

Are rational suicides possible?

And if so, what does this mean for the suicide-*nullvisjon*?

Kan sjølvord vere rasjonelle?

Og i så fall, kva vil det seie for nullvisjonen for sjølvord?

Eirik Hommedal

Masteroppgåve i FILO350

Institutt for filosofi og førstesemesterstudier

Universitetet i Bergen

Hausten 2023

Rettleiar: Espen Gamlund

Table of Contents

1. Introduction and visions of a world without suicide.....	5
2. Defining suicide.....	7
2.1. Outlining an intention-based account of suicide.....	8
2.2. The notion of autonomy.....	11
2.3. My proposal for an intention-based definition.....	15
2.4. Chapter conclusion.....	15
3. The value and disvalue of death for the decedent.....	16
3.1. The intuition that death is harmful to the decedent.....	16
3.2. The Epicurean view on death.....	17
3.3. Outlining some responses to the Epicurean arguments.....	19
3.4. The deprivationist account of the badness of death.....	22
3.5. The deprivationist account of the <i>goodness</i> of death.....	25
3.6. In defence of the neutral container theory.....	27
3.7. Who decides whether or not life is worth living.....	30
3.8. Chapter conclusion.....	32
4. Rational suicides.....	33
4.1. To contemplate suicide is to make a choice between life-courses.....	33
4.2. Making ideally <i>right</i> choices between life-courses.....	35
4.3. Making <i>rational</i> choices between life-courses.....	38
4.4. Life-courses within life-courses.....	38
4.5. Choosing between life-courses in the face of uncertain futures.....	39
4.6. Rational suicide in the face of an uncertain future.....	42
4.7. The difficulty of evaluating the worthiness of one's moment-to-moment existence...	46
4.8. A few remarks.....	47
4.9. Chapter conclusion.....	50
5. Mental illness and irrational states of mind.....	51
5.1. The empirical correlation between suicide and mental illness.....	52
5.2. Does the presence of mental illness necessarily preclude the capacity to commit suicide rationally?.....	55
5.3. Does living a life not worth living necessarily preclude the capacity to commit suicide rationally?.....	58
5.4. Chapter conclusion.....	61
6. Conclusion and returning to the suicide- <i>nullvisjon</i>	62
References.....	63

Abstract

In 2020, the Norwegian government put forward the suicide-*nullvisjon*, a zero-tolerance policy for suicide deaths. Presumably, this presupposes that rational suicide is impossible. In this theses I will argue, by showing that rational suicides *are* in fact possible, that the normative ideal underpinning the suicide-*nullvisjon* is unjustified. More specifically, I argue that death *can* be a benefit for the decedent; that a person can, at least in principle, rationally choose to commit suicide even though she can't be *entirely* certain that she's not making a mistake; and that person's who find themselves in a position where suicide *would* make sense, at least sometimes, *do* possess the capacity to suicide rationally.

Samandrag

I 2020 fremja regjeringa ein nullvisjon for sjølv mord i Noreg. I denne avhandlinga argumenterer eg for at eit normativt ideal om *null* sjølv mord ikkje kan vere riktig, då rasjonelle sjølv mord faktisk er moglege. For å underbygge dette argumenterer eg for at einskildmenneske stundom *kan* vere tent med å døy; at ein person som truleg *vil* vere tent med å døy, i det minste i prinsipp, kan velje rasonelt i å avslutte livet sjølv om han aldri kan vere heilt sikker på korleis livet ville arta seg om han ikkje gjorde det; og til sist, at einskildmenneske som er i slike høve at sjølv mord ville vore rasjonelt, *kan* vere kompetente til å ta avgjerdsla om å avslutte livet rasjonelt.

Acknowledgments

I would like to thank my supervisor Espen Gamlund, my girlfriend Kari, my family, and, most importantly, my cat.

“That Suicide may often be consistent with interest and with our duty to ourselves, no one can question, who allows that age, sickness, or misfortune, may render life a burden, and make it worse even than annihilation” (Hume 2005, 10).

1. Introduction and visions of a world without suicide

In 2002 The Storting (the Norwegian Parliament) approved the so called *nullvisjon* – a zero tolerance policy for deaths and serious injuries in Norwegian traffic. It was and is meant to serve as a sort of grand over-arching principle; an end according to which the layout of transport systems, vehicles and policies should align¹. In 2020 the government put forward a second *nullvisjon*, this time in the context of suicide deaths (Helse- og omsorgsdepartementet 2020). As in the above case it's meant to function as a sort of final aim towards which policies and resource allocation should orient.

Let's make an attempt at unpacking the moral judgements that underpin the two policies. The traffic-*nullvisjon* explicitly states that *all deaths and serious injuries² in traffic are morally unacceptable³*. Since we're at the level of national legislature, it's reasonable, I believe, to interpret this to mean something like: deaths caused by traffic accidents are bad. As a society, we have a moral obligation to reduce the number of these events to zero.

Notice how the very idea of a traffic-*nullvisjon* begs a question with a preconceived answer: are all deaths caused by traffic accidents bad and to be avoided? Yes. Notice how this moral judgement seems to follow trivially from the very concept of a zero-tolerance policy. If we restate the traffic-*nullvisjon* in a few more words this should become clearer — let's say something akin to “*all deaths caused by traffic accidents should be prevented/eliminated*”. This *normative* statement clearly contains a moral judgement along the lines of “all instances of death and serious injuries in traffic are bad”. This, you'll probably agree, seems very reasonable. A world in which *one* person dies in traffic does indeed seem to be worse than a world in which no-one does, all else being equal.

While the moral foundation of the suicide-*nullvisjon* is not as explicitly worded as in the case of the traffic-*nullvisjon*, we might garner some insights by unpacking its normative contents in a manner similar to what we did above. The two concepts are, after all, similar in structure: the phenomenon of *x ought* to be eliminated. What judgements, then, are implicit in a normative statement like “*all suicides should be prevented/eliminated*”? Certainly, I think, that all suicides are, just like traffic deaths, bad and undesirable⁴.

1 <https://www.vegvesen.no/nn/fag/fokusomrade/trafikksikkerhet/nullvisjonen/> [accessed 29.08.23]

2 The part about “serious injuries” is not all that relevant to our coming discussion; our focus will be on death. For simplicity's sake I will not include it moving forward.

3 <https://www.vegvesen.no/nn/fag/fokusomrade/trafikksikkerhet/nullvisjonen/> [accessed 29.08.23]

4 This assessment is, I believe, echoed in some of the suicide-*nullvisjon*'s formulations. Sentences like ‘every suicide is a tragedy’ and ‘we can't afford to lose anyone’ are examples of this.

Exactly *why* it is the case that all traffic deaths are bad and to be eliminated is not explicitly stated, but I think it's reasonable to assume that the lawmakers had in mind the harm that is inflicted on those who fall victim to them. Here, two categories seem most relevant. First is the decedent herself (i.e., the person whose life is cut short), and second, those direly affected by the decedent's death (i.e., the decedent's family, friends and so on). The motivation must, I think, fall principally on the first.

Notice next this important difference between traffic deaths and suicides. Deaths caused by traffic accidents are, well, accidents; they're *unintended* events⁵. You can't intend to get killed in a traffic accident; if you did, it presumably wouldn't have been an accident. Taking one's own life, however, is something one can intend to do; in fact, as we see in the following chapter, intentions play an essential part in defining the notion of suicide. Because suicides involve an axis of intent that traffic deaths do not, the former begs the question: are suicides always irrational acts? This question, it seems to me, is something a *suicide-nullvisjon* must, at the very least, take into consideration and account for. If the answer is yes, it would endorse the normative ideal of suicide eradication. If the answer is no, there's, it seems to me, something fundamentally wrong with the normative ideal motivating the *nullvisjon*, or, at the very least, other justifications for the ideal are in order.

The central question of this master thesis, then, is whether it's *possible* to act rationally in taking one's own life. To answer this question, we'll, for starters, need to get clear on what suicides actually are (chapter 2). Following this, we move on to discuss the value and disvalue of death (chapter 3). Having concluded that a person can, in fact, be benefited by her own death, we move on to the question of what it means to commit suicide from the first-person perspective and investigate under what conditions suicide might plausibly make sense from an idealized point of view (chapter 4). Following this, we consider whether mental illness or potentially irrational states of mind *necessarily* preclude a person's ability to commit suicide rationally (chapter 5). Finally, we'll return to the *suicide-nullvisjon*, armed with the knowledge to judge it disapprovingly (chapter 6).

5 Traffic deaths that *are* intended would surely fall into the categories of murder or suicide, not traffic accidents.

2. Defining suicide

Before we turn to the question of whether suicide acts could be rational acts, it seems wise to establish a general understanding of what suicides and suicide acts actually are. On the face of it, defining these terms may seem like a trivial task. Surely, one might think, it's as easy as saying that *suicide* simply means *self-inflicted death* and that *suicide acts* simply means *acts that results in self-inflicted death*. In every-day language, after all, we might say of someone that 'he killed himself' and that 'he committed suicide' and have exactly the same thing in mind. When we look more closely, however, it will become clear that these terms aren't really synonymous at all, that is, it's entirely possible to have committed suicide without having killed oneself and it's equally possible to have killed oneself without having committed suicide. Regarding the last, consider, for example, smoking. Smoking cigarettes regularly and for an extended period of time has a decent chance of causing lung cancer or cardiovascular disease, which in turn has a decent chance of being fatal. A person who dies from a smoking-induced illness was, presumably, the cause of his own death — in the event he hadn't smoked he wouldn't have gotten a smoking-induced illness and if he hadn't gotten a smoking-induced illness he wouldn't have died when he did. We wouldn't, however, say of this person that they committed suicide. It seems strange to claim that smoking cigarettes, by itself, counts as a suicidal act. Similarly, an absent-minded bungee jumper who forgets to attach the elastic cord to his feet before leaping off a bridge presumably caused his own death — had he paid more attention and gone through the safety check list he would have noticed his mistake and therefore not died when he did. We wouldn't, however, say that his death was a suicide — something along the lines of 'his death was an accident' or 'the result of a fatal mistake' seems more appropriate. Simply put: you being the direct cause of your own death does not necessarily entail that you committed suicide, you might simply have engaged in risky or unhealthy behaviours that lead to you dying earlier than you otherwise would have. The opposite is also true, that is, it's possible to commit suicide *without* having been the direct cause of one's own death. Cases of assisted dying and 'suicide by cop' are examples of this (Cholbi 2011, par. 5). In these cases the person committing suicide merely arranges the conditions that directly causes their death. Nevertheless, it seems clear to me that such cases are suicides still.

Having killed oneself, it turns out, is neither sufficient nor necessary for having committed suicide, and so simply defining suicide as self-inflicted death or self-killing clearly won't do. How, then, might suicide be defined? Consider the following two scenarios: 1. Oliver, a generally content and happy to be alive daredevil, is walking along the platform of a

train station. Suddenly he jumps down onto the tracks in front of an oncoming train, expecting the train driver to slam the breaks and for the train to come to a screeching halt mere inches from his body, forever cementing his legacy as the bravest daredevil in town. Unfortunately for Oliver, however, it quickly becomes clear that he grievously misjudged the stop distance of the train. He dies on impact. 2. Olivia, generally discontent and unhappy with being alive, is walking along the platform of a train station. All of a sudden she *purposefully* throws herself in front of an oncoming train *intending* to cause her own death. As in Oliver's case, the driver slams the breaks, but stands no chance in stopping the train in time. Olivia dies on impact. Though both Oliver's and Olivia's deaths are self-inflicted, clearly only Olivia's death counts as a suicide, and clearly, the important fundamental difference between the two cases comes down to the presence of *intent* and *desire* for death in the second, something which is totally absent in the first.

2.1. Outlining an intention-based account of suicide

An account that emphasises desire and intention rather than outcomes, then, might provide a more compelling conceptual understanding of suicide acts and suicide deaths. Fairbairn, among others⁶, makes the case for such a view. He writes:

Suicide is an act, whether of commission or omission, and whether preformed by himself or others, by means of which an individual autonomously intends and wishes to bring about his death because he wants to be dead or wants to die the death he enacts (Fairbairn 1995, 82).

Continuing that:

A person is a suicide [⁷] if he initiates an act, whether of commission or omission, and whether preformed by himself or others, by means of which he autonomously intends and wishes to bring about his death because he wants to be dead, or wants to die the death he enacts, as long as he entertains that wish and intention (Fairbairn 1995, 82).

6 I will, going forward, lean into Fairbairn's account, investigate its intricacies and challenge one of its notions. I should note, though, that a number of other intention-based models have been proposed by others (Paterson 2003, Hill 2011).

7 This is not a misspelling but deliberate terminology on the part of Fairbairn: "I want to suggest that all people who wish to die and who intentionally act so as to achieve their deaths, should be referred to as *suicides*, whether they live or die" (Fairbairn 1995, 56).

According to Fairbairn then, a person's death was a suicide *iff* the person acted autonomously so as to fulfil or consummate their autonomous intention and wish to die and did not relinquish said intention and wish before death occurred. This seems to fit well with our above intuition that, even though Oliver and Olivia both inflicted death on themselves, only Olivia's death was a suicide, and only Olivia's act was suicidal. Likewise, our chain-smoker and forgetful bungee jumper from earlier would not qualify as suicides on this current model because their actions were aimed not at achieving death, but rather at enjoying the soothing of nicotine and the excitement of a bungee jump. I think Fairbairn's account of suicide gets a lot of things right so I'll spend a few pages attempting to lay out the details of it below. That being said, it does make one significant error which I'll try to amend before concluding this sub-chapter with my own attempt at defining suicide.

One of the strengths of Fairbairn's definition is that no distinction is made between cases where the act that ultimately causes a suicider's death is carried out by the suicider herself or by some other person, so long as the act was *initiated* by the suicider. This, in other words, allows for a kind of indirect causality between a suicider and their death, thus accounting for suicides that don't involve self-killing *per se*. I'm referring here to cases like assisted dying and 'suicide by cop' as mentioned earlier.

Furthermore, whether a suicide-act is one of commission (that is, active or positive e.g. something like pulling a trigger or swallowing hemlock) or omission (that is, passive or negative e.g. something like intentionally *not* ducking to cover once you, upon hearing a series of loud bangs, realise that you have mistakenly wandered into a firing range (Fairbairn 1995, 82)) is of no matter; it's a suicide-act all the same. This seems right to me. If a person finds herself in a dangerous situation that will cause her death if she refrains from taking active measures to prevent it from doing so, it seems that, in the event that she dies, her death was a suicide so long as she held the intention and wish that the dangerous situation should result in her death. We might imagine, for example, that Olivia from above was at the train station on her way back home after having bought a length of rope she intended to hang herself with later that same day. While walking along the platform she suddenly tripped in her untied shoelaces causing her to fall onto the tracks in the way of an oncoming freight train. The train was chanceless in stopping, but nevertheless is still far enough away that Olivia would have had sufficient time to get back on her feet and jump out of its way. Instead of doing so, however Olivia stays where she, is thinking to herself "hey, I was intending to kill myself later today anyway, so I might as well save myself the trouble and let myself get hit by this train in-

stead". Though Olivia merely refrained from saving herself from getting hit by the train, I think it's clear that she committed suicide. Presumably then, so long as their intended outcome was their death, the same would be true for people who refuse medical treatment for somatic illnesses that otherwise might be cured or kept in check, instead opting to let nature take its course.

Although Fairbairn's intention-based definition of suicide seems to fit well with my, and hopefully also your, intuitive sense of what kinds of deaths should count as suicides, it comes with its own challenges and, let's say, peculiarities. First off it must be acknowledged that, because intentions and wishes are internal qualities only directly accessible to the person who experiences them, we can never be totally sure whether any given death was a suicide or not. In this sense, whether or not particular deaths were suicide deaths will remain epistemically murky. Of course, a person can, through for example a suicide letter, express her intentions and wishes to others. Also, the way in which someone dies might, at least to some extent, indicate intent. If a person dies as a result of overdosing on heroin, for example, it's going to be difficult to guess whether she died intentionally or by accident. If the way in which a person died is less likely to have happened by accident and in the absence of intent, let's say the person died as the result of a self-inflicted gun shot penetrating from the roof of the mouth through the top of his head, we might be more justified in guessing or assuming that their death was intended. Even in the presence of suicide letters or ways of death strongly suggestive of intent, however, one can never be sure if the person remained intending and wishing to die all the way to the end. A bridge jumper might have wished and intended to cause their own death at the point of leaping but had a change of heart before hitting the water. Such a death would not be a suicide on our current view (Fairbairn 1995, 59). This epistemic uncertainty about suicide deaths has obvious ramifications in the context of empirical study and statistics on suicide, e.g., counting all bridge jumpers as suicides even though they might have changed their mind before hitting the water (we'll return to this in chapter 5).

Another facet of Fairbairn's account is, and this may at first seem a bit peculiar, in so far as his definition of suicide only involves acts and intentions with no reference to outcomes, if a person initiates an act they *believe* to be fatal, when it is actually far from it, they have still, in a sense, committed suicide (Fairbairn 1995, 57), though unsuccessfully (Fairbairn 1995, 58). Consider this example: A practitioner of homeopathic medicine is fed up with life and decides to put an end to his existence. He procures a lethal chemical which he proceeds to dilute in purified water. He repeats this process until not a single molecule of the

lethal chemical remains in the solution, falsely believing that the diluent will somehow remember the deadly effects of the original chemical. Convinced that the vile of purified water in his hand will cause his death if consumed, our man proceeds to gulp it down before going to sleep, fully expecting, wishing and intending to never wake up. The next morning of course, he does. Was our unhappy homeopath's act a suicidal one? Did he commit suicide? Well, according to Fairbairn's definitions both these questions should be answered affirmatively. This may seem puzzling: all he actually *did* do was drink a glass of water and normally drinking glasses of water is not something we'd consider to be suicidal behaviour. Let's contrast this example to another failed suicide attempt where the individual's beliefs about the lethality of his act is more reasonable. A man wishing and intending to die loads a gun, puts the barrel to his temple and pulls the trigger (obviously believing that this will result in his death) whereupon the gun jams and nothing happens. Is this case different from the previous in a meaningful way? It seems to me like it's not. What matters, it seems, is that in both cases the person *believed* and intended that their act should result in their death. In short: A person's act being a suicide-act does not depend on the probability of said act being lethal, but on the persons beliefs about its lethality combined with the wish and intention that said behaviour should cause her death. Would the second case be different had the gun not misfired? Outcome wise it obviously would have, but *for* that person pulling the trigger *in* than moment, they seem to me indistinguishable. For this reason Fairbairn thinks that, according to the aspects that matters on our intention-based view, successful and unsuccessful suicides are, in the sense that matters, the same thing (Fairbairn 1995, 58). Remember: according to Fairbairn suicide being suicide does not depend on the outcomes of suicidal acts, but on a person's wish and intention to die and taking the, in her mind, sufficient steps to fulfil her intention. So long as the intention to die is present and the person acts so as to fulfil said intention, a person is, in a Fairbairn's language, 'a suicide'. While I find this novel way of talking about suicide compelling, it won't be all that necessary for our upcoming discussions. For this reason I'll, going forward, stick to the more familiar language where 'suicide' refers only to successful and not failed suicide acts.

2.2. The notion of autonomy

An aspect of Fairbairn's view that is, to my thinking, underdeveloped and in need of further clarification is the notion of autonomy as it relates to suicide. On this subject, Fairbairn writes:

A person who is non-rational is for that reason non-autonomous and [...] although a person who is not autonomous can take his life, his death was not a suicide. [...] A person can be so irrational that he is non-autonomous; however, he may act irrationally while maintaining his autonomy. Rationality is, after all, a matter of degree [...] (Fairbairn 1995, 70-71).

Continuing that:

[...] to suicide a person must at least be able to imagine what it might mean to be dead. If he does not have sufficient understanding of life and death to be able to do this, he cannot suicide though he could self kill. Nor can a person suicide if his rationality is for a time so impaired by factors such as physiological distress, intense pain, lack of sleep, the effects of drugs, or the range of conditions commonly referred to as 'mental illness' that he is incapable of the necessary level of reflection (Fairbairn 1995, 72).

According to Fairbairn, then, the capacity for some level of rational thought is a necessary precondition for a person to be able to autonomously intend to die (which again is a necessary precondition for committing suicide). More specifically a person must have an adequate understanding of life and death so as to be able to imagine what it might mean for him to be dead, for he cannot intend to be dead if he does not know what being dead would entail for him (Fairbairn 1995, 71). If a person (perhaps a child or someone seriously mentally disabled) self-kills while not being able to comprehend the most basic consequences of his self-killing, for example that once he's dead he'll remain so forever, his death would not be a suicide according to Fairbairn (1995, 71). Furthermore, an otherwise adequately rational person might have their capacity for rational thought impaired by things such as physiological distress, intense pain, drugs or mental illnesses, perhaps to such a degree to where they are no longer capable of understanding what non-existence is like, and therefore no longer able to intend become non-existent.

This, I think, showcases an important flaw in Fairbairn's account of suicide. To hold that a person must be able to *intend to become non-existent* in order to have committed suicide would place a lot of deaths we'd normally think of as suicides into the 'self-killing but not suicide' bucket. More generally, this line of thinking seems to gravely misunderstand what I'd imagine to be many suicider's motivation for wanting to die. Surely, I would think, most (though perhaps not all) people who commit suicide do so not wishing and intending to *become* dead or to *die the death he enacts* per se, but does so in order to *cease existing* as a

means to avoid or be relieved of some aspect(s) of existence they find to be in some way seriously disagreeable⁸. They don't wish to *be dead*, they wish to *no longer be alive* and there's a big difference between the two (in the first, death is an end in-and-of-itself, whereas in the second death is an instrumental, that is, a means towards the end of avoiding suffering). I'll suggest, then, that a person need *not* be sufficiently rational so as to be able to ponder the metaphysics of what it will mean for him to be in a state of non-existence to count as a suicide if he self-kills. So long as he is sufficiently rational so as to understand that his death will put an end to his suffering, I think his death will be a suicide albeit, of course, not necessarily a rational one. To reiterate, I'm not disputing Fairbairn's idea that a person must be, to some extent, autonomous to commit suicide and that autonomy presupposes some level of capacity for rational thought. What I am disagreeing with is on what questions he must be able to reflect. Consider this analogy: imagine that being very hungry puts one in a state of mental distress and tiredness, which in turn impairs one's capacity for rational thought. Furthermore, the hungrier one gets the more impaired one's rationality becomes. Being terribly and ravenously hungry, we might imagine, involves such an overwhelming sense of exhaustion, anguish and desperation that rational thought becomes close to impossible, perhaps even to such an extent that a person might no longer be able to think clearly about what it would 'mean' for him to become satiated. Without this knowledge, Fairbairn seems to suggest, he wouldn't be able to intend to become satiated and if *eating* requires intent in a similar way to committing suicide (which it seems like it does) he couldn't *eat*, though he could put food in his mouth, chew and swallow. While this may be true, I think there's no need for the famished man to intend to become satiated in order to eat, all he needs to be able to intend is to be rid of his terrible hunger.

There is, however, at least one additional and perhaps more salient way in which a person's wish and intention to die may be considered non-autonomous, this time focusing not on a person's capacity for rational thought, but on *where* his intentions come from. If a desire and intention to self-kill can be born from a chemically imbalanced or otherwise malfunctioning brain⁹ it seems, I think, most reasonable to attribute the intention to the malfunctioning brain and not to the person who possesses the brain. If a person acts on said intention she would not, I think, be acting autonomously, or at least not entirely so. Consider this analogy: Anna is moving her body in certain ways, shifting through a series of dance-like poses *with* the intention of dancing. Clearly, *Anna is dancing*: she's both intending to dance and acting out her in-

⁸ This is also pointed out by Critchley (2021, 79).

⁹ By some accounts commonly thought to be a cause of mental illnesses which in turn is often thought to be linked to suicides.

tention autonomously. Bert is suffering from some sort of strange seizure or convulsion causing his body to move in certain ways, shifting through a series of dance-like poses *without* the intention of dancing. Although Bert is moving in a dancelike manner he's not doing so autonomously and neither is he intending to dance. Clearly, Bert is *not* dancing, although he may look like it to passers-by. Carl is in a near vegetative state with no ability to move any part of his body voluntarily with the exception of his eyelids. Carl is listening to top-40 radio in the company of his caretaker when his favourite song comes on making him want to dance. Carl signals his intention to his caretaker through a complicated series of winks upon which the caretaker picks up Carl's arms and starts moving them in certain ways, shifting through a series of dance-like poses. Is Carl dancing? Well, although he merely initiated the dancing-act and does not directly perform it himself it seems to me like Carl is in fact dancing. If, on the other hand, Carl's blinks were meant to communicate his wish to have the volume of the radio turned up to eleven, a signal the caretaker mistook for the wanting-to-dance signal, Carl, it seems to me, would not be dancing. The act of moving one's body in a dance-like manner then, is neither sufficient nor necessary for dancing; the intention to dance, however, is. A person is dancing so long as they intend to dance *and* autonomously initiate their body moving in a dance-like manner, whether carried out by themselves or others. Daisy, like Anna, is moving her body in certain ways, shifting through a series of dance-like poses *with* the intention of dancing. Daisy, however, has unwillingly and unknowingly had a computer-chip implanted in her brain by a deranged neurosurgeon. The brain surgeon's chip works by sending a series of electrical pulses through Daisy's brain, perfectly tuned to alter her brain state so as to induce in Daisy the desire and intention to dance. The chip is not capable of producing a *dancing* brain-state, only a *wanting and intending to dance* brain-state. So while she is free to decide whether or not to act on the intention produced by the brain-chip, the intention itself is not hers. Had the brain surgeon not activated the chip when he did, Daisy would not have wished and intended to dance when she did, and so, in the event that she acts on said intention, would not have moved in a dance-like manner when she did. Is Daisy dancing? To me it seems like she is not, because the intention prompting her to do so is, in a sense, not her own. She certainly thinks that she is intending to dance, and she is autonomously choosing to act on what she thinks is her intention to dance, but her intention is not autonomous, it's not *hers*. In some sense, it seems to me like it's the crazed brain-surgeon who is dancing, not Daisy. After all, the necessary element to dancing is the wish and intention to dance together with the initiation

of a dancing act that's performed by oneself *or* others (as in the case of Carl, the dancing paraplegic).

If a malfunctioning brain can, in a similar way to the computer-chip in Daisy's brain, induce in a person the desire and intention to self-kill, it seems to me that such a person would not, in the event they act on said intention, have committed suicide. Whether or not a desire or intention to self-kill originates from a malfunctioning brain is, we should note, very hard and perhaps impossible to know for sure. Even if we were able to say with confidence that a particular person's brain is malfunctioning in some serious way (perhaps the person starts acting or thinking in irrational and uncharacteristic ways), presumably, a person *could* have a malfunctioning brain *and* at the same time intend to take his life autonomously. This, I think, is a bullet we're just going to have to bite.

2.3. My proposal for an intention-based definition

My take on an intention-based definition of suicide acts and suicide deaths, then, will look something like this. A *suicide act* is an act by which a person, by commission or omission and by his own hand or that of another, autonomously intends to fulfil his autonomous wish to be dead *or* be relieved of his life. A person's death was a *suicide death* if he died as a result of his *suicide act*, and remained wishing and intending to die until he could wish and intend no more. If a person died as a result of his *suicide act* but had a last-minute change of heart, he would have died *not* wishing and intending to die or to be relieved of his life, and so, his death would have been a self-killing rather than a suicide.

2.4. Chapter conclusion

We started this chapter with the goal of coming up with a descriptive definition of suicide. It quickly became apparent that such a definition should not rely on the outcomes of acts, but on the intentions that motivate them. We then accounted for Fairbairn's intention-based view and pointed out one of its shortcomings; in suicides, death may also be construed as a mean to the relief from suffering and not only as an end in itself. We then widened the notion of autonomy and argued further that acts based on wishes and intentions to die or be relieved of life born from a malfunctioning brain shouldn't count as autonomous and thus not as suicides, but as self-killings.

3. The value and disvalue of death for the decedent

Before we explore the question of rational suicide, we need to lay down some groundwork surrounding death in general, and, more specifically, whether or not death can constitute a benefit for the one who dies. If it's *always* the case that death is a harm for the decedent, i.e., that dying is bad *categorically*, the notion of rational suicide is, it seems, dead in the water (in the sense that, if all squares have four sides *categorically*, it wouldn't make much sense to draw a square and expect it to have only three). If, however, it turns out that death *can* in fact, at least sometimes, be a benefit for the one who dies, the *possibility* for rational suicides remains open.

3.1. The intuition that death is harmful to the decedent

For starters we'll need to make a few distinctions. The term *death* is, as Solberg (2019, 91) points out, ambiguous in that it might capture different phenomena in different contexts. It might, for example, be used to refer to the *process of dying*, that is, the sometimes vague period of life (of existence) that occurs leading up to the *incident of death*. The incident of death, in turn, picks out the event where an existing person ceases to exist. For our purposes we're interested in the latter meaning. With this in mind, particular deaths might be thought of as having *indirect* and/or *direct* potential effects (Solberg 2019, 99). Death's indirect effects are experienced by those who remain living, that is, the decedent's family, friends, enemies, etc. Death's direct effects concern only the decedent herself (i.e., the person who ceases to exist). We're concerned here with the latter, that is, the *direct* value or disvalue of *the incident of death*.

We'll start with the common-sense intuition that it *feels* as if our own deaths are, in some way or another, bad or harmful to us. The immense effort expended by society to delay death come illness and old age indicate the breadth of this conviction. Merely thinking seriously about the inevitability of one's own death will fill most people, I should think, with a certain feeling of uneasiness, perhaps even dread or despair. The fact that I will one day cease to exist and that things will stay like that forever while the world continues without me for literal eternity is something I struggle to wrap my head around. Frankly, the thought of this makes me a little sad and uneasy; it's the ultimate case of fear of missing out, so to speak. But what is it about death that makes it bad? One possible answer might be that it's actually just the *process* of dying that makes death bad — dying might be a painful and scary experience, and hence intrinsically bad. It seems, however, there must be more to it than this, for even

though the thought of dying painlessly while asleep certainly seems preferable to the thought of dying in terrible agony, you would probably be quite distressed were if I told you that you'll go to sleep tonight and never wake up. Presumably then, there must be something other than the potentially painful process of dying itself that makes death bad for the person who dies. In short, it seems as if there has to be something about *ceasing to exist* that's bad. Pinning down exactly why it's bad to cease existing, however, is less trivial than one might expect. In the following few sub-chapters we will take a closer look at one attempt to account for the badness of painless death for the decedent, namely, *the deprivation account*. Before this, however, we'll need take a closer look at the view on death to which the deprivation account responds, that being *the Epicurean account*.

3.2. The Epicurean view on death

The ancient Greek philosopher Epicurus (341-270 BC) raised a puzzling challenge to the apparently immediate and obvious truth that our own deaths are bad for us:

So death, the most terrifying of ills, is nothing to us, since so long as we exist, death is not with us; but when death comes, then we do not exist. It does not then concern either the living or the dead, since for the former it is not, and the latter are no more (Epicurus 1983, 8).

Contemporary readings of Epicurus typically construe two main arguments from the above excerpt: *the experience argument* and *the time argument*. In addition, a third argument (attributed to Lucretius, another ancient philosopher inspired by Epicurus) in favour of the Epicurean position, *the symmetry argument* is often mentioned in the same breath as the two above. All three advance the same conclusion: death isn't bad for the one who dies.

According to *the experience argument*, (P1) for something to be good or bad for S it must be experienced as such (because all good and evil consists in sensation). Headaches are bad because the painful *experience* of having a headache is bad, eating chocolate pudding is good because the pleasurable *experience* of eating chocolate pudding is good¹⁰. Furthermore, (P2) experience presupposes existence¹¹. Something that doesn't exist can't experience headaches or chocolate pudding or, well, anything at all. Since death consists in the complete deprivation of all sensation (experience), then, (P3) nothing could possibly be bad for the dead. If nothing can be bad for the dead, neither can *being* dead, hence; (C) *death isn't bad for*

10 Solberg (2019, 93) calls this "the experience condition".

11 Feldman (1991, 205) calls this "the existence requirement".

the decedent (Solberg 2019, 93; Kagan 2012, 216; Nagel 1970, 75-76). Another way to phrase this might be: because nothing can be bad for the non-existent, neither can non-existence.

According to *the time argument*, (P1) for something to be bad for S, there has to be a particular time at which it is bad for S. However, this particular time cannot be *before S's incident of death*, since we can't be dead while we're still alive. Likewise, it can't be *after S's death*, that would violate the existence condition. It seems, then, as if (P2) there is no particular time at which S's death is bad for S. Hence, (C) *death isn't bad for the decedent* (Solberg 2019, 94, Kagan 2012, 213-214, Nagel 1970, 76, Johansson 2013, 256).

According to *the symmetry argument*: (P1) the period of non-existence proceeding S's existence was not bad for S. Furthermore, (P2) the periods of non-existence before and after S's existence are identical. Thus, (P3) the period of non-existence after S's death is not bad for S. Hence, (C) *death isn't bad for the decedent* (Solberg 2019, 96; Kagan 2012, 225; Nagel 1970, 76).

The Epicurean view on death, then, is that it is neither good nor bad; dying is of neither value nor disvalue. By extension it seems as if it's of no matter to us whether our lives be long or short, because our lives will, by necessity, last exactly as long as we may have an interest in living (Scarre 2007, 93). Our aforementioned fear of death, Epicurus proceeds to conclude, stems from the *illusion* of the badness of death, that is, our misguided belief that ceasing to exist is somehow harmful for us — when this illusion shatters, so too should our fear of death.

On the Epicurean picture, then, our own death isn't bad for us, and our aversion to our own death is irrational. What we're left with, it seems, is a detached and neutral attitude towards the timing our own deaths. This, perhaps you'll agree, is a peculiar idea. Imagine you're walking along the street when you notice something strange. There's some kind of shadow on the ground six feet ahead of you, and it seems to be growing larger. In the next moment a voice from across the street shrieks "watch out!". You look up to see a grand piano hurtling towards the ground. It was in the process of being hoisted up by crane to the uppermost suite of a luxury hotel when the cable unexpectedly failed. Should you continue walking into the path of the falling piano or turn around and leap to safety? Well, if the duration of our lives is of no concern to us, as Epicurus seems to insist, it's hard for me to imagine how one option might be preferable to the other. Whether your life ended there and then, pancaked under a grand piano, or whether you lived another ten, twenty or fifty years shouldn't really matter to you.

Interestingly, at least for me, a rational understanding of Epicurus's argument does little to diminish my intuition that dying tomorrow would be bad for me. I'll most certainly do my best to avoid falling pianos despite Epicurus's insistence that I would not be harmed if I were to get crushed by one (assuming, of course, that my death occurred instantly and painlessly). How should we understand this apparent miss-match between Epicurus's conclusion and my continuing aversion towards my own death? In Epicurus defence, this phenomenon might be explained through evolutionary terms. It certainly seems probable that an innate fear of death, though illusory and irrational, might become prevalent through the mechanisms of natural selection. That is, beings with an Epicurean *neutral attitude* towards the timing of their own death would probably get erased from the evolutionary landscape very quickly. It could also be the case that my reluctance to be crushed by the piano is really more about a fear of the potentially painful process of dying or perhaps a fear of surviving with some horrible and permanent injury. Alternatively, of course, it could simply be that Epicurus is mistaken somehow. Many contemporary philosophers take this position. Let's briefly look at some of what I gather to be the most important challenges raised against the Epicurean view. Note that this will by no means be an exhaustive overview, I merely hope to introduce some nuance of relevance to my own discussion about suicide.

3.3. Outlining some responses to the Epicurean arguments

Against what I've called *the experience argument*, objections have been raised against *the experience condition* (EC) in the first premise (for something to be good or bad for *S* it must be experienced as such). Thomas Nagel points to the badness of being betrayed or ridiculed behind one's back: "the natural view is that the discovery of betrayal makes us unhappy because it is bad to be betrayed — not that betrayal is bad because its discovery makes us unhappy" (Nagel 2012, 4). Another example might look like this: when Caesar (the family cat at my parents' house) eventually dies, the *fact that he's dead* is bad for me *before* my mother calls and makes me aware of the fact that he's dead. If my mother opted not to call and inform me of this terrible news in an effort to spare me, thinking that "what he doesn't know won't hurt him", I would be livid. Again, this seems to be because the reason I would feel sad when being made aware of the loss of my childhood companion is that it's bad to lose something you love, full stop — it's the bad state of affairs *itself* and not the *being made aware* of the bad state of affairs that's bad. In a similar vein, Solberg (2019, 93-94) points to medical conditions like undiagnosed diabetes or dysfunctional tumour suppressor genes as potential non-experi-

ential instrumental evils (that is, undiagnosed diabetes is a bad thing to have, even though you're not aware of having it). In fact, it seems as though not being informed of bad news is itself, at least sometimes, a non-experiential instrumental evil. If your doctor opted not to inform you that the blood test you gave last week came back positive for cancer leaving you three days left to live, your ignorance of your misfortune would be bad, at least in an instrumental sense; it would rob you of the opportunity to do important death-bed things like spending your last days with your loved ones, writing your will, apologizing to that kid you bullied in middle school, etc. In short, it seems as though the old adage "what you don't know, can't hurt you", at least in certain cases, just isn't true. If this is right, it seems we have reason to think that *the experience condition* is mistaken.

In an attempt to defend the Epicurean position against such counterexamples, Rosenbaum proposes a modified version of *the experience condition* to the one we advanced above. According to this weaker version "A state of affairs is bad for person *P* only if *P* can experience it at some time" (Rosenbaum 1986, 218). The requirement for being harmed (or benefited) by something is, in other words, shifted from *S having* the actual experience of that thing to *S having the possibility* of experiencing that thing at some time (at present or in the future). Let's call this version of EC, EC*¹². Now, according to Rosenbaum, the non-experiential harm of backbiting or dysfunctional tumour suppressor genes is bad precisely because the person being subjected to them *can* learn of or be effected by them in the future, however, "[...] a person born without a sense of smell cannot be causally affected by, and thus cannot experience, the stench of a smoldering cheroot" (1986, 219). This seems plausible. Rosenbaum concludes that "we can grant that what one *does not* consciously experience can hurt one without granting that what one *cannot* experience can hurt one" (1986, 221). If death is characterized by the eternal absence of any and all experience, a dead person, the Epicurean might say, couldn't possibly be causally affected by backbiting in the future, and thus, couldn't possibly be harmed by it. The central point here is that, while the non-experiential harms of backbiting or betrayal might disprove the strong version of EC, it poses no challenge for the weaker EC*.

In an effort to modify Nagel's original betrayal example so as to also target EC*, John M. Fischer imagines the following scenario:

12 J. M. Fischer conceptualizes the two versions of the experience requirement like this: ER I: "An individual can be harmed by something only if he has an unpleasant experience as a result of it (either directly or indirectly)" (Fischer 1997, 342) and ER II: "An individual can be harmed by something only if it is possible for him to have an unpleasant experience as a result of it (either directly or indirectly)" (Fischer 1997, 342).

You are betrayed behind your back by people who you thought were good friends, and you never actually find out about this or have any bad experiences as a result of the betrayal. But now suppose that these friends were (very) worried that you might find out about the betrayal. In order to guard against this possibility, they arrange for White to watch over you. His task is to prevent you ever from finding out about the betrayal. So, for example, if one of the individuals who betrayed you should decide to tell you about it, White can prevent him from doing succeeding. [...] I simply stipulate that White is in a position to thwart any attempt by you or your friends to inform you of what happened (Fischer 1997, 345).¹³

While I think Fischer's counterexample successfully disproves EC*, the Epicurean counter is obvious. Sure, the Epicurean might as well concur, undiscoverable betrayal might constitute a non-experiential harm for an *existing* person, but surely these kinds of harms still require an existing subject for which they are harmful *towards*. When a person no longer exists, however, such things couldn't possibly be harmful because there no longer exists any recipient of the harm (this follows from *the existence requirement*). In other words, both versions of the experience condition we've discussed seem to be superfluous to the Epicurean conclusion; there's no need for the Epicurean to refuse the possibility of non-experiential harms towards persons who *are* alive, what matters is the possibility for non-experiential harm towards people who *no longer exist*. Having dysfunctional tumour suppressor genes is obviously not something that could be bad for a person who doesn't exist. If I was struck dead by lightning an hour before my cat died, his death could not cause me any harm because there no longer exists any me to be harmed by his passing. A dead person, the Epicurean might say, could not be harmed by slander any more than a piece of gravel could be harmed by being dis-favourably compared to a gemstone.

We've arrived at what I take to be the greatest hurdle for the anti-epicurean: making sense of *when* death is bad for the decedent (in other words, refuting *the time argument/the existence condition*). Five main responses to this question have cropped up in the literature

13 A similar example might look like this: Imagine that you and I are vacationing in a far-away land where no-one speaks our language and we don't speak theirs. As a heinous and malicious prank, I write a note that says "I'm an idiot" in the local tongue using a one-of-a-kind dictionary in my possession and stick the note to your back. After having written the note, I burn the dictionary. While we're walking around town, the locals snicker whenever your back is turned. The locals are, however, of a spectacularly non-confrontational constitution and make sure to instantly turn their subdued laughter into polite smiles whenever you turn to face them. Blissfully unaware of the ridicule you're being subjected to, and with no way of understand the words on the piece of paper in the event you discovered it, say in a restroom mirror, it seems to me plausible that you *are* being harmed in some way even though you had no way of becoming aware of the fact that you were being ridiculed behind your back.

(Johansson 2013, 256): *atemporalism* (death is bad for the decedent, but not at any particular time); *eternalism* (death is bad for the decedent at all times); *priorism* (death is bad for the decedent while she's still alive); *concurrentism* (death is bad for the decedent only at the incident of death); and lastly *subsequentism* (death is bad for the decedent only after the incident of death). Atemporalism refutes P1 of *the time argument* (*for something to be bad for S, there has to be a particular time at which it is bad*), while the rest refutes P2 (*there is no time at which S's death is bad for S*) (Johansson 2013, 256). Out of these suggestions, something along the lines of *atemporalism* seems to me most promising. On this view, the disvalue of death can be accounted for in the same way as that of goods or bads one never had *in* life, that is, by *comparison* to counter-factual states of affairs. Although Rosenbaum is correct in saying that “[...] a person born without a sense of smell cannot be causally affected by, and thus cannot experience, the stench of a smoldering cheroot” (Rosenbaum 1986, 219), it, presumably, *could* have been the case that he was born with a functioning smelling apparatus. It *could* have been the case that he would have been able to enjoy the smell of roses or newly cut grass, but the contingent state of the nose he was born with is such that he's precluded from doing so. In this way, being born without a sense of smell is good or bad, not at any particular time and not because of any bad experience it causes him, but because it deprives him of smelling-experiences he otherwise *could* have had (McMahan 1988, 39, Nagel 1970, 74-75). This leads us into *the deprivation account* of the badness of death¹⁴.

3.4. The deprivationist account of the badness of death

In his seminal 1970 paper “Death” Thomas Nagel put forth a positive theory, often called *the deprivation account*, to show how, contrary to the Epicurean picture, our deaths *can* in fact be bad for us even though the *state* of being dead can't be bad intrinsically. The deprivation account is fundamentally in agreement with Epicurus on the two propositions that (a) when we have died we no longer exist and that (b) when we don't exist we can't experience being dead

¹⁴ Another reason for thinking that Epicurus might have gotten something wrong becomes apparent if we considered what his conclusion would entail in the following case. Assume for a moment that murder is wrong at least in part because the murder-victim is harmed or made worse off by being killed (there are, of course, other possible reasons to hold that murder is morally wrong. A Kantian, for example, might look to the categorical imperative). Now, if Epicurus is right in his claim that death is not harmful for the decedent, murders might not be as morally abhorrent as we usually think that they are. In fact, on the Epicurean picture, if murders are morally wrong solely because they harm people, murdering might only be wrong by virtue of the pain caused to the victim during the process of death and/or the pain caused to others, that is, the pain of loss experienced by the victim's family and friends. Given this, it seems, the instant and painless killing of a man known by no-one would not harm anyone, and so, presumably, could not be morally wrong. This severely unintuitive conclusion might further evidence the feeling that there's something very wrong with the Epicurean view on death.

(or anything at all). There is however a different sense in which our deaths might be bad for us according to Nagel; namely that our deaths rob us of life we otherwise *could* have lived, that is, we presumably *could* have died later than we in fact did, and thus *could* have lived a longer life than we in fact did. This rests on the fact that the timing of our death is contingent.

[...] what we find desirable in life are certain states, conditions or types of activity. It is *being* alive, *doing* certain things, having certain experiences, that we consider good. But if death is an evil, it is the *loss of life*, rather than the state of being dead, or nonexistent, or unconscious, that is objectionable (Nagel 1970, 74-75).

In other words, it's not that our deaths are bad for us because they put us in a state of eternal non-existence *per se* (the proposition Epicurus refutes), but rather that our death is bad for us because it *could* have been the case that it happened at a later point in time than it in fact did. If death occurs later rather than sooner, a person has the opportunity to experience more of what we, according to Nagel, find desirable in life, that is to say, "*being* alive, *doing* certain things and having certain experiences, that we consider good" (Nagel 1970, 74-75). If death occurs sooner rather than later, a person is deprived of experiencing all the desirable things that she otherwise would have experienced.

The deprivationist picture is, I think, best understood through the language of counterfactuals. Let's look at an example. Imagine a man, let's call him Bob, who lives a fulfilling and mostly happy life until he gets crushed by a falling piano on his way to work. He dies instantly at the age of 40. Now consider a counterfactual version of events where Bob, by sheer fate¹⁵, narrowly avoids being flattened by the piano and lives on happily to the age of 80. Let's call this version of Bob, Bob*. Notice how Bob* gets 80 years of life, i.e., the opportunity to enjoy desirable 'states, conditions or types of activity', whereas Bob merely gets 40 years to do the same. In short, Bob* gets the chance to experience more of the things that make life desirable than Bob does. Another way to phrase this is that Bob* would, throughout his 80 years of life, (presumably) accumulate more total well-being than Bob would have the chance to do, throughout his shorter 40 years of life. According to the deprivationist view Bob is therefore *comparatively* worse off than Bob*. This is what accounts for the badness of Bob's death at forty; dying causes him to lack what he *could* have gotten had he not been crushed by the piano, that is, he lacks what Bob* got.

15 We might imagine that he was, due to briefly stopping and exchanging pleasantries with his neighbour as he was leaving the house, a few minutes delayed in relation to the original Bob.

The Deprivationist move, then, is to reject the existence requirement (*for something to be good or bad it must be experienced*). In turn, the epicurean conclusion is undermined. Let's look a bit closer at just what kinds of evil comparative bads are. Notice how the evil of lacking, functions a little differently from *intrinsic*¹⁶ and *instrumental*¹⁷ evils in the sense that the latter two obviously require an existing subject in order to be bad. For headaches to be bad, for example, they clearly require an existing subject; if you don't exist you can't experience headaches, thus, if you don't exist, headaches could not be bad for you. In fact, intrinsic evils like headaches are bad precisely *because* you experience them (Kagan 2012, 210). The same goes for instrumental evils, for they are bad only insofar as they cause intrinsic evils (Kagan 2012, 211). Comparative evils, however, function differently, for they are bad not because they involve the experience of intrinsic evils like pain or frustrated aspirations, but because they are not *optimal outcomes*. Let's consider an example: Alice buys a 5-dollar scratch card. There are three possible outcomes: (1) she wins nothing, (2) she wins 50 dollars or (3) she wins 500 dollars. Alice scratches off the opaque covering revealing that she has won the 50 dollars. Assuming that winning the money is an instrumental good (let's say Alice will spend that money to facilitate the experience of intrinsic goods, perhaps by going to a concert with her favourite band) Alice is now better off than she was before buying the scratch card. Still, it *could* have been the case that she won the 500 dollars which could facilitate a greater quantity of good experiences, perhaps by going on an all-inclusive cruise vacation. It seems Alice is worse off than she could have been because she suffered the opportunity costs of not winning the bigger prize.

According to the deprivationist picture, then, comparative evils, the badness of lacking some good you could have gotten, provide us with a reason to think that our deaths can be bad for us despite the Epicurean observation that the *state* of being dead cannot be bad in-and-of-itself¹⁸. In other words, the harm of death is that of *preclusion* (Solberg 2019, 101)

16 i.e., things that are bad for a person in and of itself, for example pain or sorrow.

17 i.e., things that, while not bad in and of itself, causes a person to experience intrinsic evils. Getting drafted into the military, for example, could be instrumentally bad (i.e., cause intrinsic evils) if it forced you to give up your aspirations to become a famous painter.

18 Recall the previously mentioned case of the painless murder of the man known by no-one. On the Epicurean picture we struggled to find reasons for thinking that such a murder would be morally wrong. This was because the man could not be harmed by being killed; after being killed he would no longer exist; hence he would no longer be able to experience anything including the badness of having been killed. This highly unintuitive conclusion, we noted, should make us wary of Epicurean reasoning. On the deprivationist understanding, on the other hand, the wrongness of such a murder would be plain: the man was deprived of life he otherwise could have lived. The fact that the deprivationist view jibes better with our intuitions in the murder case is encouraging.

3.5. The deprivationist account of the *goodness* of death

When we're alive we can enjoy the colours of spring, we can fall in love, laugh at jokes, help a friend in need, overcome hardships, go to art exhibitions, eat juicy stakes, listen to music, spend time with family, and so on. When we're dead, we can't do any of these enjoyable or meaningful things. In short, death is the end of all that is good and meaningful for us. This makes dying earlier rather than later bad because an early death deprives us of intrinsic goods we might otherwise have enjoyed.

Of course, death is not only the end of all that's good and meaningful, but also the end of all that's bad and meaningless. If dying is bad for us because it deprives us of intrinsically good experiences we otherwise might have enjoyed, dying must, by the same token, be *good* for us because it *relieves* us of intrinsically bad experiences we'd otherwise have had to suffer (McMahan 1988, 41). If I, in this very moment, fall victim to a sudden and instantly fatal brain aneurysm, my death would be bad for me because it would rob me of the opportunity to experience a myriad of things, big and small, present and future, that I find to be good and meaningful; no more morning coffees, no more hiking, no more being loved and loving others, and so on. At the same time, my death would be good for me because I won't have to experience the bad stuff I would otherwise have had to face; no more headaches, no growing old and sick, no losing my loved ones, no more tearing my hair out trying to write this text, and so on. For the time being I hope and believe that the intrinsically bad stuff my death would relieve me of is outweighed by the intrinsically good stuff my death would deprive me of. For the sake of illustration, we might imagine the totality of my future life on a scale of well-being where +100 is the best life imaginable, -100 is the worst life imaginable and 0 is a life of neutral well-being¹⁹. Let's say that the bad stuff I have in store equates to a score of -20. On the other side, the good stuff in my future equates to +40. In sum, then, the total quality of my future life is 20, and so positive. Assuming I'm right in believing that my prospect in life is such that my dying today would deprive me of more good stuff than it would relieve me of bad stuff, dying today would be bad for me. What if I'm wrong? What if life from here on turns out to be a hellish struggle? I'll save myself from the exercise of imagining all the dif-

19 The best life imaginable, I take it, is a life containing the maximal weight of intrinsic goods and minimal intrinsic bads. The opposite goes for the worst life imaginable. I won't attempt to provide a comprehensive list of intrinsic goods and bads here. That being said, pleasurable and painful experiences are probably good starting candidates. There's no need, however, to presume hedonism. Feel free to add things like accomplishments, self-knowledge, loving relationships (or the lack of these) (Kagan 2012, 256) and whatever else you think belongs on a list of stuff that's good and bad in and of itself.

ferent ways in which my own life might go down the drain. Let's return instead to the case of Bob and Bob* from above.

Remember how we imagined two contingently different scenarios: In the first, Bob lives a full and mostly happy life until he gets crushed by a falling piano and dies instantly at the age of 40. In the second version Bob* is, through sheer luck, narrowly missed by the piano and goes on to live happily until the age of 80 when he eventually dies of some other cause. On the deprivationist account we were able to explain why Bob* was better off than Bob, or, in other words, why it was bad for Bob to die at age 40. Now, let's imagine a third version: Bob** narrowly avoids getting crushed by the piano, just like Bob* does. That same day, however, Bob** steps on a banana peel which causes him to fall down a flight of stairs and suffer a severe traumatic brain injury. The injury, while not fatal, is going to result in a drastic reduction in quality of life for Bob**. He suffers from sudden cluster headaches, nausea and vertigo, he experiences violent seizures and a debilitating loss of hand-eye coordination²⁰. Because of these health issues he loses his job which in turn results in him losing his house. On top of all this, his wife leaves him, taking the kids with her. Gradually his physical condition worsens still, further diminishing his ability to engage in activities and have experiences which, according to Nagel, makes life good, and the loss of which makes death bad. Bob** now spends his days in a cardboard box in a dank back alley²¹. Meanwhile, the luxury hotel has gone through a tedious legal process resulting in the crane company being forced to reimburse the total cost of the destroyed grand piano. A new piano is promptly ordered and this time it's going to be transported by helicopter. The delivery route goes straight over poor Bob**'s dwelling and, as I'm sure you've already predicted, by cosmic irony, the piano slips from its harness, tumbles down and crushes him in an instant.

Before he died, Bob**'s life consisted of a lot of intrinsically bad stuff and almost no intrinsically good stuff. His quality of life was substantially less than neutral, and so, worse than non-existence. Let's assume for a second the knowledge that this wasn't going to change in the future; counter-factually, had he not died that day, he would have lived another ten increasingly miserable years before dying of heart failure²². Being deprived (or more accurately relieved) of these final ten years, then, was actually a blessing in disguise for Bob**. He went from a state of negative well-being to a state of no or neutral well-being (non-existence). Bob**, then, is comparatively better off being dead than being alive.

20 The parts of Bob**'s brain responsible for rational thought were unscathed.

21 Some may not agree that this is a life not worth living. Suppose, for the sake of argument, that it is.

22 Imagine that we consulted an infallible crystal ball.

3.6. In defence of the neutral container theory

In the above we have assumed that the value of living²³ a particular life is a function of its quality, that is, the balance of the intrinsic goods and bads contained within it²⁴. This view is typically called *the neutral container theory*. Here, life is understood to be a container of intrinsic goods and bads, and merely that. Being alive, that is *being a container* of goods and bads, has no value in and of itself (Kagan 2012, 258). A life of positive quality is a life of positive value and a life of negative quality is a life of negative value. A life stripped of all its good and bad contents is neither good nor bad; it's neutral. If a life (or a chunk of life) is of negative value, living it is worse than not living at all since non-existence has a neutral value. The remainder of Bob**'s life, we thought, was an example of such a (chunk of) life. Had he *not* been crushed by the piano dropping from the helicopter that day, he would have lived another ten years and this chunk of life would have been of very poor net quality. If the value of living a life is a function of its quality, the value of living these ten years would have been negative for Bob**, and so not worth living. In short, he was better off being crushed by the piano that day than not. Some, however, think there is something more to the value of life, something beyond the sum of its good and bad contents. Thomas Nagel, the originator of the deprivation account, writes:

There are elements which, if added to one's experience, make life better; there are other elements which if added to one's experience, make life worse. But what remain when these are set aside is not merely *neutral*: it is emphatically positive. Therefore life is worth living even when the bad elements of experience are plentiful, and the good ones are too meager to outweigh the bad ones on their own. The additional positive weight is supplied by experience itself, rather than by any of its consequences (Nagel 2012, 1).²⁵

According to Nagel, then, there is something about *being alive*, about *experience in and of itself* that is intrinsically good, separate from the good and bad *elements* of experience. This is a version of what's often called *the positive container theory*: there is something inherently

23 We're talking here about the value of *living* a life from a first-person perspective, and not the value *of* that life from a third person perspective. If my eyes suddenly fell out of my skull, I'd imagine my future to be less valuable *to me* than if they didn't; my blindness would preclude my experience of lots of good things I might otherwise have experienced.

24 Accounting, of course, for the quality of the goods and bads, i.e., one large intrinsic good might outweigh many small bad ones, and so on.

25 This excerpt is taken from a revised edition of the original 1970 essay.

good about simply *being alive* over and above the inherent goods and bads contained *within* life. We might say, then, that the *global value* of a life is a product of its *quality* plus the value derived from *life itself*. This additional good skews the scale of the value of life towards the positive such that a life of *negative quality* (in terms of its contents) could, nevertheless, be of *positive value*, and so, worth living. As an example, let's imagine that we tally up the intrinsic goods and bads contained within a particular life of bad overall quality. We might end up with something like this: *intrinsic goods* +20, *intrinsic bads* -30. In sum, then, this life has a quality of -10. On the neutral container view, the value of living this life corresponds directly to its quality, and so, it would not be worth living. If, however, we add some additional intrinsic good derived from *life itself*, let's say, +20 value, this life's *global value* turns out to be +10, and so, in sum positive. What we're left with is a life of bad quality that's nevertheless worth living. If this is the case, Bob**'s death might not, as we previously concluded, have been a blessing in disguise at all, but instead a harm: he was robbed of a future that was valuable even though its quality was to be negative.

There are, however, reasons to be sceptical of the positive container theory's plausibility. First off, it's hard to imagine what it is about experience itself that's somehow intrinsically good while not being reducible to the intrinsically good and bad contents of life. Even if we disregard this ontological worry, the positive container theory does not, I think, seem plausible. According to Andrew Y. Lee (2022), the positive container theory might be understood to function in one of three ways. First, the value of life itself could be a *linear* function, that is, starting at zero at the beginning of a life and growing steadily at some fixed rate in accordance with the duration of life. The longer the life, the higher the value generated from life itself (2022, 5). Second, the value of life itself could be *all-or-nothing*, that is, some fixed absolute value divorced from the duration of the life. On this version, a life that lasts for one minute generates the same value due to life itself as a life that lasts, say, 100 years. Third, it could somewhere in-between the previous two versions; an *asymptote*. Now, according to Lee, all three of the above versions of the positive container theory will produce problematic and unintuitive results about the value of the following three kinds of lives at varying lengths: *the excellent life* "is a life with an average quality very far above zero" (Lee 2022, 4), *the awful life* "is a life with an average quality very far below zero" (Lee 2022, 4) and lastly *the empty life* which is "a life devoid of any goods or bads (except for life itself)" (Lee 2022, 4).

If we think that the supposed value derived from life itself increases *linearly* we're forced to say that an eternally long *empty life* is preferable to an *excellent life* of normal dura-

tion (Lee 2022, 6). This derives from the fact that, no matter the rate at which the value generated from life itself is amassed, so long as it increases linearly, an infinite *empty life* will be a life of infinite value. By comparison, the value (both *contained within* and *of* life) generated in a finite *excellent life* might be very high, but never infinite²⁶. This, Lee thinks, just doesn't seem right; surely a very long *excellent life* is preferable to an infinite *empty life* (Lee 2022, 5)²⁷. I'm prone to the same intuition.

If we instead think of the supposed value derived from life itself as an *all-or-nothing* type of deal, it seems we're committed to the severely unintuitive conclusion that a very short (at the order of minutes) *awful life* is worth living (Lee 2022, 9). This derives from the fact that (on the *all-or-nothing* version) for the value generated by *life itself* to render a life of bad quality and of normal human duration worth living, the value due to *life itself* has to be very high (Lee 2022, 7). Imagine a life of negative quality and of normal human duration that just barely makes the cut as a life worth living on the positive container view²⁸. Let's call this life *bad life worth living*. Now, in order for the *all-or-nothing* value derived from *life itself* to tip the scale such that the value of *bad life worth living* remains positive, the extra value would have to be very high in order to counteract the badness accumulated over the entire course of *bad life worth living* (Lee 2022, 7). Remember, on this version of the neutral container theory, the 'value injection' derived from *life itself* is a one-time-thing. As a life increases in duration, however, the positive or negative net value derived from the *contents* of the life will continue to increase correspondingly (value, positive and negative, clearly relates to duration; a headache that lasts for one hour is less bad than one that lasts for two, a bad life that lasts 40 years is better than one that lasts for 80 and a good life that lasts for 80 years is better than one that lasts for 40). In other words: the one-time value injection due to *life itself* has to be very high to be relevant at the scale of a bad life of normal human duration. If we then imagine a very short but very horrible life and add to it the same large value injection due to *life itself*, we might be left with the less than plausible conclusion that a very short horrible life is worth living (Lee 2022, 9)²⁹. If I was, immediately after having been born, thrown into a vat of acid in which I lived for exactly one minute before I completely dissolved, I would rather not have been born. On the *all-or-nothing* version of the positive container theory, however, this is exactly the kind of life that would, unintuitively, be considered worth living.

26 Notice that this is equally true if we imagine the *excellent life* be extremely long, say, at the order of billions of years.

27 Lee calls this *the argument from eternity*.

28 On the neutral container view the life would be considered not worth living.

29 Lee calls this *the argument from hell*.

The last remaining option for the positive theorist, Lee continues, must lie in supposing the value derived from *life itself* to be neither purely *linear* nor *all-or-nothing*, but instead something in-between, in the form of *asymptote* (Lee 2022, 8). On this version, the value due to *life itself* increases in accordance with the duration of life (as with *linear*), but not in a straight line; it tapers off gradually until it eventually hits an upper bound (as with *all-or-nothing*) (Lee 2022, 8). In other words, on the *asymptote* version of the positive theory, the value due to *life itself* starts off at 0 at t^0 , increases by some amount until t^1 , after which it continues to increase between t^1 and t^2 but now at a *lesser rate* compared to that of the period between t^0 and t^1 , and so on. This makes a curve that tapers off and, eventually, no additional value due to *life itself* is amassed with further duration. Now, the problem with *asymptote* becomes clear when one tries to specify the rate at which the global value of life increases due to the value amassed by *life itself* (in other words how steep the curve is and how quickly it reaches the point where no additional value is amassed with duration). First off, for *asymptote* to be guarded from the previous argument against the *linear* version, the rate of value generation has to be sufficiently high, that is, the upper bound or ceiling where *life itself* ceases to amass value with duration is reached *before* an infinite *empty life* has the chance to amass more value than an *excellent life* of normal human duration (which we have already established would be an unintuitive result) (Lee 2022, 9). At the same time, for *asymptote* to be guarded from the previous argument against the *all or nothing* version, the rate of value generation due to *life itself* has to be sufficiently *low* such that a life of both very short duration and very low quality is rendered *worth living* (Lee 2022, 9) (again, we have already established that this would be an unintuitive result). The damning thing for the positive theorist is that, for the rate of value generation due to *life itself* to be sufficiently high to guard from the argument against *linear* it will fall prey to the argument against *all or nothing*, and, at the same time, for the rate of value generation due to *life itself* to be sufficiently low to guard from the argument against *all or nothing*, it will fall prey to the argument against *linear*; there is no sweet spot, and so, the *asymptote* version too fails (Lee 2022, 9).

3.7. Who decides whether or not life is worth living

Lives filled with pain and suffering such as Bob**'s seem to be the most obvious and intuitive candidates for lives that might, from the first-person perspective, be deemed not worth living. I'm not sure, however, that it's necessarily only very painful lives that might qualify. In the film *Whose Life Is It Anyway* we meet Ken Harrison, a professional sculptor, who, after a

freak traffic accident, is left paralysed from the neck down with no real chances of recovery. After waking in his hospital bed Ken spends several months grappling with his new reality as a quadriplegic and through this process it gradually becomes clear to him that he wants to die. As time goes on, he becomes increasingly resolute in this conviction. Ken derived his entire sense of identity and purpose from his work as a sculptor and being able to create and do things with his hands. “My whole being, my entire imagination speaks to me though, ehm... *spoke* to me though my fingers” “Life for me is over. I cannot do the things I want to do. I can’t even say the things I want to say. And so for me, it might as well be over”.

In the case of Ken, it’s not so much the presence of intrinsically bad stuff like pain that makes him feel as if his future life is not one worth living, as much as it is a lack of intrinsically good stuff like purpose, fulfilment, identity and creative production. I’m by no means making any kind of generalized claim that living with a seriously disabled body necessarily makes life not worth living. I’m merely suggesting that for some individuals in some circumstances, it may be so. Ultimately, whether one’s life is worth living or not is, I think, a subjective matter, the truth of which is appropriate only on from an individual’s first-person perspective. The opinion of relevance seems to be the one who’s life it is. We can’t, I think, deny Ken his well-considered opinion that the conditions for his existence is no longer compatible with what he takes a life worth living to be. At the same time, had Ken come to the opposite conclusion, we wouldn’t be warranted in saying to him that his opinion is wrong; that he *should* prefer death to life.

Let’s say you and I went for a hike in the mountains. The trek involved both good and bad intrinsic experiences. On the one hand, we got blisters on our feet and mosquito bites on our arms. On the other, we experienced the serenity of nature and a beautiful sunset. Upon returning to our car, we debrief; was the trip worth it? Did the good parts outweigh the bad? It seems to me there’s no matter of fact right opinion here. If you responded “no, the mosquito bites really bothered me, and the sunrise wasn’t even that nice”, I wouldn’t be warranted in claiming that your opinion is wrong. The qualitative experience of *what it’s like for you* to have blister and to look at sunsets is epistemic information only you are privy to, and so, it seems, only you are competent to make evaluations as to their value. Of course, if we were regular hiking buddies and I knew that you normally found treks such as this one to be worth it, I might reasonably assume that you’re just in a bad mood; that you’ll probably feel differently after a warm shower and some shut-eye. If, however, you remained resolute in your convicting that going for hikes in the mountains just isn’t worth the following week and the week

after that one, I wouldn't be warranted in dragging you along against your will, claiming that you're wrong in your assessment. The same, it seems to me, must go for life in general, for what is life but a series of blisters and sunsets.

3.8. Chapter conclusion

We started this chapter with a question: can death be a benefit for the one who dies? Without an affirmative answer to this query, we reasoned, any notion of rational suicide is doomed from the get-go: if dying is bad *categorically*, it seems impossible to rationally seek it out. To approach this question, we ventured to answer its opposite: can death be a *harm* for the one who dies? Starting with the Epicurean observation that death can't be *intrinsically* bad for the decedent, we looked to *the deprivationist account* for answers. Even though Epicurus is correct in saying that it can't be *intrinsically* bad to be dead, he's missing a crucial point: death is bad for the decedent insofar as it *deprives* him of good life he *would* have lived had he not died when he did. By the same token, if the chunk of life death deprived one of was to be a bad one, dying, it seems, would be good. In other words, death has the *potential* to be both good *or* bad, it all comes down to the comparative value of one's continued existence to that of non-existence. The value of a chunk of existence is a function of its quality, that is, the balance of its intrinsically good and bad content, and nothing else. Non-existence has no contents and thus an intrinsic value of zero. If one's continued existence is set to have a net value of more than zero it's comparatively better than non-existence (i.e., dying at that point would be bad) and if one's continued existence is set to have a net value value of less than zero it's worse than non-existence (i.e., dying at that point would be good). It seems, then, we've found an answer to our query: death *can*, in fact, be a benefit for the one who dies, just as it can be a harm.

4. Rational suicides

Of course, just because death can be a benefit for the one who dies, it doesn't necessarily follow that it's ever rational to commit suicide. In this chapter we'll explore what it means to consider suicide from the first-person perspective and try to paint a clearer picture of when it might make sense for a person to cut her life short.

David Hume famously wrote that “I believe that no man ever threw away life while it was worth keeping. For such is our natural horror of death, that small motives will never be able to reconcile us to it” (Hume 2005, 10). This is, you might agree, a strong claim to make. If we take Hume's assertion to mean something like “no man ever threw away life *while his moment-to-moment existence was worth experiencing*”, well, that seems quite reasonable. If, however, what Hume meant was something like “no man ever threw away life *while his future was worth sticking around for*” well, that seems less plausible. While it's unquestionably true that the character of a person's moment-to-moment existence might be disagreeable to the point that it's not worth *having*, the question of whether one's life is worth *keeping* depends, at least largely, not on the value of one's present existence, but on that in one's future. The future is, however, uncertain and often times hard to predict. Think of the suicide of Romeo in *Romeo and Juliet*. In the climax of the play, Romeo finds Juliet lying in the crypt, apparently lifeless. Believing that the love of his life is dead, Romeo commits suicide (the reason being, let's presume, that life without Juliet just wasn't worth keeping). This, of course, turned out to be a terrible blunder on the part of Romeo, because Juliet, well, she wasn't dead at all, she merely appeared to be. Had Romeo opted not to commit suicide, he would have seen that Juliet would awaken from her comatosed state, at which point they, presumably, would have ran away together as was originally the plan. Romeo, then, threw his life away even though it was *going to be* worth living (let's presume). Of course, his mistake was an honest one, he had good reasons to believe that Juliet was in fact dead. However, I'd venture to say, even if it *had* been the case that Juliet was, in fact, dead when Romeo found her, it's far from certain that the life he threw away wouldn't have been worth keeping nevertheless. Realistically speaking, he probably wouldn't have remained heartbroken forever and might even have fallen in love again, had children and lived a happy and meaningful life.

4.1. To contemplate suicide is to make a choice between life-courses

According to Richard. B. Brandt, a person who is considering suicide is “making a choice between future world-courses; the world-course that includes his demise, say, an hour from now,

and several possible ones that contain his demise at a later point” (1975, 69). This framework for thinking about decisions about suicide makes a lot of sense, I think, though I will alter it slightly. First, I need to point out that Brandt makes a distinction between *life-courses* and *world-courses* and goes on to employ the latter because he takes rationality to be connected to morality: “One is interested in things in the world other than just oneself and one’s one happiness” (1975, 69). Choosing rationally amounts to more, for Brandt, than merely choosing according to self-interested maximization of welfare. Now, I would certainly agree that morality plays a big part in deciding what one *ought* to do. I will, however, depart from Brandt’s perspective and treat rational and moral decisions about suicide as separate entities³⁰ that might sometimes align, and sometimes be in conflict. Going forward, then, I will take it that choosing rationally is choosing according to self-interest and choosing morally is choosing according to, well, morality. Choosing according to what one *ought* to do may, I suspect, often involve a mediation between both of these axes. At any rate, what I’m trying to investigate here is not when one *ought* or *ought not* to commit suicide, but when one would be acting rationally in doing so³¹. For this reason, I’ll stick to the term *life-course* rather than *world-course* going forward. I’ll also take the term *life-course* to be, for all intents and purposes, synonymous with the notion of *possible worlds*, that is, making a choice between life-courses is making a choice between available possible worlds.

Our framework for thinking about suicide decisions, then, will look something like this: When a person is considering suicide, they’re making a choice between *two* life-courses: one that involves their death at that point (or realistically shortly after), and one where they continue living for a while longer. The first life-course is clearly a static one: if a person decides to end her life, nothing will ever happen to her again, and she can’t make any more decisions about her future, because, well, she won’t have one. The second life-course is, at least to some extent, dynamic: if a person decides to continue existing for a while longer, things can happen to her that will alter her life-course for better or for worse and she can also make further choices about further life-courses for better or for worse. As such, the value of *being in the state she is in* is certain in the first (she won’t exist and there’s no value to non-existence, positive or negative), and uncertain in the second (she will exist and the value of her existence going forward is subject to potential changes). When considering suicide, then, it’s the value

30 Kagan (2012, 319) makes a similar distinction.

31 I won’t delve into views on practical/instrumental rationality in general here. On a basic level, choosing rationally comes down to adopting suitable means to one’s ends (Kolodny and Brunero 2023, par. 1). In our context a person’s principle end amounts to the maximization of well-being, and so the basic question has to be: which of the two life-courses under consideration best suits this end.

of the life-course where one continues living that needs to be considered to determine which of the two life-courses will be the preferable one. Importantly, this life-course includes within it, a myriad of *other* possible life-courses.

So long as one is living, one's future life-course, it seems to me, can be altered in two main ways: intentionally and unintentionally. Unintentional changes to one's life-course are those influenced by the outside world or by people who are not oneself. If tomorrow I get hit in the head by a loose roof tile falling from above, my life-course will be changed for the worse. If tomorrow I get a call informing me that I'm set to inherit the fortune of a rich uncle in the States whom I didn't even know existed, my life-course will be changed (probably for the better). Of course, I didn't intend for any of these things to happen to me and to alter the course of my life, they just happened to. My life-course can also be altered, for better or for worse, by my own intended actions. I might, for example, alter my life-course for the better by texting some friends to organize a game-night or plan out some healthy meals for the upcoming week.

The life-course approach makes sense, I think, not only in the context of suicide-acts but whenever we're deciding between doing any and all kinds of acts whose outcomes will bear on our future levels of well-being³². As such, we might say: a person who is considering whether or not to hit his thumb with a hammer is making a choice between two future life-courses; a life-course that includes a bruised thumb a few moments from now, and one that doesn't. In this way I think it makes sense to think about suicide-acts in much the same way we think about other kinds of acts. If suicide act are acts (almost) like any other, they're presumably rational or irrational by virtue of (almost) the same factors that acts in general are rational or irrational. If we figure out what makes acts *in general* rational and irrational, then, we stand a better chance at making sense of what a rational suicide might look like. Although I think it makes sense to think about suicide acts in much the same way we think about other kinds of acts that may or may not have the quality of being rational, there are undoubtedly some particularities to suicide acts that warrants special consideration (one of which we've already mentioned). We'll return to this later.

4.2. Making ideally *right* choices between life-courses

I'll start off by thinking a little about what it would mean to make the ideally *right* choice between life-courses from a gods-eye point of view. Note that making the, as a matter of fact,

32 The only difference being that other acts are choices between *two* dynamic life-course, instead of one.

right choices between future life-courses is not the same as making rational choices between life-courses. We'll return to this later.

I have on my desk in front of me a cup of coffee. If I, at this moment, reach out and have a sip of coffee, I will actualize a possible world that is available for me to actualize that involves a '*taking a sip of coffee at 12:43 on the ninth of august 2023*' event. If I act so as to actualize this possible world, I will have changed the course of my life (my *life-course*) to have involved a coffee-sipping event on the ninth of august at 12:43 2023. Whenever a person is choosing whether or not to act in a way that will alter her life-course, that is, things like whether or not to take up smoking, go for a jog, get a cat, text her crush, buy a sail boat or end her life, she is making a choice between the life-courses that would obtain if she did and did not go through with her act. She's choosing between two possible worlds that are available for her to actualize.

Of course, unless someone has laced my coffee with cyanide, the possible world that involved the coffee-sipping event will, on the whole, be very similar to the possible world that didn't. There will, however be an ever so slight difference in net value between the two. If, for example, the coffee tasted terribly bitter or if it burned my tongue, the *sipping-world* would be ever so slightly worse for me on the whole than the *non sipping-world* that I would have actualized had I opted not to sip (in other words, the *sipping world* would contain one unpleasant experience that the *non sipping world* wouldn't). As such, it's possible to say that one of these possible worlds or life-courses *would* be better or worse for me compared to the other one. This is, I think, in line with Fred Feldman's principle about comparing the overall value of possible worlds for persons:

To say that a state of affairs would be all things considered good [or bad] for a person is to say that she would be better [or worse] off if it were to obtain than she would be if it were to fail to obtain. More exactly, it is to say that her welfare level at the nearest possible world where it obtains is higher [or lower] than her welfare level at the nearest possible world where it does not obtain (Feldman 1991, 214).

If we apply this principle to the making of choices, we might get something like the following: When a person *S* is making a choice between *act* or $\neg act$ ³³ and where *act* or $\neg act$ will result in different possible worlds with different levels of well-being, it seems a reasonable

33 *Act* is meant to be a stand-in for things like *have a sip of coffee*, *buy a sail boat* and *commit suicide* and $\neg act$ a stand-in for *not having a sip of coffee*, etc. Furthermore, I'll take it that $\neg acts$ are acts just as much as *acts* are, in the sense that they are acts of omission; in other words, a $\neg act$ is the *act* of not doing *act*.

starting-point to say that: if *S* chose to *act*, she made the ideally *right* choice if the welfare level in the possible world that was actualized by *act* (i.e. the *act-world*) was higher than the nearest possible world that would have followed from $\neg act$ (i.e. the $\neg act$ -world), all things considered. Getting a cat is the right move on Cathy's part if the possible world in which she did so would be better for her than the nearest possible world in which she didn't, all things considered. Buying a sail boat is the wrong move on Ben's part if the possible world in which he bought a sail boat would be worse for him than the nearest possible world in which he didn't, all things considered. Committing suicide is the right move on Sam's part if the possible world in which he no longer exists would be better for him than the nearest possible world in which he does, all things considered.

There is, however, at least one complicating factor that needs to be taken into consideration here. If, for example, Cathy decides to act so as to actualize the possible world in which she owns a cat, she will, of course, now exist and continue to act in *that* world, forced to make a bunch of further life-course decisions became relevant and available to her precisely *because* she chose to actualize the *cat-world* in the first place. Her continuing choices about these *cat-world* related worlds will, in theory, have some bearing on the rightness of her initial decision to actualize the *cat-world*. Imagine that Cathy made a life-course choice and opted to actualize *cat-world*. Once *cat-world* is actualized, Cathy will be faced with further life-course choices stemming from her initial decision to get the cat. She may, for example, procure a litter box and train the cat to use it appropriately *or* she may not do this and simply let the cat piss on the carpet as it pleases. Obviously, now that she exists in *cat-world*, the possible *litter box-world* is going to be comparatively better than the *carpet pissing-world*, but if she fails to make the right move here, it seems to me that it may become *retroactively* true that she would have been better off in the world where she chose *not* to get a cat in the first place. In other words, *cat-world* followed by *litter box-world* might be better than the initial $\neg cat$ -world but *cat-world* followed by *carpet pissing-world* might be worse than the initial $\neg cat$ -world. All this is to say that: upon actualizing a possible world (or picking a life-course), a person will continue to act and alter this now actualized life-course further, and these alterations will, when relevant to it, bear on her initial choice between life-courses retroactively. This obviously makes evaluating the *rightness* of choosing one possible life-course over another epistemically difficult. Of course, I don't want to simply throw in the towel here, so I might propose something like this for a definition of making the ideally *right* choice between life-courses: if *S* chose to *act*, she made the right choice if the welfare level in the possible

world that was actualized by *act* was higher than the nearest possible world that would follow from $\neg act$, supposing that *S* would have continued to make the right subsequent life-course choices when relevant in the respective worlds that would have followed from the initial choice.

4.3. Making *rational* choices between life-courses

As I mentioned above, making ideally *right* choices and making *rational* choices are not one and the same. Imagine a man who wagers a month's pay with his colleague on his ability to flip a coin and have it land tails five times in a row on the first try. The odds of the coin landing tails up five times in a row is 31 to 1. Now, if the man just so happened to flip five tails in a row, it seems he did *in fact* actualize a world that was comparatively better to the one that would have been actualized if he didn't propose the bet in the first place. We wouldn't, however, say that the man acted *rationally* in proposing the bet, because the odds that *bet-world* turned out better than $\neg bet\text{-}world$ was far lower than the opposite. Choosing rationally then, is not about choosing as a matter of fact *correctly*, it's about choosing *reasonably* according to probabilities.

4.4. Life-courses within life-courses

As previously mentioned, a person who is considering suicide is making a choice between two life-courses. One (non-existence) is static, and the other (continued existence) is dynamic. The dynamic life-course can be altered further, both intentionally and unintentionally.

Imagine a man who lives in a secluded house in the woods. During a walk in the forest he steps on not one but two bear traps, one for each foot. Being a stoic, he grits his teeth, crawls back home and collapses on his bed. The next morning he makes no effort to remove the two bear traps, but instead accepts his plight and tries to go about his day like he normally would, dragging himself along the floor towards the kitchen. With each movement the teeth of the traps dig deeper into his calves. He tries to pull himself up to the coffee machine on the counter, but his feet can't hold his weight and he sinks down to the floor. He wriggles to the living room and turns on the television. "Damn" he thinks to himself "it seems I won't be able to go fishing tomorrow". He keeps this up for several months, but time moves slow when both your feet are stuck in bear traps. "Oh, why must I suffer so terribly" he might whimper, deciding there and then crawl to the bridge, roll himself over the edge, and let the weight of the iron jaws pull him down, down towards salvation.

This, you'll probably agree, does not seem like a reasonable move; in fact, it seems completely ridiculous. Surely there must have been some possible life-course available to him that involved neither living with bear traps on his feet nor the sweet relief of death. The idiom "you are the architect of your own happiness" is clearly not categorically true; infants born with cancer didn't choose to be born with cancer. There is, however, something to it in some cases; people who find themselves in a tough spot do at least sometimes have *some* power to improve their prospects for a better future existence. A person who is feeling down might see a therapist. A person who is lonely might join a quilting class at the community centre. A person who suffers from a lack of purpose in life might try volunteering at a soup kitchen. A person with both legs stuck in bear traps might call a doctor.

As such, a person considering suicide needs to account for possible ways she can alter the life-course wherein she remains alive intentionally so as to make it worth living. In a sense, the life-course where she keeps living for a while longer may contain within it a myriad of further possible life-courses, and if one of these life-courses stands a good chance of being a life-course worth living, that is of being better than the life-course in which she ceases to exist, she should obviously make the attempt to realize that life-course, and to do that, well, she has to opt to continue existing for a while longer.

4.5. Choosing between life-courses in the face of uncertain futures

If, upon asking my friend for advice about what major to pick (let's say between philosophy and business management) he answered "oh, that's easy! Just pick the option that will bring about the possible world in which the state of affairs that obtain will yield a higher level of welfare for you than the nearest possible world in which that state of affairs does not obtain", well, that's not all that helpful, is it? This is, of course, because the outcomes (especially the long term outcomes) of most life-course choices are shrouded in a mist of uncertainty. If I knew beforehand whether the life-course in which I choose to major in philosophy would be, all things considered, better compared to the one where I major in business management, choosing between the two would be easy indeed. Sadly, we're not so epistemically privileged. We can, however, make more-or-less reasonable predictions about how the future might turn out. If I major in philosophy, I imagine, my pay-check will probably be more modest than that of a businessman. On the other hand, I might cultivate some semblance of moral perceptiveness (regarding this, the jury is still out). If I major in business management, I might be able to af-

ford a nice car and vacations to the French Alps. On the other hand, I wouldn't experience the delightful joy of reading Kant.

The life-course that follows from *act* or \neg *act* always happen *after* the choice between life-courses has already been made. It's clear, then, that we're not really able to choose between life-courses as such, but between what we *imagine* or *predict* the life-courses actualized by *act* and \neg *act* to be like. To be able to choose rationally between *predicted* life-courses, then, a person must obviously form reasonable beliefs about what the life-courses that might follow from *act* and \neg *act* will look like. In other words, a person who is deliberating whether or not to *act* must (a) imagine and predict what the life-courses following *act* and \neg *act* might be like and (b) evaluate whether *act* or \neg *act* is most likely to result in the comparatively better life-course. Of course, we can't demand that the person be able to predict future outcomes and their probabilities with perfect accuracy, that would be over-demanding; it would mean that we're almost never able to rationally choose whether or not to do anything. We might say, then, that a person must be able to do (a) and (b) to a reasonable degree in order to choose rationally between life-courses. To act rationally in get a cat, then, Cathy would have to form reasonable beliefs about what the relevant possible life-courses in which she does and does not get a cat might look like. The *cat life-course* might contain a torn-up couch, loud meowing in the night etc., but also a pleasurable and meaningful relationship.

If this is right, it's clear that choosing rationally is not always possible. For starters, and this I take to be self-evident, a person must possess the mental faculties necessary so as to preform (a) and (b). Harkening back to our discussion of Fairbairn's account of suicide, it seems reasonable to think that a young child or a seriously mentally deficient person might lack the mental faculties necessary to engage in prediction and reasoning of this kind. Furthermore, an otherwise capable person might sometimes, due to factors like mental illness, intense pain, the influence of drugs, lack of sleep, etc., have his capacity for rational thought reduced to such an extent that he is no longer able to reasonably predict the life-courses following *act* and \neg *act*. Notice how several of these factors involve an element of suffering. This might strike you as a little odd. Suffering is, as we saw in last chapter, one of the things that might render a life not worth living. If a person is suffering from some terribly painful and incurable disease, then, the very thing that makes it true that he'd be better off dead is the same thing that renders him unable to suicide rationally in so far as it makes him incapable of (a) and (b). This may seem like somewhat of a catch-22 situation. I think, however, that so long as we keep third and first-party perspectives separate, it's not really an issue. While I may think that

my (luckily hypothetical) sick cats dying would be in his best interest, I recognize that my cat is not capable of forming reasonable beliefs about the future life-courses in which he does and does not exist, and because of this, could not *understand* that he would be better off dead. In this chapter, we'll restrict our discussion by presuming a normally rational state of mind. In chapter 5, we return to consider potentially irrational states of mind in regard to decisions about suicide.

Let's look at a simple and empirically straight forward example of someone who's in the position of choosing between doing or not doing some act and see how they might be rational in doing so. Imagine a man standing in the lobby of a casino deliberating whether or not he should enter to play at the slot machines. If he decides to play, he might lose *or* he might gain some amount of money. If he decides against playing, he'll do neither. These, it seems, are the possible life-courses that might obtain from playing and not playing. The slot machines are, as is always the case in casinos, rigged in favour of the house. If he decides to sit down and play, then, the *most likely* outcome will be that he loses some amount of money. Of course, if he chooses to play there is always the possibility that fortune smiles and he hits the jackpot. If this possibility comes to pass the man did, in a sense, make the right move in choosing to play. This is, as we've touched on before, only true in hindsight and thus irrelevant. What matters in the context of choosing rationally is the probability of the winning outcome at the time he made the decision to play. The probability (prior to playing) that he winds up in a possible world where he wins more than he spends is lower than the opposite possibility. This is knowledge he might reasonably acquire, say by thinking long and hard about how a business whose revenue comes from the wagers of gamblers would stand to make a profit, or perhaps by simply standing in the lobby for a while, counting the number of happy and frowny faces leaving the casino. Knowledge of this kind seem a necessity for choosing rationally in this case: if he was not aware of the fact that the slot were rigged in his disfavour, his predictions about the life-courses following his decisions would not be reasonable. If he made the right choice and made it for the right reasons, that is by reasonably predicting the possible outcomes of playing and not playing, I think he chose and acted rationally in not playing. Of course, it's possible to make the right choice for the wrong reasons. If the person next to him chose not to play on the grounds of his unreasonable belief that a demon hiding in the slot machine would spit acid in his face if he got too close, his choice would not be rational, although it would accidentally align with his best interests.

The casino example is, of course, assumptive and simplistic compared to most real-life cases of choosing whether or not to do some *act*. It's assumptive because it assumes that the man's preferred outcome is the one where his bank account balance remains the highest. This, of course, needn't be true. Perhaps gambling, even when he ends up losing, provides him with a sense of fun and excitement. If he values fun more than funds in this way, playing the rigged slot machines might be the rational choice for him after all. Similarly, if you're deliberating whether or not to accept your boss' offer of promotion, either choice might be rational for you depending on how the outcomes jibe with what you deem to be of value in life; the heightened salary and social status that comes with the promotion might be worth the increased workload and responsibility, or it might not be. It's clear that one person's wants, preferences and overall judgements about what is of value to them will not always be the same as that of another. In this way, a person's means towards overall well-being are her own, that is, subjective. In the same vein, means are plastic; a person's means ten years from now may not necessarily be the same as they are today. When we're dealing with decisions whose outcomes could affect us far into the future, it's probably necessary to give some thought to the possibility that one's future self might disagree with one's present self about *act* being the right means to maximize well-being in the long run. If, for example, I was considering whether or not I should get a neck tattoo in order to be as cool as the guys in my favourite punk rock band, I think my decision to do so would be less than rational if I completely disregarded the possibility that my future self might not share my opinions about the coolness of neck tattoos. Here too, we must be careful not to be overly demanding. Rational choice can't require perfect knowledge of our future preferences and values, this would make rational choice practically impossible. Still, I think, we must try to the best of our ability to imagine our future preferences and take them into account if we want to choose rationally. The same seems true in the context of suicide; if I find my situation disagreeable today, it's possible that my preferences might change with time such that the same situation would be less disagreeable to me ten years from now³⁴.

4.6. Rational suicide in the face of an uncertain future

The casino example is exceptionally easy to navigate because it's, given the logic of how casinos operate, no hard task to deliberate the probable outcomes of playing and not playing. If, for example, we imagine that the slot machines were rigged in such a way that the chances of the man winning more than the amount he bet was 49% and his chances of losing more than

34 Brandt makes a similar point (1975, 69).

he bet was 51%. If he chose to play only once, then, the outcome of doing so is in fact highly uncertain, but the uncertainty is calculable and *tangible*. It's because of the tangibility of the uncertainty that the rational choice was obvious. Let's apply this logic to the context of suicide to see when it might make sense to choose the life-course that involves one's ceasing to exist over the life-course that involves one's continued existence. To approach this, I'll borrow a thought experiment from Shelly Kagan (2012)³⁵.

Kagan imagines a *twisted game show* wherein the contestant must choose whether to go through one of two doors (the equivalent of life-courses). If the contestant goes through door 1, there is a very high chance (perhaps something like 99.9%) that he will be kidnapped and tortured horribly for the rest of his life (e.g., the remainder of his life will be of very bad quality). There is, however, a slim chance (0.1%), that what awaits the contestant behind door 1 is not in fact life-long suffering, but instead a wonderful vacation lasting the rest of his life (e.g., the remainder of his life will be of very good quality). If the contestant goes through door 2, he will instantly fall into a state of deep dreamless sleep for the rest of his life (e.g. the remainder of his life will be of no/neutral quality). Now, Kagan continues, it seems preposterous to insist that the contestant *ought* to go through door 1, despite the knowledge that the torture-outcome is much more likely than the vacation-outcome. To go with door 2 must surely be, at the very least, *rationaly acceptable* Kagan concludes (2012, 340).

Kagan is, I think, overly modest in his conclusion. It's hard to imagine how door 1 could be the rational choice even if the odds of the torture and vacation outcomes were more evenly matched, say, 51% to 49% respectively³⁶. There is, however, one important factor to suicide decisions that's not accounted for in *twisted game show*; the *continue living* life-course is, as we've previously noted, dynamic and contains within it the option of committing suicide *later*. Let's start by applying the logic in *twisted game show* to a case where this option wouldn't be of much relevance before altering the case to see how it would.

35 This is one part of a series of thought experiments developed by Kagan to guard against what we might call the *optimistic fallacy*: the idea that suicide never makes sense because there's always a sliver of a chance that one's future might turn out to be worth living, against the odds. I'm borrowing it for a slightly different purpose.

36 I'll mention also, as an aside, that being horribly tortured seems to me *more* bad than a wonderful vacation is good. The two are obviously meant to be symmetrical for the argument to make sense, but this does strike me as artificial. Imagine meeting a mad scientist with an experience machine. If you allow me to hook you up to this machine, he says, you'll experience the best pleasure imaginable for two hours, but for one final hour you'll be tortured with the worst pain imaginable. Would you try the mad scientist's machine? I certainly wouldn't. In this sense, I think there is a case to be made for choosing door 2 even if the pleasure outcome was *more* likely than the pain outcome. Plausibly, the same goes for good and bad lives, that is, the badness of a really bad life seems to me worse than the goodness of a really good life. I won't, however, pursue this thought further.

Imagine that you're infected by some exotic virus that makes your existence highly unpleasant to the point of not being worth living moment-to-moment (to ensure symmetry, take the quality of life you *would* have enjoyed had you not been infected and add a minus in front to get the character of your virus-existence). There is no cure for your condition, and the virus is so complex that it's inconceivable that one could be made, at least for another century or so. Furthermore, the viral illness makes no impact on your life expectancy. There is, however, one more peculiar fact about your condition: in 40% of cases, the illness spontaneously and abruptly disappears at some random point in time; it could happen a day or a week after infection, it could happen when you're on your deathbed at 80, or at any other point. Or, it might never happen. You're faced, then, with a choice between two life-courses: (1) continue living and, (2) end your life. (1) is functionally the same as door 1 in *twisted game show* and (2) the same as door 2. If you go with option (1), then, your future is, admittedly, highly uncertain, but nevertheless, the most *likely* outcome is that it will be of bad quality over-all and that you'd be better off not existing. If you decided to bet on life, you might, of course, luck out but probability says that you won't. If you were 30 when you got infected, you'll probably spend the next 50 years or so (or at least the majority of your future) in agony. Suicide does seem to be the rational move here in the same way that *not playing* was the rational move in *casino* and door 2 the rational move in *twisted game show*.

Next, imagine the same situation, but this time scientists have figured out that a small subset, let's say 5% of the population are, due to a genetic mutation, almost certain to experience sudden recovery within the first year after infection. Globally, the chances of recovery are the same as they were above. What's more, the scientists have devised a blood test that can determine if you're one of the lucky 5%. The test takes four weeks to run. Now, at the time you became aware of the fact that you've been infected with the virus, let's call this t^0 , there's, just as in the previous case, a 60% chance that your future will be a terrible one on the whole such that you're better off ceasing to exist at t^0 . However, at t^0 , suicide no longer seems to be the rational move; taking the test intuitively seems to make more sense. Why is that? Well, in this case there's an obvious asymmetry of stakes between the two life-courses (1) and (2) that wasn't there in the previous one. That is, (1) makes sense here, because you, in deciding to continue living a while longer, only risk the four weeks of suffering it takes for the test results to get back from the lab. If, at t^1 , the results show that you're *not* one of the lucky few, well you could just go ahead and end your life at that point. On the other hand, if the test shows that you *are* one of the lucky ones, the potential reward is far greater than the badness

of four weeks of suffering; it's, presumably, 50 years of a life that's worth living. The rational move here, then, seems to be to choose the *continue living* life-course at t^0 even though the odds of your future being *worth living* was only 40%³⁷. Notice that it only makes sense to wait for the test results if you know that you'll have the opportunity to end your life at t^1 . If you only had one chance to end your life and that chance was at t^0 , your situation would be the same as in the previous case where no test existed.

Let's look at one more case where the stakes are asymmetrical such that the *continue living* life-course makes sense even when the odds of having a future worth living are stacked against you. Imagine a different virus. This one will, after infection, make your life miserable and after four weeks, 60% will die. Those who pull through make a quickly recover and return to health. At t^0 , then, there's only a 40% chance that the rest of your life will turn out to be worth living. Is suicide rational at t^0 in this case? Well, it seems to me that it wouldn't be, the reason being, as in the precious case, the difference in the stakes involved. At t^0 the life-course where you remain living is most likely to be a short and pretty terrible one. However, there's a 40% chance that it will turn out to be a long one that, while terrible for the first four week's will (let's presume), on the whole, be comparatively much better than non-existence.

The cases above are, of course, deceptively easy to navigate. They're a lot like the casino case in that the outcomes of playing and not playing at the slots were, yes uncertain, but tangibly so. This makes the rational move easily calculable. In real life, people considering suicide are probably often faced with a far more intangible kind of uncertainty about their future existence. Perhaps more akin to the case of Cathy making the choice between the world-course where she owns a cat and the one where she doesn't. Cathy obviously isn't privy to the kind of knowledge we've supposed above. If she knew that the life-course in which she gets a cat has a 60% chance of being better than the one where she doesn't well, her choice would be easy. Even so, it does seem like it has to be possible, at least in some cases, to choose rationally whether or not to get a cat. We must resist the temptation to set the bar of predictability to high. People make high stake decisions about futures that are hard to predict accurately all the time. Choosing whether or not to have children, propose to one's significant

37 To make this clearer, let's retrofit *twisted game show* with the same asymmetry of stakes-logic. Again, you're faced with the decision between doors 1 and 2. Door 1 has a 99% chance of life-long torture and a 1% chance of a life-long vacation. Door 2 has a 100% chance of life-long dreamless sleep. This time, however, you have in your pocket a pill that, when swallowed, puts you in a state of life-long dreamless sleep, exactly the same as that of door 2. Now, which door should you pick? Obviously you should try your luck with door 1 because, in the likely event that you're met with torture, well, you can just pop the pill. Risking 10 seconds of torture for a 1% chance to have an excellent and long life seems like a no-brainer.

other, move to a different country, take out a big loan to buy a house, etc., all involve a high degree of intangible uncertainty. There's always the possibility that the housing market will crash or the interest rates on your loan sky-rockets without warning. Of course, there's definitely economic conditions under which buying a house just isn't a smart move. To say, however, that it's *never* rational to take out a big loan to buy a house because there's always a small chance that something unpredictable will happen, well, that just doesn't seem right. The same, I think, has to be the case for suicide. In the words of Brandt: "The argument is often used that one can never be *certain* what is going to happen, and hence one is never rationally justified in doing anything as drastic as committing suicide. But we always have to live by probabilities and make our estimates as best we can" (1975, 70).

To summarize, my suggestion for when suicide might be rational will look something like this: if it's more likely than not that the *ceasing to exist*-life-course is going to be comparatively better to the *continue living*-life-course *and* difference in stakes and alternative life-courses have been accounted for, suicide makes sense³⁸.

4.7. The difficulty of evaluating the worthiness of one's moment-to-moment existence

The uncertainty of one's future quality of life notwithstanding, there's another epistemic challenge facing the suicidal: knowing *when* life is so bad that death would be preferable to it. Exactly where the line goes between one's moment-to-moment existence being worth experiencing or not, that is, *when* the quality of one's moment-to-moment existence dips below 0, seems hard to pin down. This is, I'd wager, because there is in fact no hard line. While I feel pretty confident in saying that my existence in this very moment is one worth having, I have no trouble imagining an existence I would rather not endure. The in-between cases, however, quickly become hazy. There seems to be an element of *the paradox of the heap* to evaluating the worthiness of experiencing one's moment-to-moment existence. The fact that we can't draw a hard line between the two is, of course, no damning thing in itself. If we take a person with an impressive mane and start plucking individual strands of hair from his scalp, the person will eventually turn bald even though it's hard to pin down exactly *when* the transforma-

38 I'm convinced that there are many such cases in the real world. For one example, we might look to Inger Staff-Poulsen. Inger suffered from incurable ovarian cancer and chose to face death on her own terms. Before she died, Inger wrote an equal parts brave, sobering and fiery public suicide-letter chronicling her experience with terminal illness, suffering and a society refusing her the right to an, all things, considered good death. Inger's suicide was clearly a rational one. <https://www.dagbladet.no/kultur/vedlagte-kronikk-er-kanskje-litt-kontroversiell---fordi-jeg-er-dod/70706493> [accessed 31.09.2023].

tion happens. Even so, I take it, we can all agree that there exists both bald people and people with a full head of hair.

Right now, as I said, I'm feeling reasonably content with experiencing my moment-to-moment existence. If I suddenly got an itch on my back that I couldn't reach and a mild headache, I would feel a little worse, but I would still prefer experiencing this existence to having no experiences at all. If I was to have my appendix removed, however, I would very much hope to be anaesthetized during the procedure. One's existence while having one's appendix removed is clearly not one worth experiencing; I'd rather not have any experiences at all. It *does* seem, then, that the quality of my life would drop below 0 in the hours during my surgery.

To say of an experience that is not worth *having* is, of course, not to say that the experience is not worth *living though*, that's an entirely different matter. If it turned out that I was allergic to anaesthesia, I would, of course, still want to have my inflamed appendix removed even though it would be very painful, because I would know that, on the other side a valuable existence awaited me; one worth suffering for. If the rest of my life was to be a perpetual procedure of removal and re attachment of my appendix over and over again until I was 80, however, I'd rather be perpetually anaesthetized. Asking ourselves "would I rather be anaesthetized right now" might, I think, give us a rough measure of the value of our moment-to-moment existence compared to that of non-existence.

4.8. A few remarks

At the start of this sub-chapter, I stated, taking inspiration from Brandt, that a person who considers suicide is effectively making a choice between future life-courses. In this sense, considering whether or not to commit suicide is little different from considering whether or not to do other stuff that will positively or negatively alter one's life-course. Following this, I've presented a series of cases where a person must decide between doing or not doing some *act* (play at the slot machines, adopt a cat, get a neck tattoo, etc.), taking it for granted that choosing rationally in these cases will function similarly to choosing rationally in the context of suicide. One might object that these fail to be analogous because: when considering whether or not to, for example, adopt a cat, the possible outcomes of doing so and of not doing so could *both* be opaque and uncertain. In the case of suicide, however, things may at first glance appear different. When considering suicide, one might think, only the outcomes of *not* going through with it could be uncertain. The outcome of going through with suicide, how-

ever, is really quite clear; it's non-existence (ignoring, of course, the very real possibility of a failed suicide attempt). In other words, when considering suicide, the outcome of committing suicide seems completely predictable, that is, if you take our own life, you'll die, and if you die, you'll be dead and if you're dead, you won't exist. Non-existence, it seems, can only be one way; it's value can't be uncertain, because it doesn't have negative or positive value. This, I think, is mistaken. As we'll learn in the next chapter, a state of non-existence can actually be of negative or positive value depending on what you 'missed out on' by not existing. Even though the value of non-existence is certain to be 0, the *comparative* value of non-existence depends on the value of the life you otherwise *would* have led had you continued to live.

Some may think that considering whether or not to take one's own life is different from considering whether or not to do other things, because suicide acts can't be regretted after the fact. They're final and permanent in a very grave and basic way that getting a neck tattoo or getting cat is not; you could, after all, get a laser tattoo removal or leave the cat in a basket outside a shelter if you came to regret your initial act. In the language of possible worlds, the *getting a cat* world will, once attained, enable other possible worlds, amongst them the *leaving the cat you came to regret initially getting in a basket outside a shelter* world. Coming back to life once dead, however, is not an option. Persons who commit suicide can't live to regret their decision. This is, I think, an important point. Choosing to act in a way that will bring about consequences you won't be able to remedy should certainly demand a higher degree of certainty and confidence in the act being rational than cases where remedying the consequences of the initial act after the fact is possible. I think it won't, however, make the act necessarily and fundamentally irrational. Having a child, for example, is an irreversible act in the sense that once a child is born it *can* cease to live, but it can never *not* have been born and you'll never *not* have been a parent. This, however, doesn't seem to make the decision to have a child (to irreversibly bring into existence a person who otherwise wouldn't exist) fundamentally irrational. Furthermore, the fact that suicides can't be regretted doesn't seem to be, in and of itself, all that concerning. Consider this: some acts can be regretted and their consequences can (at least more or less) be remedied after the fact. If you dyed your hair green and then regretted doing so, you could just re-dye it or cut it off and wear a hat for a few months. No permanent damage done. Other acts can be regretted but not undone. If you decided not to wear your seatbelt while driving to the grocery store and you got paralysed from the neck down in an accident, you may well spend the rest of your life unremittingly regretting not bothering to buckle up. However, no matter how much you look back in regret at

your decision, it won't change the fact that you can't wiggle your toes. Regretting past mistakes you can't remedy may hold some instrumental value, but phenomenologically regret seems an unpleasant experience. Suicide is perhaps the only kind of act than can be neither regretted *nor* remedied after the fact. While it seems reasonable to hold that acts whose consequences are permanent involve higher stakes than acts whose consequences can be remedied, I fail to the dis-value of not being able to regret *per se* when the consequences of the act you come to regret are unamendable. All this being said, decisions about acts as drastic and irreversible as suicide (or having a child for that matter) clearly require exceedingly careful deliberation.

I feel compelled to restate here that, even though there might be cases where suicide would in fact be rational in the way we have sketched above, this does not necessarily entail that one *ought* so go through with it when it is. When standing in line at the grocery store my wish to return home swiftly might make it rational for me to skip in line (the possible world in which I skip might be better for me than the world in which I don't). That being so, I still patiently wait my turn. Why? Well, because there are other competing reasons that compel me to do so, like not wanting to be rude or to unduly cause people inconvenience. In short, if a person is deliberating whether or not to end their life, the rationality of their choice is not the only thing they need to take into account. It's very plausible, I think, that there are cases where one ought to refrain from ending one's life even though suicide would be the rational move. The moral permissiveness of suicide is, however, outside the scope of our discussion.

Likewise, there might be cases where one ought to self-kill even though it wouldn't be in one's best interest, simply because it's the right thing to do. Oates' self-sacrifice on the return journey of the Terra Nova Expedition might, for example, constitute one such case. Interestingly, such 'benevolent self-killings' would not be captured by our above intention-based definition of suicide³⁹. The same would be true of 'malevolent self-killings'. This is because the motivating wish and intention in such cases are not aimed at death or relief from a life not worth living, but instead towards benefit or harm to others. If a person jumps in front of an oncoming train in order to save an absent-minded child playing on the tracks from being killed, her intention is obviously to save the child, and not to cause her own death. The fact that the saving of the child required the sacrifice of her own life is merely an unfortunate by-product. In this sense, 'benevolent suicides' are so drastically different from the kinds of suicide we've discussed so far that I think we're warranted in keeping the phenomena separate.

39 Cholbi (2021, par. 9) makes a similar point.

You may find this distinction arbitrary. Think, however, of the difference between a murder and a killing in self-defence. Murders are acts that are intended to cause someone else's death. For killings in self-defence, however, intention is aimed at the prevention of harm to oneself or others; the killing part is merely a by-product. In this sense, murders and killings in self-defence are obviously very different acts that belong in different categories, even though both have the same consequences i.e. causing the death of another person.

4.9. Chapter conclusion

We started this chapter by noting, taking inspiration from Brandt, that a person who is considering suicide is making a choice between two future life-courses: one that involves their death at that point (or realistically shortly after), and one where they continue living for an indefinite time longer. The first life-course is static (and thus easy to predict), while the other remains dynamic, that is, open to both intended and unintended changes. In being dynamic, it's also, at least to some extent, uncertain and intangible. If it's more likely than not that one's *continue living*-life-course is going to be so bad that it's on the whole not worth living, suicide might sometimes make sense. Sometimes, however, there might be differences in the stakes involved in the two life-courses such that it makes sense to choose the *continue living*-life-course even though it's more likely than not that this life-course will be on the whole not worth living. Nevertheless, it's clear that, in some circumstances suicide does make sense, at least in principle.

5. Mental illness and irrational states of mind

Over the course of the nineteenth and early twentieth century suicidality was, through developments in psychology and sociology, increasingly considered to be caused by “impersonal social or physiological forces rather than by the agency of individuals” Cholbi (2021, par. 29). Contemporary, there seems to be broad consensus in psychology that the phenomenon of suicide is closely linked to the presence of various kinds of serious psychiatric illnesses, particularly major depressive disorder, bipolar disorder, schizophrenia and substance dependence. This general sentiment has, it seems, trickled down into the general public beliefs and attitudes towards suicide, at least in what one might call the western world: “Most people think that you’ve got to be crazy to kill yourself. Indeed, the very fact that you are contemplating suicide is *evidence* that you’re crazy” (Kagan 2012, 318).

Throughout our discussions so far, we have presupposed that it is at least possible to seriously contemplate suicide with a normally functioning and reasonably clear mind, assessing one’s future prospects for a life worth living with, at least relatively, cold reason. This picture is, of course, challenged by the aforementioned *psychiatric paradigm*. The cases we’ve imagined thus far, then, might reflect real world cases only very rarely, and perhaps never. If it’s true that you’ve got to be, so to speak, ‘crazy’ to contemplate suicide, and if it’s true of crazy persons that they’re not able to reasonably predict their prospects for a life worth living and rationally assess whether they’re more likely to be better off taking the life-course where their life ends sooner rather than later, rational suicide does seem to be rendered impossible. This would remain unchanged, it seems, even if when *is* true that death sometimes comprises a benefit for the decedent; think back to the undeniably unhinged man who opted not to gamble on the basis of his unreasonable belief that there lived an acid-spitting demon in the slot machine. His decision not to gamble was the right one, but to call his choice a rational one seems like a stretch. In the same way, we could imagine a person for whom it is in fact overwhelmingly likely that their future will be worse than non-existence and *at the same time* delirious to the point of being incapable of making rational decisions.

In this chapter I want to look at the link between mental disorder and suicide and consider whether the first precludes the latter. We’ll then consider clouded thinking in a more general sense and pose the same question.

5.1. The empirical correlation between suicide and mental illness

The common man's attitude that you've got to be crazy to contemplate suicide is, of course, not pulled from thin air; a large number of suicidological studies have indicated a strong correlation between suicide and mental illness (though, as we'll see shortly, with significant variance). In recent years, however, it has been suggested that previously well-established numbers may have been amplified in certain types of studies due to methodological faults (more on this later). Furthermore, some newer studies have suggested a significantly weaker correlation between the two phenomena, potentially challenging established truths. I'll give a brief summary of some of the larger studies in the following few paragraphs. Do keep in mind that these studies employ a less nuanced definition of suicide than the one we arrived at earlier, with virtually no emphasis on autonomy or persistence of intention.

A number of psychological autopsy studies⁴⁰ have shown that up to 80-90% of people who kill themselves suffered from some kind of mental disorder at the time the suicide was committed (Goldney and Kapur 2019, 24). Furthermore, it has been claimed that even in the remaining 10-20% of cases where no mental disorder could be established post-mortem, it is likely that the suicider was suffering from some underlying psychiatric illness that was not picked up by the psychological autopsy (Goldney and Kapur 2019, 24).

A 2011 long-term follow-up study with a sample size of 176 347 people⁴¹ found the lifetime risk of suicide in men who had been in contact with secondary mental health services for bipolar disorder to be ~7.77%, ~ 6.67% for unipolar affective disorder and ~6.55% for schizophrenia. For women schizophrenia topped the list at ~4.91% followed by bipolar disorder at ~4.78%. For context, the lifetime risk of suicide in the group of matched controls was found to be ~0.72% among men and ~0.26% among women (Mortensen, Nordentoft and Pedersen 2011, 1058).

Similarly, a large 1998 meta-study found people with affective disorders, alcohol dependence and schizophrenia to have a life-time risk of suicide estimated at 6%, 7% and 4% respectively (Inskip, Harris & Barraclough 1998, 35).

According to a Danish register study examining 21 169 suicide deaths, the PAR⁴² for suicide in persons who have been hospitalized due to a mental disorder was shown to be 40.3% (Goldney & Kapur 2019, 27). Other suicide risk factors like unemployment, disability,

40 Psychological autopsy studies (PA) are studies where information from various personal and clinical sources (e.g., family, friends and healthcare workers) are collected in an attempt to determine the causes of a suicide that has already occurred (Goldney & Kapur 2019, 24).

41 For each of these five matched controls where included.

absence from work due to illness, low-income or age pension had a combined PAR for suicide of 31.4% (Goldney & Kapur 2019, 27).

A 2022 study conducted in the U.S. examining 70 376 male suicides in the period of 2016-2018 that 60% of the suiciders had no *known* mental health conditions. Of course, this ratio may be, at least partly, explained by the health care system's failing to discover people who struggle with mental disorders. (Fowler, et al. 2022, 420)

Some concerns, however, have been raised about these findings. Firstly, worries have been raised against the use of psychological autopsy studies as diagnostic tools. Hjelmeland et al. (2012) states that PA studies are seriously methodologically flawed, arguing that “[i]t is simply impossible to assign a reliable diagnosis of mental disorder to someone by interviewing someone else” (Hjelmeland et al. 2012, 621). If this is the case, serious doubt is cast on some of the most important evidential basis for the common assertion that around 90% of those who commit suicide suffer from mental illness, a historically widely accepted fact in suicidology (Hjelmeland et al. 2012, 606).

Secondly, H. H. Maung (2021) has pointed out that the apparent connection between mental disorder and suicidality is, at least partly, circular, stemming from the fact that suicidality appears among the criteria for diagnosis in many psychiatric illnesses. This issue concerns the basic nature of psychiatric diagnosis under DSM-5 is therefore not limited to the statistics derived from psychological autopsy studies alone. Maung also questions the idea that the link between mental disorder and suicide is not mere correlation but also causal, a sentiment she claims to be prevalent in contemporary psychology (Maung 2021, 346). It's often thought that the establishing of such a causal connection is a simple empirical matter (Maung 2021, 346), verified, for example, by examining whether the set of persons who kill themselves also belong to the set of people with mental disorder diagnoses. Maung challenges this idea by pointing to its partial inherent circularity. The fact that, under the DSM-5, suicidal behaviour appears on the list of diagnostic symptoms for several of the mental disorders typically associated with suicide risk, they claim, shows that the relationship between suicide and mental disorder is, at least to some extent, *conceptual* rather than purely causal (Maung 2021, 353). That the connection is not entirely causal but at least partly conceptual owes itself to a

42 Population Attributable Risk (PAR). PAR measures “[...] the proportion of a condition that may be associated with exposure to a risk factor, or the proportion of the condition that would be eliminated if the risk factor was not present (assuming a causal relationship)” (Maung 2021, 354). That the PAR of having been hospitalized for a mental disorder for suicide is 40.3% means that if this particular risk factor was eliminated, 40.3 percent of the suicides in the examined population would have been prevented (assuming a causal relationship between suicide and mental illness).

“prior implicit judgement that suicidal behaviour is mentally disordered behaviour” (Maung 2021, 354), a judgement that, according to Maung, cannot be justified by their correlation alone (Maung 2021, 365), but stems from value judgements for instance about the harmfulness and unacceptability of suicide (Maung 2021, 359). Borderline personality disorder as well as affective disorders like major depressive disorder and bipolar disorder are examples of mental illnesses where suicidal behaviour appears on the list of criteria for diagnosis. As we saw earlier in this subchapter these specific psychiatric illnesses scored highly on the list for life-time risk of suicide. Let’s take a closer look at one of them to better illustrate the point. Under DSM-5 a person is required to exhibit a minimum of five out of the nine symptoms below to qualify for the diagnosis of borderline personality disorder (BPD). In fact, psychiatric diagnosis in these cases are based on the ticking of the appropriate number of boxes on a list of symptoms. If a person meets the threshold (they tick a minimum of, in this case, five boxes), they, by definition, suffer from the relevant disorder (Maung 2021, 350).

1. Frantic efforts to avoid real or imagined abandonment.
2. A pattern of unstable and intense interpersonal relationships characterized by alternating between extremes of idealization and devaluation.
3. Identity disturbance: markedly and persistently unstable self-image or sense of self.
4. Impulsivity in at least two areas that are potentially self-damaging.
5. Recurrent suicidal behaviour, gestures, or threats, or self-mutilating behaviour.
6. Affective instability due to a marked reactivity go mood.
7. Chronic feeling of emptiness.
8. Inappropriate, intense anger or difficulty controlling anger.
9. Transient, stress-related paranoid ideation or severe dissociative symptoms (American Psychiatric Association 2013, 663).

Notice how number five on the list of symptoms explicitly concerns suicidal behaviour.

While it’s true that suicidal behaviour is not necessary nor sufficient for meeting the threshold for diagnosis (a person who exhibited symptoms 1 though 4 and 6 would qualify) a suicidal person is, by default, *closer* to diagnosis because only four additional symptoms are required in order to meet the threshold (Maung 2021, 351). This could potentially have consequences for studies like the ones we looked at earlier. In the case of psychological autopsy studies, for example, the fact that the study population comprises persons who have *in fact* killed themselves, means that they already ticked one box by admission, and are therefore al-

ready closer to a diagnosis than would otherwise be the case. In other words, a person that, before their suicide, exhibited symptoms 1 through 4 (and therefore did not meet the threshold for diagnosis) would, after having killed themselves, be included in the set of people who had a mental disorder at the time their suicide was committed. Thus, they would be included in the dataset that supports the idea that suicides and mental disorder are correlated. In lieu of a normative justification for considering suicidal behaviour to be mentally disordered behaviour, then, it's easy to see how the data might be mistakenly skewed towards the connection between mental disorder and suicide, at least to some extent — that is, in cases where ticking the 'suicidality' box tips the scale in favour of diagnosis. For these reasons Maung believes that the common sentiment that there is a strong causal relationship between mental disorder and suicide could potentially be partly self-fulfilling and so also partly protected from scientific falsification (Maung 2021, 361).

While Maung's objections certainly call for some amount of uncertainty regarding the claims of studies purporting to quantify the correlation between the phenomena of suicide and mental illness, and also some scepticism about the common assumption that the two are causally connected, the scope of doubt is limited. Perhaps a reasonable position to take is that, while the precise degree of correlation between mental disorder and suicide might be somewhat overestimated, we're probably warranted in saying something like: while most people who suffer from mental disorders don't commit suicide, a large portion of those who commit suicide suffer from some kind of mental illness. What the true extent of correlation is we'll leave to the psychologists to sort out. For our purposes however, it's not vitally important to know whether the rate of correlation is 90%, 60%, 40% — it's sufficient to know that, in a non-trivial amount of cases, suicides seem to happen in the absence of mental illness; that is, it's just not true of all people who commit suicide that they're mentally ill. That's not to say that these suicides couldn't be irrational nonetheless, but it does preclude simply dismissing them as such on the grounds of mental illness.

5.2. Does the presence of mental illness necessarily preclude the capacity to commit suicide rationally?

Furthermore, even if we assumed for a moment that suicide was correlated with mental illness in 100% of cases, it still wouldn't follow that 100% of suicides are performed by persons incapable of suiciding rationally. While I do think it's reasonable to assume some kind of causal relationship between mental illness and suicide, it seems to me that such a causal connection

might be understood, in principle, in at least five distinct ways with very different implications.

First (i), we might imagine mental illnesses to cause suicide more or less directly, by giving rise to involuntary suicidal impulses or compulsions too forceful to resist (much in the same way that piece of dust may tickle the inside of your nose giving rise to an involuntary and uncontrollable compulsion to sneeze). Thinking back to our definition of suicide in chapter 2 (and in particular to the case of Daisy and the deranged neurosurgeon), it's clear that such a case would fail to meet our autonomy condition and, as such, wouldn't be counted as a suicide, but instead a self-killing.

Second (ii), we might imagine mental illnesses to cause suicide through cognitive distortions and the warping of a person's ability to predict the outcomes (i.e. the life-courses following *act* and $\neg act$) of their actions in a reasonably manner. If a delusional man leaped from a skyscraper, fully expecting to be able to defy gravity and dart through the sky all the way to Hawaii for a refreshing coconut water, his *intention* was to fly to Hawaii, and not to cut his life short. In a sense, he made a rational enough choice, the only problem was that the prediction on which he based his choice was completely absurd⁴³. Again, this should remind us of an example from chapter 2 (the case of Oliver the daredevil), and, again, it's clear that such a case should not be counted as a suicide on our intention-based model.

Third (iii), we might imagine mental illness to cause suicide by disrupting a person's ability and/or desire to choose rationally between reasonably predicted future life-courses, that is, making the person almost *essentially* irrational, choosing between life-courses with little to no regard for their own self-interest and well-being, perhaps even wishing for self-destruction for it's own sake, not because he thinks he would be better off dead, but because he simply doesn't care whether he's better or worse off. Such a case does indeed seem to fall into the category of suicide on our intention-based account, and rational suicide does in fact seem impossible in this case. Think back to the example with the man who bet his paycheck with a colleague that he could flip a coin and have it land heads five times in a row. If he knew that he would, in all likelihood, not be so lucky, and also knew that the life-course in which he lost a month's pay would be substantially worse for him than the one where he didn't, he clearly couldn't have acted rationally in proposing the bet.

43 What I mean to say here is that, from our delusional man's perspective, the *jump world* was in fact believed to involve a refreshing drink on a Hawaiian beach, not falling to his death.

Fourth (iv), we might imagine mental illness to cause suicide by disturbing a person's ability to reasonably appraise her prospects for a future worth living. If a person's predictions of the future is severely pessimistically skewed, the life-course in which she ends her life might falsely appear to her comparatively better than the life-course in which she remains living (because she misjudged the value of the latter). Again, such a case does seem compatible with our notion of suicide, and, again, rational suicide does seem impossible here. As we saw in chapter 2, the ability to predict that the life-courses following *act* and $\neg act$ in a reasonable manner does seem to be a prerequisite for rational decisions about them.

Fifth (v), we might imagine mental illness to cause suicide simply by making a person's future not worth living, and, as such, a legitimate reason *for* rather than a direct cause *of* suicide. Consider this selection from the list of diagnostic criteria for major depressive disorder under DSM-5:

1. Depressed mood most of the day, nearly every day [...].
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day.
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day.
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt.
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (American Psychiatric Association 2013, 160-161)

If tomorrow I woke up feeling exhausted, gloomy and miserable, unable to derive pleasure or fulfilment from the activities though which I normally do, had no energy or focus to do anything even if I had wanted to, and, on top of all this, felt like a worthless good-for-nothing worm, well, that doesn't seem like a very pleasant day at all; in fact I think I'd rather sleep though this day much the same way I would rather sleep though the removal of my appendix. Of course, if I found myself in such a state only once in a while, the experience might well be worth *living though*, that is, if the good days made up for the bad ones. If I was to be in such a

state chronically or most of the time, however, I think it's possible that the net quality of my life would be so poor that I might feel it wouldn't be worth living⁴⁴.

That's not to say, of course, that suicide would be rational in this situation, that would depend on what I could reasonably know about the future of the life-course I'm on, and on the other potential life-courses I might actualize. As to the first, research has suggested that over half of cases of major depressive disorder will remit within a year if left untreated (Whiteford et al. 2013, 1569). As to the second, if potentially effective treatments for MDD were available to me, the life-course in which I seek help may have a good chance of being better than the life-course in which I cease to exist. Such alternative life-courses would obviously have to be exhausted first (think back to the man who had his feet stuck in bear-traps). In treatment-resistant and chronic cases, then, suicide might be rational not *despite* mental illness, but precisely *because* of it.

5.3. Does living a life not worth living necessarily preclude the capacity to commit suicide rationally?

Of course, it's not only mentally ill people who can fall victim to irrational states of mind or clouded thinking; under the right conditions we all can. It may be tempting, then, to assume of people whose existence is sufficiently bad and future prospects sufficiently grim to the point that death seems to them preferable to continued existence that they're, almost by admission, not in a mental state conducive to clear and rational thought. In other words, when you're in a position where suicide *could* be rational, you're simultaneously also in a position where you're not *able to* decide between life-courses rationally. There's definitely something to this worry; if I were laying on the beach solving sudoku puzzles whereupon a crab pinched my big toe and refused to let go, my ability to decipher the sudokus would certainly diminish. If I were playing poker and got a break-up text from my girlfriend, my ability to play well and read my opponents would certainly suffer. If I was in terrible physical or emotional anguish, then, I might not be able to think clearly enough about my future prospects and available alternative life-courses. As such, it might be the case that the rational thing to do in this situation is to recognize the possibility that one's opinion is formed by a mind prone to cognitive error and simply suspend judgement. Kagan raises the same worry (2012, 341) and goes on to

⁴⁴ Research has indeed found that people who suffer from serious mental illnesses report a lower subjective quality of life compared to that of the population at large, not only directly as a result of the primary symptoms of their illness, but also due to secondary factors like the side-effects of medications, loneliness, unemployment and economic instability (Hewitt 2013, 363).

lay out a compelling thought experiment to challenge it: *the surgery analogy*. It goes something like this:

You suffer from some terribly painful and severely debilitating medical condition. There is a very small (0.001%) chance that you'll recover spontaneously, but your odds are far better if you go through a surgical procedure. The surgery has a 99.9% success rate, but alarmingly also a 0.01% death rate. You're faced, then, with a choice between life-courses: (a) have the surgery and (b) refuse the surgery. If you go with life-course (a) there's a 99.9% chance that you'll be cured of your condition and (presumably) go on to live a life worth living, and a 0.01% chance that you die during surgery. If you go with (b) there's a 99.9% chance that you won't get better and a 0.001% chance that you spontaneously do (2012, 342).

Kagan goes on to argue that, surely, having the operation has to be the right move here. The odds that it will be successful and grant you a future worth living are far greater than the odds that it will kill you and thus rob you of your very small chance of spontaneous recovery. No matter what you choose you're making a bet, and, surely, it makes sense to play the better odds. However, if we accept that being in great pain and emotional distress is preclusive of any rational decision-making whatsoever, it seems you *couldn't* rationally choose to have the surgery, and likewise, that you couldn't rationally choose *not* to have the surgery. This, Kagan thinks, can't be right; although you're in pain and emotional turmoil, it just can't be the case that a person could *never* choose rationally to have the surgery, and the same then, has to be the case for suicide (Kagan 2012, 343-4). You should, however, recognize that your situation might entail a potential proneness for errors of judgement, and, as such, not make your decision in haste (Kagan 2012, 344).

I think, however, it's possible to imagine rational decision-making even in acutely precarious situation where you don't have time to think and think again. Imagine that you and your significant other are roused from sleep one night by the sound of a blaring fire-alarm. Thick black smoke fills the room and the only thing you can make out are the bright slivers of light emanating from the slits between the door and its frame; the living room is positively engulfed in flames. This leaves the window your only possible route of escape. Unfortunately, your flat is on the third floor and jumping is sure to result in at least a few broken bones. You hunker down on the floor where the air is still somewhat breathable to consider your options. A couple of minutes go by and there's still no sign of rescue. Meanwhile, the smoke is growing thicker and it's getting progressively harder to breathe; you'll probably pass out in a minute or two. This leaves you and your SO faced with a choice between life-courses. (1) you

can sit tight and hope that the fire-fighters will reach you *just* in time, or (2) you can jump from the window, probably breaking a few bones in the process but saving your lives. Surely, we can all agree that jumping would be the right move here. What then, if, upon telling your SO that you think the two of you should jump she responded by saying “well, we’re clearly in a severely distressed mental state prone to tunnel vision and cognitive error, I think we should suspend judgement seeing as we can’t trust our own reasoning to be rational”. This, you’ll probably agree, would be an absurd thing to say in the situation you found yourselves in. Things would get even more absurd if you were awoken in a hospital bed two days later, both feet in casts, by a philosopher standing by your side loudly professing that your decision to jump simply *couldn’t* be considered rational on account of the adrenaline coursing through your veins, the pain of the smoke in your lungs and your failure to consider every conceivable alternative option.

To be sure, even though I would call the decision to jump from the window sufficiently rational for its context, it does seem reasonable to raise the requirement for rational decision-making in situations that aren’t as acute. If we think back to our previous virus example⁴⁵, you should certainly take you time to discuss your situation with doctors, family and friends to make sure that you’re not making some error of judgement thinking that the net quality of your future will most likely not be worth living, and that the stakes are such that it makes sense for you to end you life. Of course, it only makes sense to seek the guidance of others if they’re open to the idea that life sometimes isn’t worth living and that a person can be benefited by dying. Think of it in this way: let’s say you’re considering taking out a big loan from the bank to buy a house, but you know very little about the housing market and the state of the economy in general. Sometimes, the conditions are such that taking out a loan is likely to be a good move, and sometimes not. It makes sense, then, to seek the guidance of an advisor (that is, an epistemic superior in questions regarding finance and the housing market). If, however, the advisor strongly believes that no one should *ever* put themselves in debt, he’ll always advise against taking out loans, no matter the conditions of the market. Clearly, the opinion of this advisor is of little value.

45 Where you’re infected with a virus that makes your existence so bad it’s not worth living. You may spontaneously recover at some random point in time, but only 40% of those infected do.

5.4. Chapter conclusion

We started this chapter by pointing out a worry: even though it might be the case that a person might sometimes be benefited by dying, and even though it might in principle sometimes make sense for her to cut her life short, this wouldn't count for much if person's who consider suicide *necessarily* lack the ability to make the decision to kill themselves rationally. We started by looking at statistics on the correlation between suicide and mental illness, and found that a not insignificant number of suicides are performed by persons not considered to be mentally ill. We also noted that, even in cases where mental illness is present, it's not necessarily true that rational suicide is rendered impossible. In fact, mental illness might be construed as a *reason for* rather than a direct *cause of* suicide. We then turned to the potential of clouded thinking in individuals whose lives are so bad that they aren't worth living, and concluded, along with Kagan, that rational suicide does not appear to be precluded *necessarily*.

6. Conclusion and returning to the suicide-*nullvisjon*

In conclusion, I hope to have made convincing cases for the propositions that: (1) death is a benefit for a person whose future would be worse than non-existence; (2) a person can, at least in principle, rationally choose to commit suicide even though she can't be *entirely* certain that the life-course in which she ceases to exist by her own hand will be comparatively better to the life-course in which she continues living; (3) it's empirically not true that all people who commit suicide are mentally ill, and even it was, it's not necessarily true of mentally ill people that they're incapable of committing suicide rationally; and (4) the pain and emotional distress presumably accompanying situations where suicide *would* make sense, does not necessarily preclude the ability to suicide rationally. In sum, our take-away should be that rational suicides are, in fact, possible. The problem with the suicide-*nullvisjon*, then, should be plain to see; it professes a normative ideal that simply isn't justified. It sometimes *is* the case that death is preferable to life, and it sometimes *is* the case that persons can make the decision to end their lives rationally.

References

- American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: American Psychiatric Association.
- Brandt, Richard B. 1975. "The morality and rationality of suicide". In *A handbook for the study of suicide*, edited by Seymour Perlin, 61-76. New York: Oxford University Press.
- Cholbi, Michael. 2021. "Suicide". *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). Edward N. Zalta (ed.). URL = [<https://plato.stanford.edu/archives/win2021/entries/suicide/>](https://plato.stanford.edu/archives/win2021/entries/suicide/).
- Critchley, Simon. 2021. *Notes on Suicide*. London: Fitzcarraldo Editions.
- Epicurus. 1983. *Letter to Menoecus*. In *The Oxford Book of Death*. Edited by, D. J. Enright, 9 Oxford: Oxford University Press.
- Fairbairn, Gavin. 1995. *Contemplating Suicide: The Language and Ethics of Self-harm*. Social Ethics and Policy Series. London: Routledge.
- Feldman, Fred. 1991. "Some Puzzles About the Evil of Death." *The Philosophical Review* 100, no. 2 (1991): 205-27. DOI: 10.2307/2185300.
- Fischer, J. M. 1997. "Death, Badness, and the Impossibility of Experience" *The journal of Ethics* 1, no. 4 (1997): 341-353.
- Fowler, Katherine A., Mark S. Kaplan, Deborah M. Stone, Hong Zhou, Mark R. Stevens, and Thomas R. Simon. 2022. "Suicide Among Males Across the Lifespan: An Analysis of Differences by Known Mental Health Status." *American Journal of Preventive Medicine* 63, no. 3 (2022): 419-22.
- Goldney, Robert D. and Navneet Kapur. 2019. *Suicide Prevention*. Oxford: Oxford University Press.
- Helse- og omsorgsdepartementet. 2020. "Regjeringens handlingsplan for forebygging av selvmord 2020-2025." <https://www.regjeringen.no/no/dokumenter/regjeringens-handlingsplan-for-forebygging-av-selvmord-2020-2025/id2740946/>
- Hewitt, Jeanette. 2013. "Why Are People with Mental Illness Excluded from the Rational Suicide Debate?" *International Journal of Law and Psychiatry* 36, no. 5-6 (2013): 358-65.
- Hill, D.J. 2011. "What is It to Commit Suicide?", *Ratio*, 24: 192–205.
- Hjelmeland, Heidi, Gudrun Dieserud, Kari Dyregrov, Birthe L. Knizek, and Antoon A. Leenaars. 2012. "Psychological Autopsy Studies as Diagnostic Tools: Are They Methodologically Flawed?" *Death Studies* 36, no. 7 (2012): 605-26.

- Hume, David. 2005. *On Suicide*. London: Penguin Books.
- Inskip, Hazel M., Clare Harris and Brian Barraclough. 1998. «Lifetime Risk of Suicide for Affective Disorder, Alcoholism and Schizophrenia.» *British Journal of Psychiatry* 172, no. 1 (1998): 35–37. DOI: 10.1192/bjp.172.1.35.
- Johansson, Jens. 2012. “The Timing Problem.” In *The Oxford Handbook of Philosophy of Death* edited by Bradley, Ben, Feldman Fred and Johansson, Jens. Oxford University Press USA 255-273.
- Kagan, Shelly. 2012. *Death*. New Haven: Yale University Press.
- Kolodny, Niko and John Brunero. 2023. "Instrumental Rationality." *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/sum2023/entries/rationality-instrumental/>.
- Lee, Andrew Y. 2022. "The Neutrality of Life." *Australasian Journal of Philosophy* Ahead-of-print, no. Ahead-of-print (2022): 1-19. DOI: 10.1080/00048402.2022.2033284
- Maung, Hane H. 2022. “Mental Disorder and Suicide: What’s the Connection?” *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, Volume 47, Issue 3, June 2022, Pages 345–367, <https://doi.org/10.1093/jmp/jhab015>
- McMahan, Jeff. 1988. "Death and the Value of Life." *Ethics*, 99, no. 1 (1988): 32-61.
- Nagel, Thomas. 2012. *Mortal Questions*. Canto Classics. Cambridge: Cambridge University Press.
- Nagel, Thomas. 1970. «Death.» *Noûs* 4 (1): 73-80.
- Nordentoft, Merete, Preben B. Mortensen and Carsten B. Pedersen. 2011. "Absolute Risk of Suicide After First Hospital Contact in Mental Disorder." *Archives of General Psychiatry* 68, no. 10 (2011): 1058-1064.
- Paterson, Craig. 2003, “On Clarifying Terms in Applied Ethics Discourse: Suicide, Assisted Suicide and Euthanasia”, *International Philosophical Quarterly*, 43(3): 351–358.
- Rosenbaum, Stephen E. 1986. “How to Be Dead and Not Care: A Defense of Epicurus.” *American Philosophical Quarterly* 23, no. 2 (1986): 217-225.
- Scarre, Geoffrey. 2007. *Death*. London: Taylor & Francis Group.
- Solberg, Carl T. 2019. “Epicurean Challenges to the Disvalue of Death.” In *Saving people from the harm of death*, edited by Espen Gamlund and Carl T. Solberg, 91-104. New York: Oxford University Press.

Whiteford, H. A., M. G. Harris, G. McKeon, A. Baxter, C. Pennell, J. J. Barendregt, and J. Wang. "Estimating Remission from Untreated Major Depression: A Systematic Review and Meta-analysis." *Psychological Medicine* 43, no. 8 (2013): 1569-585.