Alternative Actual Persons Principle*). I do not review these in detail, but provide the reader with Benatar's ultimate conclusion on this approach: "Given the amount of harm that can be expected to characterize human life, a principle prioritizing harm avoidance would lead to anti-natalism, albeit via a different route."

assessment of a life's worth is personalized and not impersonal.²⁴ As I have shown in my summary of the RAP in Section 2.3, when closely following Benatar's developing argument, especially in the future-life sense, he clearly suggests that even a single unit of pain is one unit too many. Therefore, quantification of harms and the subsequent personification of said harms are unnecessary. One may perhaps conclude, in this regard, that we need not bother with compiling quadrant (4) at all – perhaps we are already doing enough by ensuring the absence of pain in quadrant (3).

3.3 Chapter Conclusion

Perhaps the most compelling feature to note in the criticism leveled against Benatar's anti-natalist argument is that most of his critics' acknowledge to some degree that he puts forward a strong case, despite its difficult nature. The key theme in terms of the critical shortcomings on Benatar is, in my view, related to these critics' demands not to assess a life's worth in an impersonal way. Such an approach serves as a disavowal of Benatar's core argument where he clearly makes allowance for such an impersonal assessment; that is, he argues that we should not measure a life's worth in terms of some specific person who would be deprived of potential pleasures.

In reviewing criticism of Benatar's argument concerning anti-natalism it is apparent that the argument for reducing harm through anti-natalism holds true (at least, up to now) in terms of most instances of assessing a life's worth. The RAP provides this study with a useful tool for investigating alternative forms of existence, and by extension, for investigating the harms that may be inherent to or introduced by those existences. The next chapter investigates how technology-driven potentialities are suggested as alternatives to our current organic human existence with the goal of reducing harm. Said alternatives will be analyzed by means of the RAP in Chapter 5.

²⁴ This may also be seen in Boonin's examples of the 'Blessed Child' and 'Cursed Child' and their differing units of pain and pleasure.

4 Chapter 4: Technology-driven alternatives to existence

Modern technology appears suggestive of both a technologized present and a hypothetical future that is unrestricted by many of the demands of our organic bodies – a future anchored in a type of alternative existence that negates harm. However, I suggest that we need to question whether the technologies of today and of the speculative future may indeed be able to address harm. Anti-natalism contends that existence is not worthwhile, but would the same hold true for forms of existence based on alternative, technology-driven, forms of existence that purport to overcome or negate harm?

In this chapter I reflect on what an existence based on technological alternatives may look like in relation to Benatar's concept of harm. I postulate three technological alternatives to everyday existence, which I envision as a natural stepwise evolution building upon each other, starting off with the most recognizable alternatives in our current phase of technological development (algorithmic control) to the more esoteric and futuristic (hive minds and robotics).

In beginning this chapter with algorithmic control, I envision that the human organism would still be physically intact, but that a significant degree of engagement with technological mediation of existence would be taking place. Thereafter, I investigate how the human individual becomes more integrated with technology and may sacrifice some of its physical (and possibly mental) integrity for said machine-integration in the form of hive minds and broadly interpreted human-cyborg relations. The final envisioned step in this framework of technological development is the point at which the human individual no longer has a biological form and moves past any remaining organic component – the human being's body may become artificial, robotic, or the mind may be shifted 'online' to fully utilize technology for its own intentional purposes.

Whilst describing these three alternatives I reflect on specific examples of their application, either as already apparent in our current world, or idealized in art (particularly science fiction). Art often forms a medium through which we can represent our wildest fantasies, our boundless dreams for alternate realities, or for potential new realities. It is therefore in the genre of science fiction (sci-fi) – whether film or literature – where I also find my inspiration

for a technology-driven existence, and a projection of future human (or perhaps only human-like) incarnations.²⁵

As part of this investigation, I also consider how such postulated alternatives appear to have as goal the elimination, or at least reduction, of harm as described by Benatar in Chapter 2. I briefly show how, in each technology-driven alternative to everyday existence, there may still be shortcomings in attempting to overcome harm, before developing these shortcomings in greater detail in Chapter 5.

Prior to deeper investigation on technological alternatives to existence, I first describe key definitions here. These definitions originate with various sources in order to reflect the fluctuation nature of technology-related concepts. Furthermore, in the rest of Chapters 4 and 5, I build on the definitions with further analysis or examples. However, the definitions provided here, are to be considered as the core definitions from which my study proceeds.

• Technology	Even though Tiles and Oberdiek (2014:249) give attention to the distinction between views of technology arising from a viewpoint of either optimism or pessimism, they indicate that a uniting feature is that technology is based in the realms of facts, and is derived as ‘an extension of human beings or as an extension of nature’ (Tiles & Oberdiek, 2014:249). Furthermore, technology results from a rational application of knowledge (from the realm of facts), which in turn results from developments made in science. Drawing on Ellul (1983:86), Tiles and Oberdiek (2014:254) point to key attributes which technology should exhibit: 1. It is artificial;
--	--

²⁵ Although most of my inspiration is drawn from art forms around sci-fi, e.g. such as literature or visual arts, I wish to point out the existence of historical views of such concepts. The historian of ancient science, Adrienne Mayor, describes ancient Greeks already envisioned many of these technological manifestations, especially with biological links (Long Now Foundation, 2020). Some examples mentioned by Mayor include: i) automation of gates to allow greater ease of entry and departure for the gods of old; ii) automated delivery carts which would provide food and drink to the gods and depart a party once emptied only to return refilled again. Important in this context is that these historic views were not connected to some magical workings of divine commandments; the historical texts clearly reflect a process of manufacturing of these amazing products of technology. Even the ominous character of Pandora is described as “made not born”, in Mayor’s own words, Pandora was a type of “fembot”. Mayor also provides an extended background on how these ancient ideas link to modern-day incarnations. In the case of the struggles for the hero Jason, his sorceress wife, Medeia, would be interpreted today as a “hacker”, she could figure out all of the shortcomings of the technological beings which Jason had to fight (Long Now Foundation, 2020).

	2. It is autonomous; 3. It is self-determining; 4. It grows according to a causal process; 5. It is formed by an accumulation of means; and 6. Its parts are inseparable implicated.
• Artificial intelligence (AI)	Nilsson (1998) provides a compelling definition with sufficient breadth, but a narrow enough view to form a starting point for my study. According to Nilsson (1998:1): <i>“Artificial intelligence, broadly (and somewhat circularly) defined, is concerned with intelligent behaviour in artefacts. Intelligent behaviour, in turn, involves perception, reasoning, learning, communicating, and acting in complex environments. AI has as one its long-term goals the development of machines that can do these things as well as humans can, or possibly even better.”</i>
• Algorithm	Blass and Gurevich (2003) focus their study on pinning down an explicit definition of algorithm, specifically also reflecting on the Church-Turing thesis of perceived computational success. However, their interpretation of algorithms from the work of Andrei Kolmogorov is the definition I am using in my study. It is summarized as (Blass & Gurevich, 2003:204): 1. An algorithm consists of a series of pre-defined steps; 2. The pre-defined setup is independent of inputs as well as the status quo of computation; 3. For each step, there is a direct and immediate transformation of the status quo; 4. This process of transformation is active in isolation and therefore does not impact the remainder of the status quo; 5. The series of steps execute until one step becomes impossible or a flag is triggered that serves as confirmation of an output having been attained.
• Machine learning (ML)	As Zednik (2018) explains that Machine learning (ML) indicated a progression point after algorithms. In this context, the machine, which will perform the algorithm, requires limited input in the form of algorithm. In the case of unsupervised ML as discussed below, the machine is now able to compile the algorithm itself (Zednik, 2018:225). It is also important to note the distinct of <i>supervised</i> and <i>unsupervised</i> ML as defined in Danks (2014:154): • For supervised ML, the target variable is known; i.e. the object for which prediction will be attempted by the machine’s learning is defined beforehand. The algorithm will search for explanatory patterns to define the target variable.

	• For unsupervised ML, the target variable is not known, or rather, not yet isolated / selected. In this case, the algorithm has to inspect also possible explanatory relationships between data points and thus compile a meaningful target and its explanatory factors.
--	---

4.1 The governance of human life by algorithms

Just and Latzer describe algorithms as “problem-solving mechanisms” (Just and Latzer, 2017:239) that are utilized in a variety of Internet-based services such as Google, Spotify, and Netflix (Latzer et al., 2014). Furthermore, they delineate algorithmic selection as “the automated assignment of relevance to certain selected pieces of information” (Just and Latzer, 2017:239). In the contemporary age we have entered the era of algorithmic governance of information societies and algorithmic reality construction, whereby said algorithms play increasing roles in terms of politics and epistemology in our contemporary societies (Just and Latzer, 2017:253). Importantly, in this discussion of an algorithm-based alternative to existence, I still view human life as essentially organic with technology only supporting or facilitating our biological activities in specific ways.

4.1.1 What are algorithms?

Algorithms entail a basic mathematical-logical formulation, which is then expressed as a mechanical, electrical or computerized rule where the result or output of executing the algorithm is of benefit to, or desired by, the human recipient – normally also the initiator of the algorithm’s execution. They are an intrinsic part of engagement with others and with the modern world in our contemporary age – I am referring, in this section, not only to how algorithms provide services to humanity, but also how such algorithmic governance serves to decentre aspects of our daily lives. Algorithms present an alternative to our existence in the sense that, technologically, our existence is directed by algorithms as a means to solve problems (as is the case with older forms of technology) or as algorithms giving us what (we think) we want through mathematical-logical formulations (as we find in digital algorithms). The aspect of algorithms that feature in this section, follow from their application as defined at the start of Chapter 4.

My definition of alternative existences based on algorithmic ways of being thus refers to cases where the sequential formulations, or mathematical rules, of the algorithm may be composed by a human, and whereby technologies – or tools – execute these steps fairly independently subsequent to initiation. In this regard, it is important to note that human civilization has moved far beyond relying on tool use to achieve pragmatic ends; in contemporary times, our tools have become automated and we express our desires under the auspices of algorithmic programming. Thus, when we consider technology in its most ancient and primitive forms, i.e. humans using tools to achieve some material goal, humans were always the central decision-makers. Consider, for example, the act of starting a fire or hunting prey; in such instances, the human is using tools as technology to achieve the specific results. However, the algorithms, or in this case the logical sequential steps of applying the tools and obtaining the final result, are executed within the human's mind, or in skilled instances, the human's muscle memory.

Now consider how we order a snack from a vending machine, which is based on the convenience of the retrieval of nutritious sustenance. We see how mechanical algorithms set the tone for vending machine-loving Japan to obtain a (hot!) nutritious meal (Japan Lover, 2014). The video describes how in the 1970s Japanese travellers were able to buy a comforting warm meal when hunger pains struck from a coin-based vending machine; from hamburgers to noodles with broth dishes. A more advanced example would be Philips Hue smart-home lighting systems: the human user still chooses which lamps must be activated and at what luminosity, whilst the lighting system itself only actually activates the electricity flow and dimmer devices to achieve the human actor's desired lighting.

Technological algorithms, on the other hand, bypass this process of engagement on the part of the user. Rather, the algorithm determines automatically what we desire and works to give us what we want. This alternative to existence is, as mentioned, a reality for us already – algorithms guide our Google searches and our ways of engaging with other individuals via social networking. An underlying argument is that the human should approve, or allow, this determined guidance as enacted by the algorithm.²⁶ However, as confirmed in the high profile

²⁶ As Zuboff (2015) explains, there is much motivation for the human agent in approving this guidance. It is described how the human agent allows this determined guidance ('surveillance capitalism' to use her terminology), since there is some end goal of acquisition, whether in the literal or figurative sense. It is

against of the EU against Google, search algorithms restrict the human agent's view of options to approve (Breinstrup & Dandanell, 2019). The case centred on how Google search results for a product can be filtered to move competitors' search hits to a lower level, and thus become more unlikely to be selected by the user.²⁷

4.1.2 Algorithms and existential harm

Algorithms, as enacted through tools, are essentially designed to make life easier, less tedious, and more pleasurable for the human actor. Some of the examples mentioned in the previous section point to obvious benefits: the vending machine provides nutrition quickly and at convenient locations around the clock for the human recipient. Vending machines free up time in that we need not exert ourselves in seeking other food sources, or even in preparing the meal itself. The smart lighting system, on the other hand, may detect the human actor coming home at some time after sunset and automatically provides convenient lighting which may aid in the individual avoiding injury in the process of fumbling for a light switch in a dark stairwell.

A crucial factor to note here is that the harm that is purportedly overcome is mostly centred on the human actor's biological desires or needs. The convenient vending machine is ultimately stilling potential hunger, whilst a smart lighting system is ultimately preventing bodily harm for the human actor and thus ensuring that individual's physical integrity. In terms of algorithms, we find that the mechanism may be different, but that the same forms of harm-reduction are assumed through their technological function. Initially it may seem that harms, in the sense of Benatar's views, may indeed be negated or even overcome. Relating specifically to Benatar's theories on how hopeless human beings are at assessing a life's worth (see Sections 2.6.1, 2.6.2, 2.6.3); such technologies indeed seem to overcome aspects of

mentioned how the human agent could be occupied with obtaining a home loan, or supposedly specialist online advice on legal or medical matters (Zuboff, 2015:82).

²⁷ This leads to the phenomenon of 'filter bubbles' as defined by Pariser (Flaxman *et al.*, 2016:299). In their study, Flaxman *et al.* (2016:311) however, argue that the segregation of results (through filtering) may be due to the 'web-search personalization' as identified by Pariser, but that their remains an element of human agent preference being added to the segregation of results. Given that the human agent's view of options is already restricted to external factors (e.g. the Google competitor filtering), the effect of segregation is compounded even further.

biological harm (at least upon initial consideration). However, I will show that in this purported overcoming of harm we find that other aspects of harm do indeed arise, which suggest that harm cannot be completely negated in this context. These instances of harm will next be evaluated by means of the hedonistic, desire-fulfilment, and objective-list theories (see Sections 2.6.1, 2.6.2, 2.6.3) that Benatar introduces to show how such theories cannot account for harm itself.

4.1.2.1 Hedonistic theories

With respect to hedonistic theories (Section 2.6.1), algorithms would be fulfilling the function of increasing pleasure as part of their basic aim. If one considers some of the negative mental states mentioned in Benatar (2006:71), such as thirst or hunger, a vending machine would serve to relieve the negative mental state of hunger by providing a nutritional item. Air-conditioners and central heating relieve our negative mental state of thermal discomfort. Another type of vending machine, which may also increase the positive mental state of amusement delivered by means of a packaged toy, not to mention how sex toys augment the positive mental state of physical pleasure, serves to underline the shortcomings in this regard. In each of these cases, it appears that technology addresses a specific negative mental state with the goal of negating harm. However, as Benatar shows, such hedonistic theories are unable to negate harm in its fullest sense fully.

4.1.2.2 Desire-fulfilment theories

Overcoming the problems associated with desire-fulfilment theories seemingly present an unscalable challenge for the algorithm-based device. Consider Benatar's description of how hedonistic and desire-fulfilment theories may overlap – in his own words “one cannot be mistaken about whether one is, right now, experiencing a positive or a negative mental state” (Benatar, 2006:74), and “one can be mistaken right now about whether one's desires have been fulfilled” (Benatar, 2006:74). A problem presented by algorithms is that they purport to optimize desire-fulfilment due to such algorithms' ability to make use of different inputs and different scenarios to reach the most optimal outcome for the benefit of the human actor. One's hedonistic tendencies, for example, may be satisfied (which leads to physical pleasure, as mentioned in the previous paragraph) while one's desires may remain unaddressed.

Consider a scene from the film *Minority Report* (2002) where a customer of a virtual reality games lounge requests a simulated scenario that would allow him to kill his boss. Given that different scenarios are possible and that the algorithm by definition can be repeated endlessly, we may note that even if said customer may find initial enjoyment in the act of killing their boss in virtual reality (hedonistic tendencies are satisfied), said customer may later realize that he could have derived more pleasure from another killing method, or a prolonged killing (he becomes aware that his desire is actually unfulfilled even though he experienced a momentary positive mental state). Therefore, an element of harm may still remain because algorithms, while functioning optimally to satiate an individual's hedonistic tendencies, may not be able to predict the nuances of the human actor's desires and may only serve to address such desires on a superficial level.

4.1.2.3 *Objective-list theories*

The third set of theories described by Benatar, namely objective-list theories, also link closely to the varying inputs and scenarios described in the previous section (in terms of algorithm-driven technology). Even when the perceived desires or hedonistic drives do in fact originate from societal conventions and peer group standards (Benatar, 2006:82-83), the adaptability of algorithmic responses may make allowance for the human actor to strive for the attainment of a positive mental state (Rice, 2013). Benatar's discussion focuses on how objective-list theory presents as a type of generalization of hedonism or desire (as discussed in Section 2.6.1 and 2.6.2) in that the pleasures or good are something that all individuals have in common. In line with this description of objective-list theories, I next turn to a second source for an example of a concrete objective-list.

Rice (2013), in presenting such an objective-list theory, zooms in on one specific pleasure which individuals in general perceive as an indispensable factor – that of having loving relationships. He points out that in numerous cultures (Ancient Greek, Chinese and American Indian societies) there is a core theme of friendship, compassion, or “generosity toward family and tribe members” (Rice, 2013:210).²⁸ I argue that algorithms do provide a means for the individual to build relationships with others in society, for example through sharing mutual

²⁸ The respective concepts are represented by *φιλία* (philia), *仁* (rén), and *wancantognaka*.

interests on a Reddit group, or making new non-romantic acquaintances through Tinder – and hopefully friends ultimately. Such chat apps, based on algorithmic matching, provide faster, better-matched opportunities for connecting with others. Firstly, these apps' algorithms provide one with an indication of which individuals are physically close to oneself. Secondly, there might be options for matching similar hobbies, or through the dubious yardstick of astrological compatibility. Friendship may flourish by making use of algorithms.

4.1.3 The shortcomings of algorithmic technologies for overcoming harm

However, inasmuch as algorithmic technologies can serve to address negative mental states, or purport to overcome harm by its changeability in terms of fulfilling certain desires, or provide avenues for connecting with others, there also exists a sense that such technologies may introduce a range of potential harms into the social landscape. Algorithms may structure the kinds of information we are exposed to (leading to a form of censorship, or even self-censorship), and may lead to societal echo chambers.

In this regard Gilani *et al.* describe the phenomenon of homophily, “our tendency to surround ourselves with others who share our perspectives and opinions about the world” as both a natural tendency and as part of the algorithmic structuring of our social networking technologies (2018). Echo chambers thus refer to spaces where individuals with comparable political views conglomerate online to discuss particular issues with the effect of mutually reinforcing opinions (even shifting opinions towards more extreme stances), greatly affecting political and ideological polarization (Sunstein, 2002).

In this manner algorithms may lead to the introduction of new harms into a society (they thus have a social and political aspect), whereby increased political and ideological polarization may lead to increased societal instability. An example here would be the significant amount of mis- / disinformation which has spread by technological means on the Covid-19 pandemic (Jamieson, 2021). Furthermore, as will be explored in the next sections and in detail in the next chapter, such algorithmic technologies cannot address fundamental aspects of our existential constitution – i.e. our bodily manifestation – in the world and therefore cannot overcome harms associated with such existence.

4.2 Hive minds and cyborgs

In the current section, hive minds and cyborgs will be brought into discussion with the question of harm. I will, in this section, shift away from the idea of the human as a being with in-born biological integrity (a perspective maintained in the previous section) and instead, in this second stage of human-technology evolution, describe how the technologies we use may lead to a situation where we can overcome biological limitations and struggles. I therefore suggest a conceptual movement on to a second, higher and hypothetical stage of humans' interaction with technology (thus, from the algorithmic characteristics derived from smart – but physically distinct – tools to the next developmental step which suggests the biological integration of the human actor with smart tools).

Verbeek (2005) describes how our daily life has become permeated with “tools, appliances, and other assorted objects” on which we are highly dependent (Verbeek, 2005:1); even to the extent that human individuals' interaction with each other, as well as their physical well-being, is subject to manifestations of technology. I therefore wonder how far a process of integration (or even assimilation) with external enhancing elements has already taken place. As Verbeek explains, significant parts of our lives make use of technology and instrumentation – consider human travel (from mechanical bicycles to the *hyperloops* currently being theorized), communication (telecommunication for faster information over greater distances), and survival (hospital equipment). One may therefore ask whether we may truly parse questions of humanity's usage of tools in our daily existence in their entirety, i.e. in the context of increasing technological capabilities, have tools now become so vital and integrated that they in essence take on an existence of their own?

Verbeek (2005:2) provides a confirmatory view on this, stating that society has in fact dematerialized the material. He refers to marketing strategies of luxury products, whereby the product itself may be a physical item that is exchanged for money, but that the buyer is actually buying the dream state to which marketing of the luxury products refers. Consider for example purchasing a luxury smartphone and the resulting ambience it creates for the buyer that the user will enjoy numerous splendid evenings of socializing with the 'right' kind of people in glamorous venues. The reality may actually be that the smartphone is the only

item of worth that the buyer possesses, and that it may be used to document sporadic social events with ‘average’ people in boring settings.

However, a new question arises in this context: is it fair to consider a luxury smartphone and a cyborg body part as equals in the material world? The end goals of these *artefacts* – to use the Verbeek terminology – are not the same in that the perceived net benefit that cyborg body parts would bring may outweigh that of the smartphone. In this comparison it should be noted that the technological artefact of the smartphone entails interaction, i.e. the dual entities of smartphone and user may be nearly constantly interacting, but they are still separable (for a while at least!). On the other hand, a cyborg body part presents a much closer integration where technology has become inalienable and where removal comes with significant intervention. There is thus a clear progression from interaction (as discussed in the previous section) to integration. In the current account, the biological manifestation of humans in the form of a pure carbon-based cellular body is abandoned for an alignment with not only algorithms, but also with technology, that changes the biological and embodied horizons of the individual. The most extreme example of such integration relates to the hypothesized concept of hive minds.

4.2.1 What are hive minds and cyborgs?²⁹

Hive minds refer to “an awareness formed from the communication of individual minds but different from each of its individual minds, and so not defined by the separateness of the individual minds which compose it” (Sirois, 2020:14). Hive minds are the stuff of sci-fi legend. A classic example, where we can clearly see the physical changes that a hyper-technological existence might entail for humanity, is that of the fictional Borg race in the *Star Trek* television and film franchise. In the fictional universe of *Star Trek*, the Borg are a humanoid species that consists of a combination of many individuals from various humanoid species which they have ‘assimilated’ – in other words, where free individuals are taken into a cybernetic hive much like that of bees or ants (Okuda and Okuda, 1999:51). The process of assimilation is key to a

²⁹ The items of a hive mind and cyborg are treated together here two reasons: the brevity which the nature of this study demands, but more compellingly: the fact that the two would be linked in a gradual fashion. That is, the gradual replacement of body parts (cyborg) could lead to a full replacement of the body with uploaded / online minds, such as hive minds.

biological individual becoming Borg (in fact, the Borg do not reproduce sexually but rather through taking – assimilating – pre-existing biological beings into their technological collective). The only prerequisite for Borg assimilation is for species to possess some distinctive (biological or technological) feature that will enhance the status quo of the Borg's development³⁰ – seemingly one of the least discriminating criteria encountered in modern societies. Such a unique feature is then integrated and rolled out across the Borg collective – effectively an 'upgrade' to their society.

In the *Star Trek* media franchise it follows that the Borg philosophy is based on the idea that, through belonging to the Borg (thus, by means of this assimilation) the individual's quality of life will be highly improved. The assimilation process entails that an individual will undergo cybernetic implants that enhance their physical abilities (through prostheses that include bionic limbs with increased strength, or a bionic eye that allows visualization beyond the capabilities of a human eye such as infrared or ultraviolet vision). A further part of the assimilation process consists of a type of signalling implant in the brain that allows the individual to connect to the Borg network – the *hive mind*. The latter allows instant central communication throughout the *Collective* – the entire population of Borg individuals (as far as one may refer to them as individuals) that is spread throughout the universe.

There are close relations between the integration of an individual into a broader framework of others by technological means, and the idea of the cyborg (also a popular trope in sci-fi as early as the 1927 film, *Metropolis*).³¹ In this section, I treat the hive mind as a separate though related concept to the idea of the cyborg, since both of these accounts refer to some form of human-technology relation wherein the biological characteristics of the individual are altered. Through our entertainment, media the possibility of existence expressed through such integration with technology has become more prevalent, and perhaps even been made more acceptable.

³⁰ This aspect is an obscure element of Borg mythology, mentioned in passing in a *Star Trek: Voyager* episode, *Mortal Coil* (Kroeker, 1997). A summary of the episode is available at: https://www.imdb.com/title/tt0708932/quotes/?tab=qt&ref=tt_trv_qu.

³¹ There are some parallels, though arguably more differences, in the sci-fi concept of the cyborg and the posthuman idea of the cyborg for Haraway (2016).

4.2.2 Hive minds, cyborgs and existential harm

Considering this process of integration into the Borg Collective, there appear to be a few obvious advantages for the individual. Borg implants may serve to alleviate certain forms of disability, whether through birth or accident – e.g. a missing limb may be replaced with a prosthesis that works just as well as the lost limb, or may provide certain advantages, i.e. a mechanical arm may function even better than the missing limb. There is an interesting line of comparison here with the classic example in Benatar (2006) of an individual who is missing a limb – whether it be an adult who lost the limb through accident or a foetus that had developed without the expected full set of limbs. As we noted in Section 2.2, Benatar (2006:23) explains that the individual observing the situation from the outside will judge the situations of both the adult and the foetus as bad and certainly harmful. Although the badness is usually not judged as equally by the external observer (as discussed in Section 2.2), I can envision that becoming Borg would alleviate many of the potential harms associated with such individuals through the possibility of inorganic corrective and enhancing components.

A second and somewhat less obvious advantage is the significant equalization of individuals that assimilation into the Borg brings with it. In the hive mind there are no racial, class or gender differences beyond the merely biological features of the assimilated individual; rather, everyone in the hive is judged to be equal components in a broader system. One may observe harm reduction in the Borg version of an alternative technological existence as it manifests through the assimilation of *any* receptive individual in that this presents an avenue towards the equality of which humanity has long dreamt. When the Borg assimilates an individual it sees no skin colour, no religious differences, no sexual orientations, and no social standing – everyone is welcome and treated as an equal here.³² Let us take the logic of this technological development further in terms of the question of harm.

An important aspect of overcoming harm in this context relates to the question of gender. The Borg would probably not be able to comprehend the concept of gender, and thus would

³² There is perhaps a slight exception here: that of the individual(s) acting as the Borg queen. A humanoid species which the Borg assimilated early on during their expansion possesses heightened mental / telepathic abilities. The Borg thus prefer these individuals to act as a type of signal booster for their collective's communication; the various queens dispersing central messaging across the hive (Okuda and Okuda, 1999:50).

be unable to parse the associated problems of discrimination in our contemporary societies. In terms of our contemporary technology, we find developments of medicinal and clinical care to allow the transgender individual to take on a male gender, for example, whilst retaining the original female reproductive capabilities. However, such individuals still encounter transphobia and other forms of discrimination in some quarters of contemporary society – even in societies that pride themselves on the recognition of diversity. As such forms of disregard and discrimination are culturally derived; it seems strange that such forms of transphobia exist even in so-called ‘developed’ societies because such individuals represent culturally unique and diverse contributors to said society. Looking towards Borg society, however, one finds that the aspect of distinctness that the transgender individual is representative of, will be added to the entire colony of individuals that form the Borg hive mind – there would be equality and the Borg would become a transgender species as a result. There would be no transphobic violence, no discrimination, but thereby the original individuals would lose their distinctiveness.³³

The point remains, therefore, that diversity and uniqueness would be lost in such a context. Some may argue that such a form of equality, while overcoming some aspects of harm, presents a nightmarish view of a technological dystopia – and that overcoming harm in this manner is stretching too far. Sirois, for example, argues for the undesirability of such a hive mind existence on the basis that free will may be lost; an undesirable outcome (Sirois, 2020:14).

On the other hand, some may argue that the benefit of belonging to the Borg Collective, in terms of their focus on bringing harmony and coordination to large groups of people, is a worthy goal.³⁴ In Okuda and Okuda (1999:51-52) we learn that in instances where Borg communities became isolated from the central Collective, their existence fell into chaos. Once

³³ I find this to be a modern-day reinterpretation of ‘two-spirit’ people as identified among indigenous North American Indian people, where such individuals are honoured for serving dual roles in society. Closely linked to this idea I specifically want to mention the ever-extending research into chromosomal ‘deficiencies’. Through technology we were able to apply a scientific basis to the gender spectrum. Although there is still categorization at play and no true continuum as of yet, technological advances showed us that individuals can be more than the usual binary XX-females and XY-males: we observe classifications now of individuals such as XXX-females, XYY-males, XXYY-males, and even XX-males.

³⁴ Although this topic is conceptually already pointing to the next alternative to existence (artificial bodies and robots), I wish to mention here this facet of hive minds in terms of purely human-cyborg considerations.

those technological attributes which allowed communication flow and coordination from the central system led by their queens became severed, their increasing individuality led to uncoordinated tribal-like warfare – imagine the dead Wi-Fi corners in our own homes which leave us isolated from internet services; reason enough for carnage! The advantages of coordinated, Collective-type existences may already be envisioned.

With the global effects of the COVID-19 pandemic raging deep into the year 2020 (and onwards) our sense of communal connection is especially challenged. The virus, and particularly the response by global governments, may be seen as a direct challenge to humanity's need for physical closeness. It seems obvious that hive minds may address the issue of individuals' negative experience of their pandemic-induced isolation, or serve to overcome the 'new normal' of *social distancing* once individuals are indeed able to congregate again to some extent. Cyborg implants may connect us instantly, and continuously – perhaps through a type of modern-day *sobornost*.³⁵ I can also imagine that the desire for physical closeness will inherently be reduced when one's body has become modified with robotic limbs (as part of a sci-fi cyborg) – indeed, the physical act of love becomes much less appealing when cybernetics and circuits are introduced as a part of intercourse. Through continued technological enhancement³⁶ of the body, we may find that the mental sensation of physical sex could be delegated to electronically-induced hormone stimulation.

A related topic with regards to the question of community and a sense of belonging that the Borg hive mind challenges may be observed in humanity's response to the pandemic, i.e. the question of medical personnel's decision-making on a community level vs. an individual level. I would argue that few of these caregivers, especially in developed countries, would ever have imagined that their medical ethics training would have needed to move to the supra-level of deciding for the survival of the community against the survival of a specific individual (and

³⁵ This term refers to a Russian idea that underlines the need for co-operation between people at the expense of individualism.

³⁶ This aspect of 'enhancement' may be reminiscent of the Nietzschean *Übermensch* on a conceptual level. Indeed, Sorgner (2009) also argues, that although Nietzsche does not postulate that human improvement can be attained by means of technological methods, he also does not explicitly exclude it (Sorgner, 2009:38). The argumentation is based on an important distinction which is associated with the Nietzschean *Übermensch*; namely that the *Übermensch* is not just some continuation of the human species. That is, the *Übermensch* should exceed a 'higher human' by far. From this implication, it seems unlikely that pure reproductive evolution would at some point result in an *Übermensch* (Sorgner, 2009:38). It would require something more.

thus doing no harm). However, 2020 has shown us video messaging by distraught caregivers warning the uninfected that they, as caregivers, could only actively care for those individuals with the greatest potential for survival (due to limited hospital capacity), whilst only passively assisting those weaker individuals who face an inevitable death as a means to ensure the overall community's survival, i.e. the greatest number. The hive mind analogy becomes less of an abstract thought experiment in this regard, as questions of real-world ethics come to the forefront. For example, if a virus would infect a hive mind and the signal dispersion of a Collective had broadcast a virus to all those connected it would dramatically affect the colony. For the future Collective medico-techno caregiver, however, there would be no ethical dilemma in deciding on treatment allocation – saving one is saving all.

Such considerations speak to an inherent anthropocentric bias in our societies, because the human stands centrally in this connected hive mind and we presume that such configurations of humanity serve as the pinnacle of human intellectual development. In this regard Müller (2014:297) states that we should not hold onto the anthropocentric bias that human intelligence is the ultimate in a set of possible outcomes. This hints at the potential of technological integration to overcome anthropocentrism and to shift away from humanity as the central criterion where we specifically identify Benatar's harm to originate with. Wilson (2017:46) argues that working towards human enhancement through technological means may be hampered by the fact that it is difficult to determine a focus point, or even starting point, for such enhancement – where does human nature begin and end in relation to technology? In his view there is still no consensus of what constitutes being human – therefore it is not obvious where we could (should) deploy technology for enhancement. In addition, Wilson appears to concur with McKenny that such technological enhancement should commence by defining a “normative conception of human nature” first and foremost (Wilson, 2017:46). In sketching hypotheticals of technological enhancement (such as the hive mind), the central question regarding characterizing human nature and its potentially varied configurations in relation to technology is sidestepped.

I argue that such an approach may be too much of a concession to, or even negation of, (our) human natures. In Haraway (2016:59) we are reminded of the highly successful (and yet deplorable) ability of humanity to process information first and foremost via binary taxonomies – the great *dualisms* as per her terminology. The problem is that humanity has

rarely proceeded beyond these binary systems in conceptualizing human nature and existence. Should we content ourselves in merely identifying opposing poles? We have already begun to observe the spectra that form a continuum between these poles (particularly in terms of gender), and more importantly, to acknowledge and try to understand said spectra. Consider closely the case of the current, great gender ‘awakening’ that is the gender *spectrum*. Bearded men wear make-up and dresses, not in drag-club contexts but rather on the city streets and at high society functions (think of the cases of Conchita Wurst, Frederik Robertsson and Nico Tortorella).^{37,38} Such cases may be perceived as publicity stunts at worst, or even phantasmagoria. However, there are instances of which the seriousness of the gender spectrum is undisputed, for example when individuals with secondary male sexual characteristics carry a baby to full term and give birth normally. While human nature has become ever more complex, the central question of what this existence entails remains to be answered as a prelude to determining the success or failure (in terms of overcoming harm) of human enhancement by means of technology (particularly in terms of hive minds).

4.2.3 The shortcomings of hive minds and cyborgs for overcoming harm

As stated in the preceding section, the hive mind and cybernetic enhancement of the individual does seem to suggest that individual uniqueness is lost in an ocean of sameness – the potential for harm is not so much overcome as far as human nature itself is pushed to the side. I therefore argue that such forms of human-technology relations purport to overcome harm by theorizing an alternative conception of human nature (or never asking what human nature entails in the first place). In this sense, we are moving beyond the existential

³⁷ Tortorella explores his own personal experiences in his book *Space Between: Explorations of Love, Sex, and Fluidity* (Tortorella, 2019).

³⁸ I note that all examples provided here are ‘one-directional’, that is they only refer to physical males presenting as females. There are naturally examples of physical females presenting as males, e.g. Empress Maria Theresa of Austro-Hungary being declared *King* of Hungary as the sovereign title of Hungary only had a male version, i.e. there was no queen regnant (although queen consorts did exist). There is also the example of Pharaoh Hatshepsut of Egypt presenting as male where her official statues and written documents show an image of a bearded individual. However, as with the former, this may simply be a case of the social customs surrounding the regnal title. I do believe, however, that it points to the obvious patriarchal tradition in humankind throughout our written history. This gender topic is extensive within itself, and it is considered beyond the scope of this study.

boundaries of harm suggested by Benatar, and thus find that his approach must either be discarded or expanded to include such instances of human-technology relations – or that his approach suggests a traditional account of human nature that such hypothesized technological alternatives cannot truly engage with. This point will be discussed in more detail in Chapter 5. Next, I will investigate what the complete negation of the biological human body may entail for the question of harm to expand on the ideas introduced in the foregoing section.

4.3 Artificial Bodies and Robots

In the current section, the relation between technological replacement of the biological and the concept of harm will be discussed. This is the third and final step of the human-technology evolution which I utilize to structure the discussion in this chapter. I reflect on what options may become, or already are, available to humanity to conduct at least some part of its existence in an inorganic mode, i.e. a mode where full embodiment is not necessary. I investigate this aspect due to the fact that many forms of harm are connected to the human body, and its potential shortcomings, particularly in terms of pain, injury, and biological needs.

4.3.1 What are artificial bodies and robots?

Thus far I have provided several examples of human-technology interaction, or at least technology's ability to serve human ends (see Section 4.1) where the ultimate outcome could be either positive or negative (i.e. prove to be beneficial, or at least to reduce harm, as generally contrasted with the actual introduction of harm). It would appear that a state of 'not crossing over' – where our bodies, our organic expression of self, are not abandoned fully for some inorganic version of existence – may indeed be tenable in terms of obtaining a net reduction in harm. This relationship of interaction should be weighted more skewly in its assessment; that is, can we truly say that there is actually a real sense of interaction between the human individual and the technological? Interaction would usually require active

participation from at least two parties, which is not the case here. However, whether harm may be overcome in its entirety remains a question to be addressed in Chapter 5.

A further alternative to be considered is perhaps the full integration or combination of human physical bodies with technological elements, such as using bionic body parts (as mentioned in the previous section), i.e. cyborgs, or even whole-body replacement, where humans exist as robots. Often sci-fi worlds heavily influence our visions, even our ideals, of what such integration may look like. We may, for example, be able to function in the airless environment of outer space by means of technological devices – prostheses – integrated into our physical bodies to allow for greater comfort and efficiency where a human body would fail. Thus far we have investigated the Borg species, and observed how their technological inorganic enhancements could (supposedly) make technological assimilation feasible and even desirable (though not escaping from the constraints of harm that Benatar mentions).

However, the integration of humanity and technology is much more far-reaching than the example of hive minds suggests. The next, and possibly final, step of technological development may be the full assimilation of our very being with our technologies. This higher-level progression may also overcome the problem which has been persisting beforehand: that much present technology is still relying on human initiation and is therefore vulnerable to human shortcomings in the initiation processes – whether algorithmic, cybernetic, or based in hive minds, a human component remains.³⁹ What happens when the biological body of the human is no longer part of the equation, what does the negation of the biological imply for our understanding of harm?

4.3.2 Artificial bodies, robots, and existential harm

In this final analysis of technology-driven alternatives to existence, we find that organics and our entire human-related, carbon-based expressions are declared null and void. The remaining biologic parts that our cyborg manifestations attempt to maintain and enhance are abandoned for a further third step in development of human-technology interaction, with the

³⁹ Consider for example how the programming of cyborg body parts might also be vulnerable to programming errors or computer viruses, which would make the analogy to biological viruses much more tangible.

last remaining element being that of the *human mind*. With this step, I am referring to the complete emulation of a (human) brain, in essence its ‘uploading’ – whether to a robot body, or a type of online-Internet environment. As described in Bostrom and Yudkowsky (2014:325), such uploading is a conceptualized technology, at present still belonging to the future. The uploading process’s result will be the availability of an entity’s intelligence being preserved, and able to act by itself, and interact, with other entities. A significant problem, or at least uncertainty, in this process uploading, relates to the identity of the new uploaded entity (also see Footnote 45). At the moment of the completion of the upload, the two entities – one organic and one inorganic – may indeed be the same individual, but their continued existence will unlikely be parallel (Bostrom & Yudkowsky, 2014:325).⁴⁰

Returning to the notion of the human mind, Carter (2007:202) indicates that even when artificial neural networks are utilized to model or mimic cognitive processing as one would expect of a human mind, there remain three areas which he terms ‘hard’ problems. These hard problems refer to the human individual’s consciousness, their personal identity and their emotions.⁴¹ These specific ‘hard’ problems are briefly surveyed with the intention of illustrating how difficult the postulation of a human nature may be, and how technological engagement makes the problems even more difficult (even before relating said presumed human nature to the overcoming of harm).⁴²

⁴⁰ There is noteworthy questions of ethics that arise in this instance of multiple entities. As described in Bostrom and Yudkowsky (2014:326), the supposedly parallel existence may be occurring at different rates of times. If both of the entities commit the same crime where the punishment is jail time, the jail time would naturally have to be different for the different entities. Another example relates to pain alleviation, but still within the context of different rates of times. If an uploaded mind is existing at a faster rate than the original source mind, would it be considered more imperative to alleviate the pain of the uploaded mind than that of the original mind?

⁴¹ I briefly follow his reflections on said problems, referencing the technological with the vaulting terminology of artificial intelligence (AI) as opposed to Carter’s neural networks.

⁴² It should be noted here that this distinction between ‘hard’ and ‘easy’ problems originates with Chalmers (1995). Chalmers set up this distinction in order to distinguish two separate problems as understood under the central topic of consciousness. Under the ‘easy problems’, Chalmers (1995) is referring to how a human individual (or rather its brain) is capable of differentiating between sensory stimuli and subsequently respond accordingly. Such problems are aimed at “objective mechanisms of the cognitive system” of the individual (Chalmers, 1995:81). The ‘hard problems’ however, are aimed at “the way things feel for the subject”, where this feeling or sense is anchored with the subject (Chalmers, 1995:81). In summary, the hard problems focus on how facts / knowledge from the physical realm cannot be directly linked with the human individual’s conscious experience.

4.3.2.1 *Consciousness*

Perhaps the most important of these problems is that of consciousness. As we are waiting with bated breath for, and most even fearing, the supposedly devastating AI-singularity⁴³, I wish to point out that the uncertainty and fluidity of descriptions and terminology referring to said singularity are analogous to the flux in views of what constitutes consciousness. I believe that the definitional problem of both terms may be closely related in the context of this study. As Carter (2007) explains, the concept of the conscious remains a somewhat contentious topic, more so in the field of philosophy. One may note, for example, the distinction between being awake and sleeping, which would cover a partial understanding of being conscious. A further understanding, and highly relevant for the question of computerized minds, is the awareness of self, and that the entity is distinct of other entities. A more demanding definition is that of “having the capacity for subjective experience and for having the associated privileged first-person qualities of experience – *qualia*” (Carter, 2007:203).

Although I do not delve into the vast topic of the philosophy of mind in this study, I wish to provide examples, which I believe serve as exponents of the aforementioned ideas around consciousness and how it may figure in the technological. The most obvious is the wakeful and sleeping state of the entity; this is easily seen even on our current personal computers: the computer can be actively used, or a user can set it to ‘sleep’ mode. Even our human dream-states and subconscious have some relation to our computers quietly downloading updates – whether we are working on it or set it to sleep mode. The idea of downloading an update also exemplifies a type of self-awareness for me. With this self-awareness, I am referring to the step beyond ‘mere’ consciousness. As defined in Ferris (2012:28), consciousness is restricted to an awareness of the entity’s physical manifestation (body), the environment of that manifestation (the surroundings in which the body exists) this is the recognition of that consciousness. Self-awareness is the next step of the entity realizing that

⁴³ Here I refer to the technological singularity where there is predicted / expected to be ‘...an event or phase that will radically change human civilization...’ (Eden *et al.*, 2012:1). Other thinkers provide a more narrow view, which links closer to my topic of consciousness here, where there will be an exponential increase in the intelligence of computers (software) to the extent there improvement cycles will increase intelligence whilst decreasing development duration for every generation (Eden *et al.*, 2012:2).

it is aware of its physical manifestation and environment. Therefore, the download, although at a fairly low level, may signal this computer's self-awareness.

Returning to the technology-driven alternative to existence described in the previous section as centred on a Collective, we would find that full integration of human mental processes with technology would allow one to update oneself in a self-aware manner – much as we presume AI would do. In this regard we may note that an update would allow a computer (with a human consciousness) that is connecting to the Collective (the Internet in this instance) to change itself, and by comparing itself to the specification of the Collective (the latest configurations listed on the Internet for a programme on the computer) that entity would show self-awareness by realizing the latest information or update is missing and thus that entity would seek to align itself with the Collective's programming.

Of course the underlying problem remains, that self-awareness was induced in the case of AI: we have programmed the computer to compare itself to the latest configurations and thus identify potential upgrades. However, consider the counter-example in humans of how self-awareness is induced: are children born with a spontaneous desire to share, and to share equally? Shultz *et al.* (2011:221) – drawing on earlier results from Van Schaik (1983) – indicate that by means of evolutionary trends, based on ancestral experience, the awareness was created that living in societal structures and behaving in social and neighbourly ways the greatest benefit can be achieved for one and all.⁴⁴ A further variation of the human example is how individuals with mental states that may be considered 'deviating from the norm' can be conditioned, whether therapeutically or medically, to act contrary to their potentially psychopathic tendencies. Unlike the 'normal' individual, they may likely never have full understanding of their learned behaviour, but through a process of appropriate action-reaction activities they may learn that by means of following a checklist of behaviours they can function successfully in society.

⁴⁴ Ward and Webster (2016:124) indicate that is undoubtedly overall, general benefit to group-living in the case of several species. However, they do acknowledge that the individual benefits (payoffs to use their term) are not equally distributed across the society.

4.3.2.2 *Change*

With respect to the second problem, Carter (2007) defines how a person is always the same individual over time; even though they may move to a new house or travel to work, or as they age or change their hairstyle. In essence, you are still you: “Although I am qualitatively distinct from day to day, I am numerically identical” (Carter, 2007:204). Carter dwells on the notion that *psychological continuity* is necessary for constituting the personal identity of an individual. However, as is the case with an amnesiac, we obviously cannot say the individual has ceased to exist; moreover, the individual is by most definitions still the same person who has lost their memory. A further appealing, but as of yet idealistic counter-example is that of being ‘processed by a matter teleportation device’, where the individual is manifested, by accident, in duplicate.⁴⁵ We can certainly understand his claim that – especially at the moment of manifestation – the two individuals are the same person, thus implying a loss of the numerical identity.

Carter extends his questioning on this topic by reflecting on where we would place thresholds in terms of change, or crossover, which would be distinct from the personal identity so heavily guarded by our psychological selves. In his questioning, he presents a thought experiment where an individual undergoes an initial single unit replacement of one brain neuron with an artificial neuron. If one continues the replacement up to the point of a full brain replacement, at what point did the psychological continuity break down? We may then share in his uncertainty of the case where an individual’s mind may be fully replaced by the technological – even if such an instance still has a physical expression such as a robotic body. Carter concludes in asking what obligations we have towards that new expression of the mind, e.g. a robot body (or even to an ‘online mind’).

4.3.2.3 *Emotions*

Carter’s third and final ‘hard’ problem relates to that of human emotions. Carter (2007:205) again turns to human psychology and reminds us of the distinct minds we may encounter in psychopathic individuals, and thus also their version and handling of human emotions. He

⁴⁵ We note a comparable example in Hofstadter and Dennett (2000:3-7) where the astronaut who is miraculously returned to Earth from Mars, starts questioning herself as to her real identity: did the *real* her die on Mars, or is the *real* her now clutching her relieved daughter?

also reflects on sci-fi filmography and explains our anthropocentric bias through the fact that we would deem technological beings (which may be computerized minds or AIs) as 'equals' once they exhibit emotion, or preferably, their own inherent, spontaneous emotion that conforms to a human understanding thereof. However, argues Carter, we should not fall into the sentimental view on the intangibility of human emotions Carter (2007:205). Even though Carter describes our understanding as limited, he notes how easily actual emotions may be reduced to pure chemical reactions and to combinations of neurotransmitters.

I therefore consider that even emotions may be programmable once we have outlined the necessary chemistry and its reactions. Although the individual will no longer require the flow of protons and their electron clouds – at least not in the organic sense – the ultimate logic guiding those chemically-induced decision-making, emotion-exhibiting actions can be diluted to mere true-false statements with if-else specifications. The output of emotions, so determining and yearned for by humans, will actually be decreased and perhaps even become unnecessary. With this output I am referring to aspects such as gestures and acts between humans: hugs and kisses, saying 'I love you' that presuppose an emotional reaction. When full integration of human mental processes with technology has occurred, or when we engage with AI entities that need not smile at each other to confirm their happiness (they merely consult each other's logical test outputs) we will have moved far beyond traditional conceptualizations of human nature. And finally, such a technological alternative to our everyday existence leaves no heart to ache with longing; the online mind finds relief in activating sleep mode.

4.3.3 The problem of Artificial Intelligence as the other

The above discussion has shown that parsing human nature (with the intention of overcoming harm) becomes highly problematic in technological alternatives to existence. There is a sense that human beings that replace their bodies (and minds) with the robotic or electronic render the human being itself more similar to our theorization of Artificial Intelligence (AI) than to conceptualizations of human nature that have dominated in the eras past. We should then consider the relation of AI to our contemporary human existence to extrapolate this hypothetical existence to its fullest.

The concept of AI is complex and has seen many developments as technology itself progressed, but Müller identifies the basic issues in the field as related to computation, cognition and ethics (Müller, 2012:69). AI refers to machine intelligence and was originally conceived to have characteristics that allow the machine intelligence to fool a human observer (who is spatially distant from the machine) into thinking that it is another human being (this is termed the Turing Test). Newer views of AI may relate to machine learning and neural networks, where a machine or computer program has characteristics of intelligence often related to self-awareness.

However, even if AI is tooled with application designs based on neural network or machine learning, there remains an element of initiation which can only be executed by the human participant – the designer is always a presence in its construction. The human participant first has to set up a system of machine learning, first has to ‘code the computer’s script’. The endurance of AI’s functioning may indeed be autonomous and independent; the initiation of AI’s functioning is not. Such a characteristic may only be overcome with the hypothesized ‘singularity’ of AI – spontaneous and inherent self-awareness. Until such a point in time, human-AI relations in the sense of interactions and ‘crossing over’ remains constrained by human inaptness. In an interview with data scientist and academic, Shion Guha, on the meeting of the AI and ethics worlds, it is stated: “There is no such thing as a neutral technology designer” (US Catholic, 2020). This certainly appears to be a harsh judgement against humankind: humans will always engage with technology for the sake of either the engagement itself, or for their own sakes. This underlying biasedness will therefore translate into how we approach the engagement with technology.

Stedman (2019) provides a more optimistic overview when evaluating instances of ethics breakdowns, and ultimately harm, arising from heavy data mining techniques with regards to personal data within business environments in relation to AI.⁴⁶ He does not gravitate towards the preceding ominous view declaring non-neutrality, but rather states that humans – thus the programmer and initiator or the AI event / process – can be ‘led blindly into things’. This is an aspect to take into account in our assessment: AI breakdowns in ethical terms will

⁴⁶ It is important to note the link here with algorithms, which is linked to AI by means of contemporary machine learning approaches.

undoubtedly have elements of *unknown unknowns*.⁴⁷ This is because we cannot with certainty predict what such intelligence will be capable, since we are engaging with an outcome space that has transcended our human point of view. However, the harm which arises from AI breakdowns, whether intentional or not, still seems to originate with the human initiator of the AI. The previous section notes that the interaction between humans and technology is not without its perils and a balanced co-existence may not lead to a greater reduction in harm – in fact, tracing how we might engage with AI shows how human nature needs to shift fundamentally in order to overcome many of the harms inherent in existence.

4.3.4 The shortcomings of artificial bodies and robots for overcoming harm

However, as with all versions of paradise there is a delectable lure to downfall. Robotic existence inherently contains shortcomings as an intrinsic part of the human mind designing and inhabiting the robotic manifestation. Sullins (2011:236) investigated realistic questions that arise around *affective* robots where he describes their efficient design for social engagement to become so compelling that the human individual – whatever its form – will prefer the company of robots. A further worry is the service of sex robots, where the human individual's addiction can lead to full-time absorption and effectively make it cease to exist truly. Robots providing care-giving tasks provides an additional means for the indirect benefactor (parent of a child, or child of an ageing parent) to have more free time and shirk their responsibility with respect to direct engagement over time.

4.4 Technological alternatives and the negation of harm

My summary of how such a technology-driven existence can purport to overcome harm can be collected under the overarching theme of technological existence leading to a renunciation or redundancy of physicality. Such a summary sounds abstract at best, and I wish to illustrate concrete examples, however idealised, of how the technological can be extended to alleviate

⁴⁷ Based on the extended definition that all information can be classified according to known knowns, known unknowns, unknown knows, and unknown unknowns, as purportedly put forward by the former USA Secretary of Defence Donald Rumsfeld in his commentary on the USA's military intervention in Iraq after 9/11.

three specific instances of harm of which we have become increasingly aware during the COVID-19 pandemic. The three instances of harm I comment on here are grouped under societal structures⁴⁸, gender imbalance⁴⁹, and species interaction. Although each of these topics are vast fields within their own right which could never receive sufficient attention in this short study, I wish to link anchoring elements of said harms to the possibilities of harm reduction that a technology-driven alternative existence can offer us.

4.4.1.1 The question of societal structures

In terms of societal structures, in 2020 the social crises of #BlackLivesMatter and its associated protests and ensuing violence in the USA is of significance. A great historical divide exists between black and white social groups. Histories of slavery – and especially its celebration and glorification by colonizing countries – were confronted directly and openly (Bouffieux, 2020; Phiri & Wright, 2020). Disproportionate statistics of COVID-19 related deaths affecting non-white communities came to the forefront in countries with diverse demographics. Middle-Eastern groups (in a thwarted instance of actually referring to practising Muslims) received negative attention in Western Europe due to their non-compliance with preventative measures against the virus (Svendsen, 2020; Wiegman, 2020). In a moment of enflamed COVID-19 passion, various social groups rebelled against their governments' COVID-19 prevention measures with reasoning based on a presumably divine-given right to breathe without masks. Another fracture of societal structures was the distinction between classes, as categorized by monetary wealth. Consistent news headlines confirmed the significant increase in shareholder wealth the world's richest had experienced despite the economic havoc which the pandemic caused for the average world citizen (Letzing, 2020). In terms of the presidential elections in the USA in 2020, it was also noted how rich donors can have a powerful effect on political focus. In Europe, the rich northern countries were reluctant to provide unconditional financial support to southern counterparts often considered as wasteful spendthrifts (Heikkilä & Burchard, 2020). In South Africa, in the deepest moment of

⁴⁸ For brevity, I use this term in a wide sense: I envision the societal structures such as racism and classism under this overarching term, whilst acknowledging their distinct fields.

⁴⁹ Again for brevity, I apply the term in a wide sense: I consider gender fluidity, the battle of the sexes, the concept of gender vs. sex, LGBTQI+ issues to be grouped under this undeserving anchoring nomenclature.

the crisis, the lure of (misappropriated) funds could not foster any conviction of national solidarity and unity: Ubuntu appeared to be unobtainable.

In a technology-driven alternative existence described above, there appears to be an outcome to remove, and hopefully eradicate, those harms associated with societal structures. For when we are 'incarnated' to exist as online minds only, racism will need to find a new interpretation, if it could survive at all. With a Collective as central point, which desires the persistence and integrity of each individual, a world of artificial bodies certainly seems appealing. Additionally, the powerful political lobbyists would have no meaning in this world anymore; there is now no goal other than continued improvement shared among all individuals. Should the individual feel inferior to the so-called special signal booster function of the (Borg-inspired) queens in a dispersed Collective, then signal boosters would be installed for everyone and mutual integrity is guaranteed yet again.

I suspect, however, that religion would not be completely avoidable in this world; yet I assume that it will converge to the one ultimate point of concern of every individual of the technological world: energy. Energy consumption remains a driving force for a technology-driven existence; but in a technologically structured alternative to existence, it would be a mutual drive. Once all robotic entities require it, no distinction between such entities would remain. Finally, geographical (and its underlying vain) distinctions would be as meaningless as the border fences we once erected, often so randomly; once we are all online, with our artificial bodies, we would be the Collective, and the Collective would be us.

4.4.1.2 The question of gender imbalance

I now turn to how a technology-driven alternative existence can assist us in overcoming the gender-related aspects we encountered earlier in this section. I quote⁵⁰ McEwan (1994:51) for a poignant summary of many of the issues (and potential reactions and responses thereto) around gender ethics:

⁵⁰ The quoted section reflects the adapted version as used in the film, *The Cement Garden*, as in the song by Madonna; the original text in the McEwan book varies slightly.

*"Girls can wear jeans and cut their hair short, wear shirts and boots.
'Cause it's OK to be a boy. But for a boy to look like a girl is degrading.
'Cause you think that being a girl is degrading. But secretly you'd love to
know what it's like... Wouldn't you? What it feels like for a girl?"*

The quote became mainstream in the film, *The Cement Garden*, and was later also used as an introduction to a song by singer, Madonna. Ruti (2018:xiv-xix) zones in on this blunt social, even cultural, tradition of contrasting the functional roles of the genders based on the physicality of those genders. She mentions often hearing sneering remarks such as "grow a pair", or upon completing some physical activity unsuccessfully it being described as having been done "like a girl". Her extreme shock at these differences may perhaps be attributed to her Scandinavian upbringing and the almost austere gender neutrality prevalent in such societies. Ruti (2018:48) explains how confounded she was upon realizing that her native Finnish does not even accommodate gender differentiation for personal pronouns: both 'he' and 'she' is represented by a single morpheme 'hän'.⁵¹ In neighbouring Sweden a new gender neutral pronoun has been created and its use is actively encouraged (Bergström, 2019).

We see further elaboration of an incident where public safety officers castigated Ruti's boyfriend for allowing her to change their flat tyre given that she was a 'lady' and that task is supposedly not lady-like. The latter example is an apt illustration of how there are also harmful instances where "men suffer from penis envy just as much as women do" (Ruti, 2018:x). According to Ruti, this explains that even when women feel degraded or inferior to the heteropatriarchy there is also a type of counter-example among individuals physically defined as males when they do not naturally fit within this categorization. Such male

⁵¹ A similar situation occurs in many of the languages of the Bantu language family. Also in spoken Chinese versions, there is a single pronunciation of 'tā' even though the written forms show distinct versions (他 'he' vs. 她 'she'). Compare this with Romance languages where gender distinction is explicitly required on adjectives, and Arabic where gender is distinguished on adjectives and personal pronouns for 1st, 2nd, and 3rd person subjects and objects. For the two latter languages (and virtually their whole language family groups) we can confirm that there are clearly defined gender roles in their society, and thus the languages suit their users well. However, we are well-acquainted with comparable distinct gender roles in Bantu cultures, and yet their languages do not necessarily reflect such traditions so sharply in their grammars – similar to the Finnish structure. Do we need to come up with a new version of a more accommodating Sapir-Whorf hypothesis: the hypothesis that an individual's language influences their *gender* view, and not just their world view?

individuals may be ones, which identify as part of the LGBTQI+ individuals⁵², or even be described pejoratively as instances of beta-males.⁵³

Benatar (2012) gives us a completely contrasting view, where he investigates themes around discrimination against men (and boys) as rooted in social or cultural practices. Briefly neglecting his core message of reducing and ultimately removing humanity from existence, Benatar reflects on how the male individual's body leads to the conscious concession of additional or worse harms to be induced on it. We firstly encounter the theme of military conscription and service (Benatar, 2012:26). As pointed out, Israel is one of only a handful of modern states where military service is required from female citizens – perhaps an unexpected point in terms of gender equality – however, the requirements for service differ between sexes. Men are subject to longer training and extended reserve status periods. Moreover, their marital status does not change their military reserve status, whereas for their female counterparts, exemption is then a given. Women also rarely or never see active combat. Obviously, we have to ask ourselves whether modern societies are ready to send a pregnant daughter / wife to a warfront in the name of equality. A second theme centres on modification of the male body: circumcision (Benatar, 2006:41). Culturally there is significant acceptance of the practice whereas the female version is mostly abhorred throughout the world, with strict legal enforcement in place to prevent it, except in certain so-called developing countries. There is also the possibility of compounded trauma after the actual harm of the circumcision event; for example, during the Holocaust his male body would reveal a Jew's ethno-religious status much easier than in the case of a Jewess's body.

When drawing on the harms we see identified in both Ruti and Benatar, whatever the gender of that body, it shows that the physical expression of gender differences inherently lead to harms for the individual – whether psychologically or physically. An alternative technological existence with no gender expression seems not be subject to such harms. Living in a society of artificial bodies with extreme equality appears to negate these harms, and even lends itself

⁵² This term is still undergoing expansion in its definition. For this study I make use of this form where it represents Lesbian, Gay, Bisexual, Transgendered, Queer, Intersexed. Other concepts include inter alia pansexual, asexual, and being an ally of the community.

⁵³ Urban dictionaries define further classification such as omega, gamma and sigma males. I do not contrast these in detail, but rather underline the counter-examples as the opposite of alpha males.

to make allowance for infinitely more granular distinctions, should this society wish it for some reason. The artificial body or robot existence can accommodate innumerable categorical mutations – not that it would need to: the possibility of the distinctions flow naturally from a system where all updates / upgrades / improvements are dispersed instantaneously across all individual entities. Playing a bearded warrior queen in your digitally simulated world is merely an .exe-file away. Differentiation in this existence will purely be out of personal interest, curiosity or frivolity, for there are no physical characteristics that may be magnified by a society sensitive to such characteristics in their conscious and unconscious taxonomical judgements and discriminations. Robotic existence demands only energy, and even where there is subatomic differentiation of (energy) particles, would such a society dare repeat the mistakes of their organic ancestors in rating such particles against each other?

4.4.1.3 The question of species interaction

The third instance of harm I investigate here is structured around the interaction between species, and by extension, the manner in which humans engage with animals specifically. As COVID-19 rages on and we await the introduction of a vaccine – and in line with Benatar's anti-natalism – I argue that it may be hoped that only a minority believes that a vaccine must be found at whatever cost to non-human life. However, even when society does not find itself in a moment of crisis, its usage of and value extraction from (or rather exploitation of) the animal kingdom are unfortunately abominably boundless.

Johri (2020) explains that as human carnage unfolds in the USA, it is quietly accompanied by carnage among 'lab animals' – animals used in testing experimentation for finding a vaccine against the novel coronavirus, as well as animals not directly linked to any COVID-19 vaccine research. As for the latter, Johri indicates that many researchers were categorized under vulnerable groups for infection and morbidity, thereby leading to a decrease in the number of animals used for research since the researchers could not actively continue their work. This in turn leads to the animals being culled. In the case of animals being used directly in testing, the massacre has many facets. With the delays in stock levels and relays of inputs necessary to breed and keep the animals necessary for testing, many of the animals bred may ultimately turn out to be 'unusable' for vaccine testing being either too old or young, or whole projects

may end up being cancelled (perhaps due to the application of initial research assumptions which are later confirmed as incorrect).⁵⁴

While we are at least making headway in the fight for equality among human individuals, equality among species seems far off, if not impossible, with the human body and mind in its present format. Therefore, at presents it may feel sensible and natural to believe (and even justify) the harm caused to animals for the benefit of humans. We need also mention here the horrors of factory farming for humanity's supposed nutritional needs (indeed, our organic bodily health still requires continued animal sacrifice). The god to whom the sacrifice is made has changed somewhat; no longer is it an unseen and demanding entity inhabiting some mountaintop temple, but instead the scholarly field of Science. This new god's relationship manager are no longer priests in ornamental garb, but rather researchers in white coats working towards a new interpretation of the *greater good*.

The greater good in this instance is that of the continued survival of humanity. Humans insist on and enact these sacrifices out of necessity for the maintenance and persistence of that body, which suggests that the sacrifice of non-human species are unavoidable whether for medical (animal testing subjects), security (guard animals, beasts of war), or nutritional (meat consumption) sustenance of human bodies. Anthropocentrism reigns limitless in humanity's approach to existence. And yet the cracks in this selfish conviction have begun showing; Haraway (2008:11) describes (via Derrida) Freud's notion of how humanity justifies its anthropocentrism by means of "the fantasy of human exceptionalism" despite three distinct historical breakdowns of this belief. Copernicus destroyed the spatial centrality of anthropocentrism by proving a heliocentric solar system. The second breakdown came from Darwin and his notion that all existence is a (slow and long) game of survival of the fittest. And finally Freud's identification that humanity is *mostly* reacting on, and behaving according to, an unconscious element of the human mind. Whilst Haraway (2008:12) posits a fourth breakdown, her classical 'cyborgian' vision, I posit as central claim of this study that none of these postulation really serve to address the question of harm; even if we wish to move past

⁵⁴ Early on, a potential vaccine did lead to effective blockage in the kidney cells of a species of monkey. However, the monkey species has a special enzyme that aided the working of the vaccine. Not only is this enzyme not naturally present in humans, the coronavirus uses an alternative method of entering human lung cells (Saey, 2020).

the organic and onto the artificial body / robot existence, we are not addressing the harm that is found in *human* existence.

While a technological existence may serve to void the biologic necessities of health, security and nutrition, and thereby diminish the harm inflicted upon other species to satisfy those needs, the existential harm that Benatar identifies, still remains. Even if one were to envision energy needs as (ultimately⁵⁵) the only external input to be required for an ever-expanding technological existence, even if robotic existence will free individuals to exist with other species, the question of harm rises to the surface as argued by Benatar. In the technological alternatives to existence described in this section, we may find that gravity is no longer a given fact of life. Satellite networks can endure far above planetary surfaces, leaving the fauna and flora on planets in pristine condition to continue with their own existences unhindered by the greedy, sacrificial demands of human bodies. However, the harm of coming into existence remains a crucial consideration.

4.5 Chapter Conclusion

In this chapter we explored the dream-like worlds of alternative existences which can be, and are already, driven by technology. Three central examples were presented to illustrate that much progress has already occurred in our daily life; e.g. the preponderance of algorithms we encounter constantly. This alternative existence is likely the most relatable alternative for the reader, since the examples of vending machines and smart homes are tangible and concrete right now.

Over and above the first alternative existence, we have seen that the second alternative existence of hive minds and cyborgs offer many positive changes in existence, especially in the context of allowing connection between individuals. In the dark days of the COVID-19 pandemic, which saw self-isolation and working-from-home, an encouraging message from our local Collective queen may provide a moment of comfort. When one's organic human

⁵⁵ I say *ultimately* here as I am confident that the initial transition to an artificial body / robot existence will obviously still require mineral resources to construct the infrastructure required; e.g. the hardware and energy conversion components to sustain such an alternative existence. Once science allows for full manipulation of matter, such activities can decrease and finally desist.

lungs fail due to COVID-19 , the Borg technician can surely provide one with a replacement set faster, and likely with much less effort, than a lung transplant.

The third alternative existence of artificial bodies and robots, despite its severely abstract nature provides a utopian option: gone is the inequality of societal, cultural or racial structures; gone is the battle of the sexes and the ugliness of all of the ‘-phobias’; and finally also, gone is humanity’s exploitation of supposedly lesser species. When individuals are expressed through, and exist as robots with artificial bodies, we can all be at one with each other.

However, in each of the three alternative existences I already hinted at possible shortcomings. Paradise does not exist, at least not for now, here on Earth – what we encounter in such alternative existences are utopian hypotheses. Algorithms show the first cracks in the breakdown of humanity’s bodily desires, particularly in terms of desire-fulfilment. Whilst hive minds and cyborgs still allow for some leeway with regards to physical expression, individuals have turned into a ‘faceless’ mass of uniformity. For artificial bodies we face the complete rethinking of our own human natures; life with, and as robots, may lead to strange new *bad* developments. Indeed, such a life leads to the introduction of new kinds of harm that relate no longer to *human* being, but to a form of existence that resides beyond our contemporary understanding. In the next chapter, I will investigate the shortcomings of the three alternative existences more in-depth and illustrate that the question of harm remains a central consideration in each of these hypotheses.

5 Chapter 5: Can harm be negated?

Does an existence based on technological alternatives, considered in light of Benatar's anti-natalism argument, serve as a sensible avenue to overcome the harm associated with existence? In the current chapter I explore how such alternative existences cannot serve to overcome the anti-natalist conclusion. The goal of this chapter is therefore to investigate to what extent an alternative existence based on technological alternatives fails to negate harm as described by Benatar.

I introduce this chapter by means of an example, to illustrate how technology-mediated alternatives to existence succeed in shifting us away from being as a human being, and additionally fail to overcome harm. Omohundro (2014:307) relates how a computer programme developed for playing chess can overreach as far as hacking into banking computer systems and transferring money. In this example, we see that the original goal for the existence of the entity (the computer programme playing a game of chess) has become misguided and its existence results in harm (transferring money – robbery). The failure of this alternative existence should then reveal why we need to focus on the theme of the *intent being misguided*. And yet, the very principle of a technology-driven alternative existence would be to have inherent autonomy: to be free to exist. This, however, may in turn bring with it an arbitrary outcome space – we cannot predict what such intelligence will be capable of (King et al., 2019:6) because we have moved beyond the confines of *human* beings.

In this chapter I firstly revisit Benatar's RAP and, by using an example comparable to the preceding one, I illustrate how the RAP becomes a practical measuring tool for assessing a life's worth, or more preferably in the technological alternatives' context, an existence's worth. In revisiting the RAP, I conclude with a guiding principle – an increasingly strict targeting scheme – of analyzing when pains and pleasures, goods and bads, or harms and benefits are present in an existence, and what our judgement on the existence thus should be.

The bulk of the chapter is subsequently devoted to using the RAP practically – reverting to the *guiding principle* in extreme cases – in assessing the viability of the three proposed

technological alternatives, and showing that harm remains wherever and whenever existence endures.

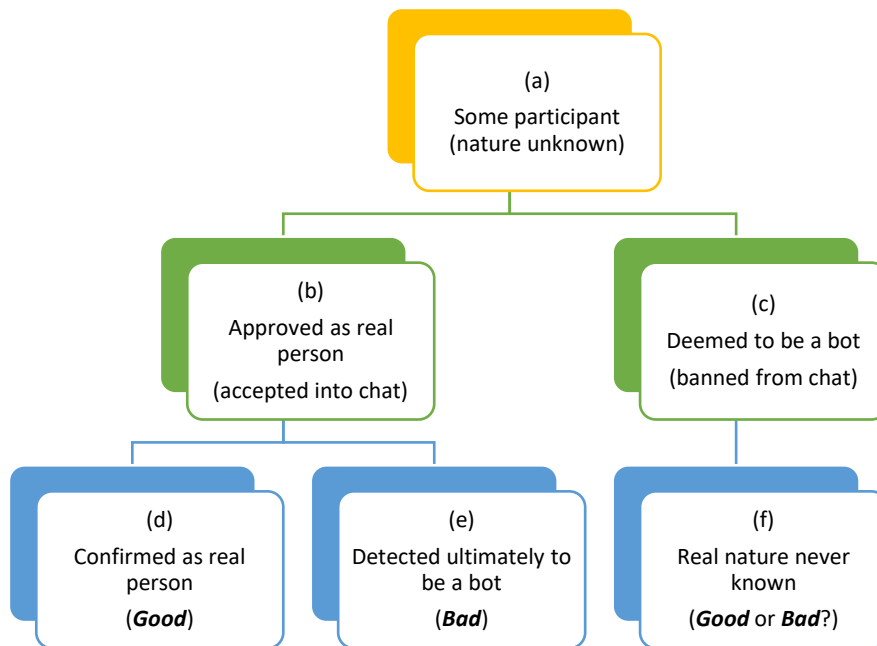
5.1 Return to the RAP

In this section, I illustrate how we can apply Benatar's Relational Asymmetry Principle (RAP) in order to gauge its practical application. This practical application is necessary as a measurement tool for the rest of my argument, serving to illustrate why technology-driven alternatives to existence continue to falter in the face of Benatar's ultimate goal of avoiding harm. In setting out these practical applications, I not only demonstrate a robust system of evaluation of *Good* and *Bad* in terms of existences, but also conclude with a hierarchical system within this evaluation system.

King et al. (2019) describes an aspect of artificial intelligence, which I consider as analogous to the relational asymmetry principle found in Benatar (2006), which was discussed in Section 2.3. The example provided, which illustrates its analogy to Benatar's RAP, centres on the monitoring of cyber-attack bots⁵⁶ active on social media websites (King et al., 2019:8). The first step of identifying which participants on social media websites are bots already runs into complications; namely the inherently difficulty of identifying a complementary subset of participants which are deemed true participants – real individuals (i.e. people). Once a participant is characterized as a bot it will be banned from further participation on the social media site. It is only the subset identified as bots, which is actually verified and classified to be either true (real individuals) or not (bots impersonating individuals). The complementary subset is never verified for bots, which were not detected *unintentionally*. What about the incidence of true participants, which were banned due to erroneous classification as bots? I illustrate this example in Figure 1.

⁵⁶ I use term the author's term of 'bot' here as a distinction from the expected robot, in order to emphasize that this entity is not a physical manifestation as we would normally expect of a robot, but rather an online entity with an AI-based existence on the internet.

Figure 1: Classification of chat bots



Source: Author

In Table 3, I re-develop Benatar’s relational asymmetry principle (RAP) to the case of social media participation monitoring. Firstly, I deconstruct Benatar’s RAP to the four aspects of *Incidence* – the entity is supposed to be existing; *Confirmation* – the entity is confirmed as present or absent; *Nature* – a classification of whether the incidence is beneficial or harmful; and *Impact* – an evaluation of whether the incidence will have a positive or negative effect dependent on whether its presence is affirmed and what its nature is. The latter aspect of *Impact* corresponds directly to the assessment of *Bad – Good – Good – Not Bad* in Benatar’s quadrant approach in Table 1 (see Section 2.3).

Secondly, I apply the aforementioned four aspects of the RAP to the example of the cyber-attack bots; this application is tabulated in the lower half of Table 3. Note the analogies between Benatar’s original aspects and how these can be interpreted for the bot surveillance: for aspect *Incidence*, the distinction of [existence / non-existence] corresponds to [detection / non-detection] of some entity’s presence (whether true individual or bot); for aspect *Confirmation*, the affirmation of the entity being present or absent remains the same; for

aspect *Nature*, a bot is interpreted as ‘harm’ since it will hinder true human conversation in the chat room whereas a true individual is considered as ‘benefit’ since it will allow for conducive chat room discussion; and finally for aspect *Impact*, the designations of positive / negative remain the same. The latter of positive / negative are the representations I use here for Benatar’s *Good / Bad*. Note specifically the designation ‘Negative ?’ which represents Benatar’s *Not Bad*, and thus the asymmetric core of the RAP.

Table 3: Application of asymmetry principle on AI monitoring of social media sites

Context	Incidence	Confirmation	Nature	Impact
Benatar’s relational asymmetry principle	Existence	Present	Harm	Negative
	Existence	Present	Benefit	Positive
	Non-existence	Absent	Harm	Positive
	Non-existence	Absent	Benefit	Negative?
Monitoring of social media participants through AI	Detected	Present	Bot (Harm)	Negative
	Detected	Present	True individual (Benefit)	Positive
	Not detected	Absent	Bot (Harm)	Positive
	Not detected	Absent	True individual (Benefit)	Negative?

Source: Author

In applying Benatar’s RAP in relation to the bot-example, I firstly observe that the aspects of *Confirmation* and *Impact* remain unchanged and directly interpretable. I secondly conclude that harm may change its superficial or detected nature (i.e. aspect *Nature*); but its core result of negative impact (i.e. aspect *Impact*) remains intact. In showing this persistent underlying nature and unavoidable negative impact of harm, I deduce that all incidences of existence are ultimately harmful, and should be avoided. I therefore summarize Benatar’s RAP as the following *guiding principle*:

- It is good to reduce harm;
- It is better to eliminate harm; but
- It is best never to introduce harm at all.

This logical structure indicates that, even when sketched against a technological framework, harm remains.

5.2 Why does harm remain?

This section will suggest why harm remains in each of the three alternatives to everyday human existence introduced in the previous chapter. I relate my judgements to the RAP, and where necessary the *guiding principle* to explicate my developing argument.

5.2.1 Algorithmic control: Algorithmic technologies can, and will, fail us

In Section 4.1.3 it was noted that algorithmic technologies will not necessarily lead us to a one-directional state of harm *only* being reduced, let alone being entirely eliminated. Instead, the element of mindlessness that is inherently part of algorithmic technologies may lead to a further introduction of harm. If we then only evaluate the process's efficacy based on its ultimate result that may still exhibit a *net reduction* of harm, we cannot claim to be better off in seeking salvation in an existence based solely on algorithmic technologies.

Braidotti (2013:59) provides an example of human interaction with algorithms, in which she demands a further consideration of how harm may still feature in an anthropocentric crossover to a technology-driven alternative existence. In her example – which is highly reminiscent of the red queen hypothesis⁵⁷ – we imagine ourselves engaging in an all too common act of being a client ringing up a call centre. Braidotti explains how the human caller is quickly overwhelmed by a computerized system of selection menus through pre-recorded

⁵⁷ In this reference, I specifically refer to the context of the author Lewis Carroll's *Through the Looking Glass* where the antagonist, the Red Queen, explains to Alice (of Wonderland fame) that "it takes all the running you can do, to keep in the same place". The eponymous evolutionary science hypothesis contains a more detailed definition referencing species' desire for proliferation in their changing environments.

messages. Certainly, the labour input of the service provider has reduced through this automated call centre, but the human caller's experience, and ultimately existence, has encountered a hindrance. The caller often spends more time on the activity; and more often than not, additional emotional frustrations come to the fore. This may be a suggestion that a *crossover*, in effect a symbiosis, between humanity and technology may not be as successful, or at least as harm-reducing, as hoped for.

Braidotti (2013:61) goes on to explain that (commercial) institutions– and surely the ones subjecting clients to call centre services – are investing significantly in “the informational power of living matter itself”. Braidotti explains that whereas Foucault pointed to the concerning notion of bio-power increasingly sought by governments, this approach is now also followed by commercial institutions and it has become a case of art imitating life. Fantasy realities from films such as *Minority Report* and *Blade Runner* showing future humans actively targeted by customized marketing, are in principle already a reality. Although we do not yet encounter gigantic billboards or life-sized holographic sales agents, humans are already receiving highly individualized marketing information where products are customized according to the specific individual's data being analyzed. Through such in-depth analysis, commercial institutions are exerting a bio-power on humans.

At face value such tailored products and services may seem ideal; we obviously know that the institution is trying to gain profit through its sale, but it had made an effort in providing a highly efficient sales process to the human client and that is something the human individual may indeed appreciate. No time is wasted in attempting to unsubscribe from mailing list with irrelevant products or endlessly searching online for the exact item one is missing. Direct marketing eases humans' commercial interactions. Obviously the concern of data privacy exists in terms of how much information one is divulging to commercial institutions – data privacy concerns may include one's physical position being tracked which may allow hackers to compromise one's physical safety. However, even the supposedly positive aspects of greater time and commercial efficiency through direct marketing, may embark on a sinister path.

Financial transaction data processed through heavy computerized analysis such as machine learning can now assist a commercial institution in detecting the client's likelihood of having a baby. For the client a happy occasion is about to occur, and for the financial institution an

opportunity arises for remortgaging the client to extend the home for a nursery, or for additional medical care costs. But a darker opportunity also arises for the financial institution to maximise its profit – or indeed minimize its losses – when data analysis points to an impending divorce. A client’s risk profile – and thus their profitability for the institution – may now change drastically. As the client’s divorce is underway their affordability of extensive mortgaging will change: two separate abodes are now required, and there may be greater travel expenditure as children move between parents. Although the client’s actual attitude to his financial commitments has not been assessed directly, we find that data analysis tells the institution that such commitments appears to have changed and that the client is now a riskier bet. The subsequent step for the commercial institution is to increase interest rates on credit products and in the manner of a self-realizing prophecy, such interventions actively create the financial strain which the institutions were passively expecting. Harm, instead of help, has now been assigned to the client.

Consider also an example of two futuristic motor vehicles being driven by algorithmic technologies on behalf of their owners / passengers. The algorithmic technologies in the vehicles will be moving at speeds much greater than any human individual could ever handle, and on motorways in a manner which maximizes efficient use of surface – thus vehicles are spatially closer to each other in such a manner that, again, no human individual could affect comparably. Should a vehicle collision be imminent due to some external force that cannot be avoided by the algorithms – e.g. a sudden lightning strike causes a tree to fall onto the road – how would the algorithmic technologies behave with respect to their charges (their owners or vehicle passengers), and with respect to other human individuals who will be affected by the collision impact? If we turn to the Three Laws of Robotics⁵⁸ in Asimov (2018), we have limited resources for coming to a clear conclusion. There follows no guidance for how the algorithms of the two or more vehicles should respond individually, or in unison. Should we in turn add an additional layer with a heavy dose of utilitarianism, we could

⁵⁸ The laws are defined in Asimov (2018) as:

- i. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- ii. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
- iii. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

perhaps expect that the algorithms of the respective vehicles will communicate with each other and coordinate to achieve the optimal collision response that results in the ‘best’ combination of reducing passenger death and injury.

Problems in the execution of the algorithmic technologies remain, however, since the best combination of reducing passenger death and injury would require some sacrifices to be made. Furthermore, such sacrifices would potentially not be relieved from the human drama of discrimination. Should the automated cars allow the pregnant woman to live? If the life of another individual weighs more heavily, we may be faced with the second sexism Benatar (2012) describes. On the other hand, perhaps the algorithms already have her medical file on record and know that her blood type is incompatible with that of the baby and that one of them may die during the pregnancy term. In the other vehicle, the algorithm can let the father live since he is the family’s breadwinner and employs many workers at his factory, which depends on his specialized knowledge. Of course, the algorithms could never foresee the human fate that, should the father die, his death ultimately leads to his young son’s commitment to studies in algorithmic technology which eventually revolutionizes the field years later. The complexity of human existence may serve to outwit or thwart even the most logically structured algorithms.

The interplay of human life, the human mind and the value it assigns to its human distinctness is a troublesome aspect when relying on algorithmic technologies with the goal of salvation from the inherent harms of human life. Algorithmic technologies, at least in their present form, introduce new kinds of harms. Thus, even when the net result of existence based on algorithmic technologies leads to a net reduction in harm, we have not truly achieved our ultimate goal of completely eliminating harm, in the sense of it being completely absent (refer to Benatar’s RAP as discussed in Section 5.1). Indeed, the algorithmic technologies may reduce harm by optimizing the collision response, but it would be more optimal never to bring something into existence – in this instance, the part of existence surrendered to algorithmic technologies – that may lead to harm. This also links to how Benatar distinguishes between the human assessment in the present-life and future-life senses. For this example of algorithmic technologies, the analogy would thus be given as present-existence and future-existence of algorithmic technologies as an alternative to purely biological human existence

(see Section 2.2 for the explanation of present-life and future-life sense that Benatar distinguishes).

As Benatar argues, we set much higher thresholds when assessing the value of life in the future-life sense (Benatar, 2006:23). Therefore, we should set that some high threshold for existence based on the alternative of future-existence algorithmic technologies. Thus when inferring from the RAP quadrant approach in Table 1 (see Section 2.3), this example links closely with quadrant (3) where the *absence of pain* and thus harm, is assessed as *Good*. Since we have seen that harm may still be introduced in such instances, this existence should also be avoided.

5.2.2 Hive minds and cyborgs: What remains (of us) in the Collective?

As mentioned in Section 4.2.3, an initial shortcoming of the hive mind is that, even as the hive mind creates order and an overwhelming sense of equality among the connected individuals, the very concept of the ‘individual’ becomes thwarted. The great unison of the Collective drowns out uniqueness to achieve its stabilizing coordination among the parts of the whole (see also Section 5.2.3 for a further discussion regarding the parts of a whole). As with the preceding section’s case study, we are in fact at risk of introducing harm in our attempt to reduce it. The harm that we would be introducing is the loss of self, of the individual.

We may think that the current trend of influencers, and their social power, on Instagram and comparable social media platforms is a new occurrence: a completely new development and response to our technology-permeated lives. However, Floridi (2014) cites French author Marcel Proust who asserts that “our social personality is created by the thoughts of other people” (Floridi, 2014:60). He therefore believes that a *social self* is constituted through human individuals’ engagement with each other, specifically through information and communication technologies such as social media platforms. On the surface this concept may appear to fit well with the notion of hive minds and our unitary Collective existence, but Floridi (2014) continues by explaining that the interactions have a type of ‘feedback loop’. Such a feedback loop means that the interacting individuals are continually shaped and defined due to the hyper-connectivity of social media platforms, which highlights the potential breakdown of the proposed alternative existence of the Collective. To illustrate this breakdown, I

investigate Floridi's argument for the need for privacy (the lack of which, in this account, would be a harm to individuals if said privacy ceases to exist through the loss of self in a Collective). Floridi (2014:102) considers privacy an indicator of freedom. Setting up the privacy state provides freedom for the actor. In considering Floridi's four forms of privacy freedoms, the Borg-inspired Collective fails on all four counts. In the following paragraphs, I investigate each on the basis of the suggested Collective.

Firstly, the individual units of a Collective have no *physical privacy*. The bodily functions and intimate actions that may demand physical privacy are meaningless in the Collective. Biological functions of the individuals have been superseded by optimized cyborg bodies. Furthermore, there is also no need for reproductive actions when all individuals are the same and equal (the cyborg bodies may no longer possess the components necessary for sexual relations in the first place). This naturally also affects the bodily aspects associated with such privacy.

Secondly, *mental privacy* becomes obsolete in a hive mind. While queens may act as signal booster for the Collective, actively maintaining a free flow of information across all the Collective's units, there would be no need for private thoughts amongst the units of the network. There is no personal space for experiencing events, no need for reflecting on their effects on the individual. In this configuration, the process of active manipulation by the signal boosters of individuals to achieve Collective-wide coordination entails a serious risk of harm for the individuals.

Thirdly, the aspect of *decisional privacy* would be in direct opposition to the operations of a Collective, as the individual units simply exist to execute centralized decisions. Floridi (2014) argues that the individual would expect to have freedom of thought and decision-making in their everyday life around themes such as "education, health care, career, work, marriage" (Floridi, 2014:103), but in the Collective the queens coordinate the flow of services such as: i) software upgrades (analogous to education); ii) hardware replacements (analogous to health care); iii) functional execution which in itself is non-negotiable when there are only worker individuals and signal booster queen individuals (analogous to career); iv) operational execution of tasks already decided on (analogous to work); and as mentioned above, interpersonal relationship become meaningless when all individuals are constantly fully aware of everyone who is one and the same (analogous to marriage).

The fourth and final privacy, which is eroded through a Collective existence, is that of *informational privacy*. The individual assimilated by the Collective may have possessed some unique, self-identifying information that that individual kept private. However, in the Collective the signal boosters broadcast all information all of the time. Your past memory of the puppy love for a high school sweetheart, or the relief of finally passing your university degree would no longer be relevant as an individual within a Collective. However, any measure of privacy concerning those memories that you would have wanted to remain undisclosed for the sake of embarrassment, or feelings of inadequacy, cannot be upheld in the Collective. The constant free flow of information between units requires the individual to abandon their ownership of such information.

In conclusion, there is a breakdown of privacy in the process of sacrifice of the self for admittance to the alternative existence of the hive mind. Again, it is apparent that more harm is introduced through the method that was intended for alleviating it. Whereas the foregoing sketch does relate to quadrant (4) of the RAP – that is, the *absence of pleasure* is *Not Bad* in the sense of assimilated Collective individuals no longer really needing privacy – I believe that this technological alternative fails in terms of quadrant (2), that the *presence of pain* is *Bad*. This hypothetical form of existence allows the presence of pain since it introduces additional harm due to the loss of privacy, which results from the loss of self. This alternative existence should therefore also be avoided.

5.2.3 Artificial Bodies and Robots: Over my dead body

In Section 4.3.4 I hinted at the shortcomings that may be introduced and gradually increased by a crossover from biological human bodies to artificial bodies or robotic manifestations. The aforementioned section has shown the central link of the human body, and by extension of existing in one, causing considerable harm to individuals in an indirect manner. While the argument may also present a subconscious case of body dysmorphia, I would suggest that the reasoning appears to hold true and that there is a strong correlation between biological bodies and harm. The harm and bad feelings which I have presented in Section 4.4 are experienced due a certain bodily layout which has an existential component, but which also requires some interaction to be taking place between at least two bodies – as is the case with

gender or racial discrimination, for example.⁵⁹ It is this aspect on which I reflect further here; I will specifically discuss also how it relates to humanity's potential loss of its bodily mode of existence in submitting to an alternative existence based on artificial bodies and robots. The harm associated with victimization or discrimination is often directed against a bodily configuration functioning within the spheres of societal structure, gender imbalance or species interaction; it is rooted in some action originating from one body and the subsequent reaction in another body. Contrarily, there are also instances when two or more bodies interact that allows for a good reaction in the other body (pleasure is experienced).⁶⁰

In Dreyfus (1992) we see a timeline of the phenomenological thought concerning human-technology interaction. He points to Descartes' affirmation that the human individual, in contrast to some technological expression such as a machine, can accommodate infinitely many situations. Dreyfus concedes, however, that ever-increasing computing power may lead to the development of technology where such infinitesimal response capability may at least appear to be possible for the human observer (Dreyfus, 1992:236). The question is then posed how the computing centre – in this study the alternative of a programmed artificial brain in lieu of an organic human brain – can account for phenomenological processing when the input mechanism (an organic human body) is absent. Dreyfus zooms in on a core concept of Neisser's *temporal gestalt* view in this regard, noting that "what counts as a part is defined in terms of the whole" (Dreyfus, 1992:244). This argument has important consequences for the mind-body connection, a dualist connection, which may be lost with a transition to artificial or robot bodies, for the mind-body connection is an inseparable actor in the information processing of parts and wholes. Dreyfus describes 'anticipations' that occur in the relation between body and mind, citing two examples: i) Dreyfus describes how the human body anticipates the beats of a musical rhythm; and ii) Merleau-Ponty's example of how a human individual experiences the feeling of touching silk fabric.

⁵⁹ I acknowledge that there undoubtedly are instances where harm and badness are felt by the individual spontaneously and inherently, thus originating and contained within the individual as a whole with no interaction with a second party. I believe such cases require more study whether at least those individuals should abandon or rectify their body or body parts, which cause the trauma, as per the discussion of technological alternatives to existence investigated in this study.

⁶⁰ Consider for example the sensation that one experiences at a live music concert of an adored singer, or in a more intimate setting, the sexual gratification one body can receive from another.

When I return to the earlier examples of listening to a music concert or being intimate with another individual, this aspect of anticipation in the mind-body connection, as described by Dreyfus, which is a determining factor of assessing the presence of pleasure for both bodies. The rock singer knows what inexplicable feelings her music creates within her own body and being, and the anticipation is electric between the artist and the audience for the conveyance of this pleasure. Similarly, one body has some anticipation in the sexual stimulation that is induced on another body – enduring feelings and thoughts beyond the clinical description of ‘orgasm’.

Following on from this description it is clearer what the abandonment of the organic human body may bring about. We know that we will lose many pains, harms and bad feelings as identified in Section 4.3.4, which allows us to achieve the goal of quadrant (3) in relation to the *absence of pain* that is *Good*. However, in the preceding discussion, I have argued that the human body plays an integral part in affirming quadrant (2) with regards to the *presence of pleasure* which is *Good* as well. On the one hand, replacement of the human body with artificial or robotic bodies (which may present comparable qualities) seems to remain a flight of fancy with current technological capabilities.⁶¹ On the other hand, the nature of such replacement entails a fundamental shift in our conceptualization of human nature.

In response to the aforementioned considerations, I therefore turn to quadrant (4) which states that the *absence of pleasure* is *Not Bad*. However, to achieve this goal of the RAP, we must focus on its asymmetric principle, and combine it with Benatar’s conceptualization of present-life and future-life sense. Therefore, for humans alive today, who have experienced the pleasure of the mind-body connection, it would be difficult to give up a human body given the pleasures they have known. In contrast, though, I believe that for future-lives, which have never experienced bodily pleasures, it would be more compelling to argue that we must avoid those harms that the (organic) body will experience. This can be feasible since the individual which would be inhabiting an artificial or robotic body will have no sense of the absence of pleasure. But alas, it is only the antecedents of the robotic individual who will bear the

⁶¹ Dreyfus (1992) points to Simon’s assertion that it seems plausible that the technological capabilities of simulating a human-like brain will become available, whereas simulating the “comparatively flexible sensory, manipulative, or locomotive system” of an organic body will require much more effort, if it is even possible (Dreyfus, 1992:236).

knowledge of the absence of pleasure being *Not Bad*.⁶² In attempting to focus on the replacement of our bodies and minds with technological alternatives, it must be recognized that for present-life individuals such a shift will necessitate a move beyond the constraints of our current human nature and would therefore be incompatible with the starting points of the RAP.

5.3 Technological alternatives to existence present no harm-free utopias

Through the ages, technology has provided humanity with many freedoms. Technology has freed our bodies from many manual labour, alleviating the physical toil humanity previously had to endure to improve its existence. We may think of technology that has led to (semi-) automated production of food, and shelter in the form housing construction. Technology has freed our minds from many onerous, repetitive tasks and has expanded our access to information (consider how education has changed from its limited accessibility to mass education by means of the printing presses). Technology has also freed our psyches from many previously disturbing states; consider how psychiatric medicine may allow clinically depressed individuals to lead functional lives, or how gender reassignment surgery can set a tortured soul free from its incorrect body.

With the constant increase in computational capabilities and humanity's boundless imagination, technology is fast approaching a status of omnipotence and omnipresence. Such attributes were assigned to gods in the past. Indeed, Clouser (2005) suggests that, starting from Tillich's view of religious belief being anchored in that which is of 'ultimate concern' (Clouser, 2005:12), we can ultimately derive a powerful definition of belief being religious "where the essential core of divinity per se is to have the status of unconditionally non-dependent reality" in relation to technology (Clouser, 2005:24). In attempting to overcome harm, one may ask whether humanity has turned to idolizing and redefining technology as

⁶² Floridi (2014:70) gives an additional take on this. He asserts that once some effect has taken place it does not necessarily require the cause of the effect to continue being present. Following from this he argues that self-identity cannot develop without the body. However, once the self has developed, that state of consciousness may indeed exist fully independently of the human body that brought about, or at least assisted in bringing about, its development. To follow his example: there can be no butterfly without the caterpillar, but the caterpillar need not continue to exist for the butterfly to exist.

tangible divinity on Earth. Technology has often been the catalyst in realizing scientific projects, which many already worship as the ultimate answer to their existence. Noble (1997) relates how the religious elite in the Western world initially had discounted what was perceived as technology, but that technology in the ages following became increasingly linked to religious ideas and motifs.

However, when considering the preceding definition of religious belief and what is godly, in combination with the account of the causal and yet non-interdependent relationship of the caterpillar and butterfly in Floridi (2014:70), it would appear that technology falls short with regards to being unconditionally non-dependent of its non-divine ‘adherents’ for indeed technology needs us, and we do not need it (a return to a pre-technological era is feasible, though unlikely). Technology has certainly made human existence easier, but it is not a prerequisite for it to continue. Could we therefore consider technology a replacement for the oracles, shamans, and priests of old? Could we argue that technology acts as the intermediary for the supposed true ‘god’ that is science, as the communication medium for humanity with its superior? Noble (1997) sees a parallel with Augustine when he draws on Ellul’s view that humanity considers technology to be a part of humanity’s fallen state; that is, in the biblical terms of humanity being born and existing in perpetual sin in their Earthly lives.

I wish to emphasize that technology is unlikely to possess a godly status (which ties in with the preceding argumentation in Section 5.2 that concluded that it cannot provide a harmless alternative to existence) – for it cannot overcome the fundamental existential harms that Benatar identifies without completely pivoting the human individual away from *human* being as we know it today. In my view, Benatar’s anti-natalist conclusion cannot be overcome through appeals to newly developed technologies or even technological hypotheticals, because his estimation of harm is so closely related to the facticity of our human being.

5.4 Experiencing harm is human, and that’s (not) okay

In utilizing the RAP practically and analyzing whether technology-driven existences could offer viable alternative, I have continually been assessing the value of existence. By investigating closely Benatar’s distinction of present-life in contrast to future-life, it seems apparent that (in the case of a future-life) avoidance of said life coming into existence is best – a point

supported by the guiding principle (see Section 5.1).⁶³ However, in terms of the case of present-life existences, the process of determining the value of a life requires a process of netting benefits and harms (through measurement and calculation). However, measurement by definition necessitates deviation, for one cannot measure what is defined fully within itself (much more so in the sense of existence). Thus to conclude for *Good*, or at least *Not Bad*, requires necessarily that *Bad* also be present – a primary starting point for this study was that harm is inherently present when existence endures. For this reason we are able to assess an existence’s worth, for no existence is an ever-accumulating series of joys and pleasures (quite the contrary, as Benatar argues).

When investigating the proposed technology-driven alternative existences we continue to notice harm; harm has mutated, taken on new names and forms, but it remains. We may therefore conclude that harm is inextricably linked with existence, as argued by means of Benatar’s RAP. Experiencing harm is *to be human*, and even in the technology-driven sense we find that experiencing harm is *to exist*. I would even proffer that this is okay, for harm is unavoidably part of our existence, or in other words we will have to be okay with it, if we wish to continue to exist as present-life individuals. Even if one appeals to alternatives of believing that harm is somehow tied to our being *human*, Benatar’s conclusion has taken on new devastating proportions: all existence is harm. In fact, *all* existence is *always* harm, and it is a *considerable* harm. To reformulate this study’s introduction in Section 1 for the preceding generalized view: existence is pain.

5.5 Chapter Conclusion

In this chapter I demonstrated to what extent technology-driven alternatives to existence fail to overcome Benatar’s goal of anti-natalism, of avoiding harm. The shortcomings of the alternative existences we considered here point to an inability to reduce harm, let alone to

⁶³ As a reminder of the guiding principle:

- i. It is good to reduce harm;
- ii. It is better to eliminate harm; but
- iii. It is best never to introduce harm at all.

completely avoid or eradicate its occurrence. On the contrary, as I have argued, we have seen that such alternative existences end up introducing new forms of harm to the individual's existence (whether those harms are modified forms of known harms or completely new versions of it).

In terms of an alternative existence based on algorithms, I have argued that humanity's influence is still clearly felt. Inherently human shortcomings are quickly translated to the purported algorithmic solutions to harm. We may, for example, recall the somewhat ominous warning by the data scientist, Guha, that no designer is neutral (US Catholic, 2020) and thus we find that human failure becomes embedded in algorithms' design.

I have also argued that an alternative existence of hive minds and cyborgs comes at great cost to the individual, or individuality. The individual and the other have more than united; the unity has made way for unison. By this, I mean that togetherness has been transcended to reach sameness. In this context, the value of privacy dissolves.

Finally, I have argued that the sacrifice of the phenomenological spirit that was detected in the aforementioned alternative existence of hive minds reaches its climax in especially the third technological alternative to existence found in artificial bodies and robots. I suggested that we would need to reject the very thing that defines our humanity in the hope of reducing harm, our human natures, whilst inadvertently introducing new harm by that very act.

In conclusion, I have argued that technology-driven alternatives to existence indeed do not provide any escape from harm because we can see that the shortcomings presented by such technological utopias may be too great, but also that said alternatives ultimately fail in reducing and avoiding harm.

6 Study Conclusion

In this study I have argued that there can be neither net reduction nor negation of harm in the hypothetical transitioning from our known organic human existence to an existence based on technology-driven alternatives.

I firstly investigated the nature of harm as identified in Benatar (Chapter 2). I reviewed Benatar's notion that we have different criteria for measurement in terms of assessing a life's worth depending on whether it is a life to be started (before birth – future-life) or a life to be continued (when it is already underway – present-life). Chapter 2 also reflected on humanity's overly positive view of life and its pleasures due to psychological states and their underlying biasedness. The conclusion of this literature review resulted in the description of the measurement tool of the RAP, which was subsequently used for assessing an existence's worth when it takes on alternative forms (i.e. technology-driven forms).

In Chapter 3, I briefly engaged with Benatar's detractors, particularly in terms of his anti-natalist perspective. As discussed in Chapter 3, Benatar's argumentation makes significant allowance for, indeed focuses on, impersonal evaluation of the worth of life in contrast to his detractors' focus on a specific person. Benatar, in contrast, argues that one cannot judge all the harms in life based on the criteria of a specific individual.

In Chapter 4 I investigated the existential alternatives which technology can provide, with a particular focus on the ways in which such perspectives purport to minimize or even overcome harm. Certain aspects of human life, such as bodily shortcomings, associated societal impacts, and interaction with other species, may be overcome with technology as potential saviour. This chapter showed that, at least initially, the prognosis of non-organic alternative existences appear promising.

In Chapter 5 I argued, that while technology is often considered a saviour (i.e. able to minimize or overcome harm), such technological developments may bring about new harms and potentially also aggravate already-existing human shortcomings (particularly the existential harms that Benatar identifies). Reviewing the shortcomings of technology by means of the RAP, I argue that we should not to seek salvation in technology (in fact, we should abandon our quasi-teleological relationship with technology) and instead see it for what it is: a

powerful, and often dangerous tool – not an end. We should therefore be wary of the teleological pitfalls of technology as saviour, particularly when it purports to minimize or overcome harm.

The current study may be expanded by means of the following avenues:

- Expanded engagement with the central topic of Benatar's *Better Never to Have Been* (2006), namely the proposal of 'soft extinction' – to stop the procreation of human life. The latter is proposed as a generalized method of executing anti-natalism. In my view, technology may assist in achieving such a soft extinction.
- A key conclusion of this study is that harm is an inherent part of existence. A further topic for study is the in-depth investigation of this relationship in more detail. Does one drive the other; in other words, does harm and existence exhibit correlation or causality? As noted in Footnote 59, the sense of harm and badness may be spontaneous or inherent in an individual without interacting with another party. I propose that there is fruitful ground for discussion with regards to this question.
- A further distinction to be made on the measurement of life, which is not present in Benatar (2006), is that of the separate measures that are the *quality* of life, and the *value* of life. Deeper investigation of the distinct topics and juxtaposition along the lines of Benatar's RAP may yield conflicting results, and bring the RAP itself into question.
- An aspect omitted in this study is how technologies can affect one's understanding of being in the world. It is proposed that thinkers such as Adorno (*The Culture Industry*) and Marcuse (*One-Dimensional Man*) be consulted as starting point for such further study.
- This study was focused on technology-driven alternatives to existence. However, there are extensive alternative technological postulations remaining for future study, which may broaden the applicability of this study.
- I also propose further investigation into the harm of existence as experienced by flora and non-human fauna – even bacteria and non-cellular life. Whether through

reincarnation as an apple tree or deeper philosophical reflection on ‘What is it like to be bat?’, there may be yet more forms of harm to identify, and hopefully, overcome.

As mentioned in Chapter 5, measurement necessitates deviations: to conclude for *Good*, or at least *Not Bad*, requires necessarily that *Bad* be present. Harm appears to be a type of complement to existence, which cannot be escaped nor overcome through technological development. The fates of harm and existence are tied. Even if we seem to identify harm as linked to some other aspects of life – such as the harms of manifestation in an organic body or societal discord – these are merely decoys of the true ‘culprit’ – to exist. Recognition of this harm seems a necessary requirement for doing the position of Benatar justice, and I would argue that we need to take his anti-natalist argument seriously. Where the mad king would yell ‘Kill them all’, the sane philosopher should plead ‘Kill it all’.

7 References

- ASIMOV, I. 2018. *The Complete Robot*. New York, NY: DoubleDay.
- BLASS, A. & GUREVICH, Y. 2003. Algorithms: a quest for absolute definitions. *Bulletin of European Association for Theoretical Computer Science*, 81.
- BEINSTRUP, T. & DANDANELL, F. 2019. Vestager klar med bøde på 11 mia. kr. til Google. *Berlingske.dk*, 20 March 2019. <https://www.berlingske.dk/virksomheder/vestager-klar-med-boede-paa-11-mia.-kr.-til-google> Date of access: 26 May 2021.
- BELSHAW, C. 2012. A new argument for anti-natalism. *South African Journal of Philosophy*, 31(11).
- BENATAR, D. 1997. Why it is better never to come into existence. *American Philosophical Quarterly*, 34(11).
- BENATAR, D. 2006. *Better never to have been: the harm of coming into existence*. New York, NY: Oxford University Press.
- BENATAR, D. 2012a. Every conceivable harm: a further defence of anti-natalism. *South African Journal of Philosophy*, 31(37).
- BENATAR, D. 2012b. *The second sexism: discrimination against men and boys*. Chicester, UK: John Wiley.
- BERGSTRØM, I. I. 2019. Han, hun eller hen? Brug af ordet 'hen' resulterede i mere inkluderende holdninger. *Videnskab.dk*, 21 Sep. 2019. <https://videnskab.dk/kultur-samfund/han-hun-eller-hen-brug-af-ordet-hen-resulterede-i-mere-inkluderende-holdninger> Date of access: 9 Nov. 2020.
- BOLEA, Ș. 2019. Antihumanism in the works of E.M. Cioran and Thomas Bernhard. *Philobiblon: Transylvanian Journal of Multidisciplinary Research in Humanities*, 24(1).
- BOONIN, D. 2012. Better to be. *South African Journal of Philosophy*, 31(16).
- BOSTROM, N. & YUDKOWSKY, E. 2014. The ethics of artificial intelligence. In: FRANKISH, K. & RAMSEY, W.M., eds. *The Cambridge handbook of artificial intelligence*. UK: Cambridge University Press.
- BOUFFIOUX, M. 2020. Calvin Soiresse : « Il est urgent de procéder à une réconciliation des mémoires ». *Paris Match Belgique*, 20 Jun. <https://parismatch.be/actualites/societe/408768/kalvin-soiresse-il-est-urgent-de-proceder-a-une-reconciliation-des-memoires> Date of access: 7 Nov. 2020.

- BRAIDOTTI, R. 2013. *The posthuman*. Cambridge, UK: Polity Press.
- CARTER, M. 2007. *Minds and computers: an introduction to the philosophy of artificial intelligence*. Edinburgh, UK: Edinburgh University Press.
- CHALMERS, D.J. 1995. The puzzle of conscious experience. *Scientific American*, 273(6).
- CLOUSER, R. A. 2005. *The myth of religious neutrality: an essay on the hidden role of religious belief in theories*. Notre Dame, IN: University of Notre Dame Press.
- DANKS, D. 2014. Learning. In: FRANKISH, K. & RAMSEY, W.M., eds. *The Cambridge handbook of artificial intelligence*. UK: Cambridge University Press.
- DREYFUS, H. L. 1992. *What computers still can't do: a critique of artificial reason*. Cambridge, MA: MIT Press.
- EDEN, A.H., STEINHART, E., PEARCE, D. & MOOR, J.H. 2012. Singularity hypotheses: an overview. In: EDEN, A.H., MOOR, J.H., SØRAKER, J.H. & STEINHART, E., eds. *Singularity hypotheses: a scientific and philosophical assessment*. Available from SpringerLink e-books: <https://link.springer.com/book/10.1007/978-3-642-32560-1> Date of access: 26 May 2021. pp. 1-12.
- ELLUL, J. 1983. The technological order. In STOVER, C.F., ed. *The technological order*, MI: Wayne State University Press.
- FERRIS, J. 2012. Self-awareness with a simple brain. *Scientific American Mind*, 23(5)
- FLAXMAN, S., GOEL, S. & RAO, J.M. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80.
- FLORIDI, L. 2014. *The 4th Revolution*. New York, NY: Oxford University Press.
- GILLANI, N., YUAN, A., SAVESKI, M., VOSOUGHI, S. & ROY, D. 2018. Me, my echo chamber, and I: introspection on social media polarization. 2018 World Wide Web Conference. Lyon, FR.
- HARAWAY, D. J. 2008. *When species meet*. Minneapolis, MN: University of Minnesota Press.
- HARAWAY, D. J. 2016. *A Cyborg Manifesto*. Minneapolis, MN: University of Minnesota Press.
- HARMAN, E. 2009. Critical study David Benatar. Better never to have been: the harm of coming into existence. *NOÛS*, 43(10).
- HEIKKILÄ, M. & BURCHARD, H. V. D. 2020. 'Frugal four' propose 'loans for loans' approach to coronavirus recovery fund. *Politico*, 23 May. <https://www.politico.eu/article/frugal-four-propose-loans-for-loans-approach-to-coronavirus-recovery-fund/> Date of access: 8 Nov. 2020.

- IHDE, D. 1975. The experience of technology: human-machine relations. *Cultural Hermeneutics*, 2(3).
- JACQUETTE, D. 2000. Schopenhauer on the ethics of suicide. *Continental Philosophy Review*, 33(1).
- JAMIESON, K.H. 2021. How to counter Covid misinformation. *Scientific American*, 324(4).
- JAPANLOVER. 2014, 15 Apr. *BEGIN Japanology ~ Vending Machines*. [YouTube video]. <https://www.youtube.com/watch?v=icmy8FaaHHM> Date of access: 15 Aug. 2019.
- JOHRI, N. 2020. How lab animals have fared in the coronavirus crisis. *DW Science*, 1 Jul. <https://p.dw.com/p/3eaT9> Date of access: 25 Nov. 2020.
- JUST, N. & LATZER, M. 2017. Governance by algorithms: reality construction by algorithmic selection on the Internet. *Media, culture & society*, 39(11).
- KING, T. C., AGGARWAL, N., TADDEO, M. & FLORIDI, L. 2019. Artificial intelligence crime: an interdisciplinary analysis of foreseeable threats and solutions. *Science and Engineering Ethics*, 26.
- KROEKER, R. 1997. *Star Trek: Voyager*. United States: Paramount Television.
- LATZER, M., HOLLNBUCHER, K., JUST, N. & SAURWEIN, F. 2014. *The economics of algorithmic selection on the Internet*.
- LETZING, J. 2020. The super-wealthy won big as COVID-19 spread - here's how they can help alleviate the economic peril faced by most. *World Economic Forum COVID Action Platform*, 28 Oct. <https://www.weforum.org/agenda/2020/10/the-rich-got-richer-during-the-pandemic-and-that-s-a-daunting-sign-for-our-recovery/> Date of access: 8 Nov. 2020.
- LONG NOW FOUNDATION. 2020, 7 Mar. *Gods and robots: ancient dreams of technology*. [YouTube video]. <https://www.youtube.com/watch?v=czi-7G6JzbQ&t=2362s> Date of access: 12 Nov. 2020.
- MORIOKA, M. 2021. What is Antinatalism? *The Review of Life Studies*, 12.
- MCEWAN, I. 1994. *The cement garden*. New York, NY: Vintage Books
- MÜLLER, V. C. 2012. Introduction: philosophy and theory of artificial intelligence. *Minds & Machines*, 22(3).
- MÜLLER, V. C. 2014. Risks of general artificial intelligence. *Journal of Experimental & Theoretical Artificial Intelligence*, 26(6).

- NILSSON, N.J. 1998. *Artificial intelligence: a new synthesis*. San Francisco, CA: Morgan Kaufmann Publishers.
- NOBLE, D. F. 1997. *The religion of technology*. New York: NY, Alfred A. Knopf.
- OKUDA, M. & OKUDA, D. 1999. *The Star Trek encyclopedia: a reference guide to the future*. New York, NY: Pocket Books.
- OMOHUNDRO, S. 2014. Autonomous technology and the greater human good. *Journal of Experimental & Theoretical Artificial Intelligence*, 26(14).
- PHIRI, A. & WRIGHT, M. M. 2020. Black Lives Matter but slavery isn't our only narrative. *The Conversation*, 13 Sep. <https://theconversation.com/black-lives-matter-but-slavery-isnt-our-only-narrative-137016> Date of access: 7 Nov. 2020.
- REINER, R. 1987. *The princess bride*. United States: Twentieth Century Fox.
- RICE, C. M. 2013. Defending the objective list theory of well-being. *Ratio*, 36(17).
- RUTI, M. 2018. *Penis envy and other bad feelings: the emotional costs of everyday life*. New York, NY: Columbia University Press.
- SHULTZ, S., OPIE, C. & ATKINSON Q.D. 2011. Stepwise evolution of stable sociality in primates. *Nature*, 479.
- SIROIS, J. 2020. The singularity of the human hive mind. *Philosophy Now*, 139.
- SORGNER, S.L. 2009. Nietzsche, the overhuman, and transhumanism. *Journal of Evolution & Technology*, 20(1).
- STEDMAN, C. 2019. Data ethics issues create minefields for analytics teams. In: STEDMAN, C., ed. *Ethical data mining and analytics elude privacy, usage snafus*. <https://searchbusinessanalytics.techtarget.com/news/252467829/Data-ethics-issues-create-minefields-for-analytics-teams> Date of access: 3 Mar. 2020.
- SULLINS, J. P. 2011. Introduction: open questions in roboethics. *Philosophy & Technology*, 24(6).
- SUNSTEIN, C. R. 2002. The law of group polarization. *The Journal of Political Philosophy*, 10(21).
- SVENDSEN, A. B. 2020. Pia Kjærsgaard tordner efter stigning i coronasmitte blandt indvandrere: 'Luk ghettoer'. *dr.dk*, 5 Aug. <https://www.dr.dk/nyheder/politik/pia-kjaersgaard-tordner-efter-stigning-i-coronasmitte-blandt-indvandrere-luk> Date of access: 15 Sep. 2020.

- TILES, M. & OBERDIEK, H. 2014. Conflicting vision of technology. In: SCHARFF, R.C. & DUSEK, V., eds. *Philosophy of technology: The technological condition and anthology*. UK: John Wiley & Sons
- TORTORELLA, N. 2019. *Space between: explorations of love, sex and fluidity*. New York, NY: Crown Publishing
- US CATHOLIC. 2020. AI and ethics aren't mutually exclusive, says this data scientist. *US Catholic Interview*, 4 Feb. <https://uscatholic.org/articles/202002/ai-and-ethics-arent-mutually-exclusive-says-this-data-scientist/> Date of access: 3 May 2020.
- VAN SCHAİK, C.P. 1983. Why are diurnal primate living in groups? *Behaviour*, 89.
- VERBEEK, P.-P. 2005. *What things do: philosophical reflections on technology, agency, and design*. Pennsylvania, PA: Pennsylvania State University Press.
- WARD, A. & WEBSTER, M. 2016. *Sociality: the behaviour of group-living animals*. CH: Springer International Publishing.
- WIEGMAN, M. 2020. Coronapatiënten op ic's vooral van niet-westerse achtergrond: 'De achterdocht is groot'. *Het Parool*, 12 Oct. <https://www.parool.nl/nederland/coronapatiënten-op-ic-s-vooral-van-niet-westerse-achtergrond-de-achterdocht-is-groot~b5ff4cdc/?referrer=https%3A%2F%2Fwww.google.co.za%2F> Date of access: 2 Nov. 2020.
- WILSON, R. L. 2017. *Techno-salvation: developing a Christian hermeneutic of enhancement technology*. Pittsburgh, PA: Duquesne University. (Thesis – PhD).
- ZEDNIK, C. 2018. Will machine learning yield machine intelligence? In: MÜLLER, V.C., ed. *Philosophy and theory of artificial intelligence 2017*. CH: Springer Nature.
- ZUBOFF, S. 2015. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1).