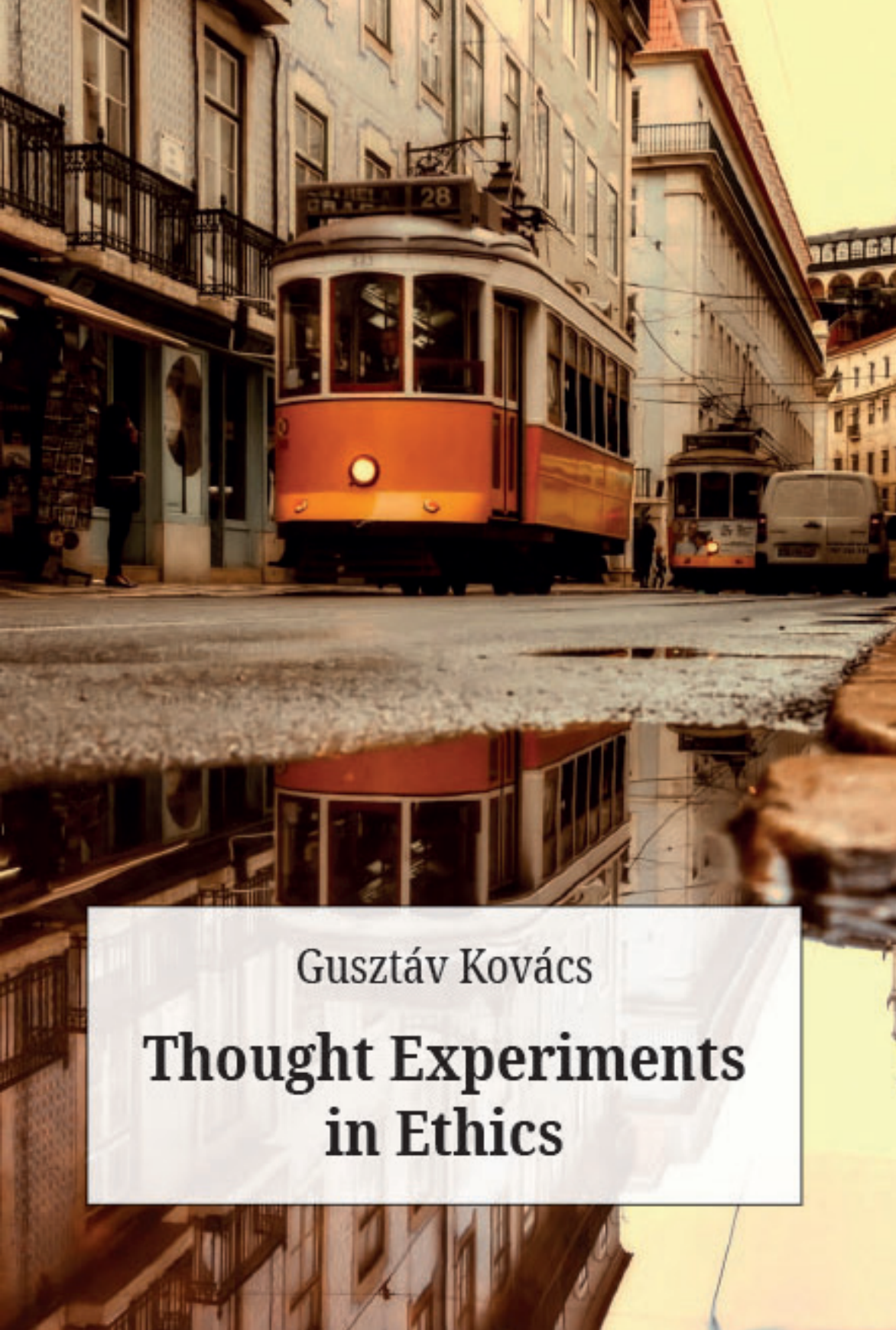


Gusztáv Kovács

# Thought Experiments in Ethics

Gusztáv Kovács

# Thought Experiments in Ethics

*FOR DOMI*

Gusztáv Kovács

# Thought Experiments in Ethics

Episcopal Theological College of Pécs
Pécs

Published by
Episcopal Theological College of Pécs
Pécsi Püspöki Hittudományi Főiskola
11 Hunyadi János St, Pécs 7625, Hungary
www.pphf.hu



ISBN 978-615-5579-28-8

Press: Molnár Nyomda és Kiadó Kft.
Director: Csaba Molnár
Layout: Attila Jakab

# CONTENTS

# PREFACE

I remember the moment my good friend and colleague, Attila Szücs, asked me point-blank: "So, who would you choose? Tomoceuszkakatiti or Gyugyu?" We were driving back home from Vienna after giving a seminar on reproductive ethics, and the image of the road curve where he posed this question is still vivid in my mind. It has become a flashbulb memory, and the puzzle about the tyrant and the slave continues to gain momentum in my mind. Moreover, it has proven to be a key element in the genesis of this book.

The question above is only one of many that I have used in teaching ethics over the last decade.[1] They have proven to be useful tools not just for illustrating questions in ethics, but also for shaping the moral thinking of students and providing them with a chance to learn something new about themselves. The solid-seeming ethical ideas of high school and college students have been challenged by the Trolley Problem, the Experience Machine, or the Violinist Thought Experiment, leading to numerous fruitful conversations about what is right and wrong. All of these discussions, including some alternate versions of standard thought experiments, have contributed to the formulation of this book.

Another important source of motivation was the interest of my colleagues who came to my lectures, seminars, and con-

ference presentations on diverse topics in ethics, and have provided valuable feedback on the thought experiments I regularly use to help the audience join the discussion in a deeper way. I am especially grateful to my colleagues beyond the borders of Hungary who gave me the opportunity to present my ideas in diverse cultural settings. I am deeply indebted to Sigrid Müller, the head of the Department of Systematic Theology and Ethics, who was a wonderful host during my visits to the University of Vienna and inspired me to think about intuitions in a more critical way. Piotr Morciniec from Opole University fired up my enthusiasm to make teaching more playful and appealing to the minds of my students. My annual stays in Poland and his regular visits to Pécs have provided us both with the opportunity to develop the occasional classroom use of thought experiments into a more encompassing academic venture. Our brotherly conversations have been important sources of inspiration. The invitations from Roman Globokar to the University of Ljubljana and from Dominik Opatrný to Palacký University in Olomouc were ideal opportunities to test the ideas formulated in this book in an international context.

Particular thanks are given to Noémi Najbauer (University of Pécs) who worked tirelessly to polish the language of this volume. In her person I have not only found an excellent proofreader, but, as her comments show, the first genuine reader of the book. I also would like to add a word of thanks to Emma McDonald (Boston College), András Mészáros (St Patrick's College, Maynooth) and Dominic Whitehouse OFM who generously gave of their time and energy to revise chapters with a tight deadline.

This book is published as part of the work of the MTA-PPHF Religious Education Research Group,[2] where I am especially indebted to Ottilia Lukács, Katalin Asztalos, and István Csonta for converting my philosophical ideas into surveys conducive

to empirical research. Sponsors such as the Hungarian Academy of Sciences and Renovabis Foundation, and the scholarships provided by CEEPUS and Erasmus mobility programs, provided generous financial support for the research and the writing of the book. The constant encouragement of Archbishop György Udvardy made it possible for me to complete the text while attending to my duties as the rector of the Episcopal Theological College of Pécs.

Above all, I want to say thanks to my family who have supported and strengthened me in a special way throughout the project.

My hope is that this book will bring abstract philosophical questions closer to the reader and help professors and teachers to work with a broader concept of ethics, reaching out not only to the minds, but also to the hearts of their students.

Kovács Gusztáv

Episcopal Theological College of Pécs

## Chapter I

# THE STORY IN YOUR HEAD: TOMOCEUSZKAKATITI AND GYUGYU

One of the most memorable conversations in Hungarian filmography takes place in the 1976 film *The Fifth Seal*.[1] The scene is set somewhere in Budapest in the autumn of 1944, when Hungary was already under German occupation and the rule of The Arrow Cross Party[2]. Five men gather in a pub somewhere in the capital. They are all simple people whose aim is to survive the war: Gyurica, the cynical watchmaker; Kovács, the deeply religious and uncorrupted carpenter; Király, the snobbish bookseller; and Béla, the money-hungry, brawling tavern-keeper. They are later joined by Károly Keszei, the crippled photographer just home from the front. The warm and cozy pub is a symbolic place for the simple but emblematic characters; it provides them with the sense of safety. It's a safe haven amid a barbarous world, where they can talk freely and sidestep the callous course of history. Still, the sounds of terror and tyranny filter in from the outside world and disturb the conversation around the table.

The friends are engaged in a discussion about the perfect way to prepare brisket. The starting point of the conversation

is the bargain made by the bookseller who managed to secure a portion of this rare delicacy in exchange for an album of paintings by Hieronymus Bosch. The conversation is interrupted by the watchmaker who reveals the subject of his daydreaming while sitting and gazing at the ceiling: "I don't know whether I should be Tomoceuszkakatiti or Gyugyu"![3]

Seeing the lack of understanding in the eyes of his friends, the watchmaker unmasks the identity of the two mysterious people behind these names, and also reveals his reasons for choosing between them. Tomoceuszkakatiti is the tyrannical ruler of the incredibly wealthy island named "Lucs-Lucs". Gyugyu is a slave. The tyrant enjoys all the benefits resulting from his position and the wonderful attributes of the island. He treats Gyugyu with extreme cruelty: having his tongue torn out when he dares to smile; taking away his daughter and exploiting her sexually; lopping off his wife's nose and putting out her eyes. The humiliated slave finds consolation in the thought that his conscience is clear: he has done others no harm but merely suffers cruelty at their hands. The twist in the story is that Tomoceuszkakatiti, the tyrannical ruler who humiliates and executes people without any particular reason, considers himself the most decent person on earth. He was raised in accordance with the morals of his age and does not perceive his deeds to be evil according to his conscience.

After setting the imaginary scene the watchmaker challenges his friends to make their choices:

> "Now, you have five minutes, Mr. Kovács, to decide whether you want to be Tomoceuszkakatiti or Gyugyu!"
> "How come five minutes?" asked the carpenter looking at him.
> "Just as I said! After the five minutes are up, you die,

> and ten seconds after that, you will be resurrected either in the body of Tomoceuszkakatiti or Gyugyu. Do you understand now? Make your choice according to your conscience!"[4]

## WHAT CHOICE WOULD I MAKE?

Viewers are probably pestered by the hypothetical question long after watching the film. What choice would I make? Shall I be Gyugyu, the blameless but humiliated slave? Or Tomoceuszkakatiti, the tyrant crippling innocent people, all in good conscience? The company gathered at the pub is just as unsettled by the question. After leaving the pub, they struggle all night trying to find their own personal answers. There is only one of them, the crippled photographer, who insists on choosing the fate of Gyugyu. Still, he is the one who reports the others to the police, causing them to end up in the prison of the Arrow Cross leader embodying Tomoceuszkakatiti. Thus, the film turns the parable into reality. Reality functions as the test of the parable too, due to the disjunction between the characters' answers to the parable and their individual actions throughout the story. Those members of the company who shied away from the fate of Gyugyu in their thoughts,

> are able to face death when things turn serious, and when there is a need to prove their morality—even Király, the weakest character. The reader needs to find an explanation for the dichotomy of theory and practice when seeking for a proper understanding of the story. The question raised by Gyurica was first answered incorrectly by all actors, namely in the abstract-theoretical situation. Keszei opted for [the char-

> acter of] the slave but reported his friends the next day. The three craftsmen choose life as a tyrant, but on the third day they turn out to be incapable of hitting the dying Communist. Thus it is in a concrete situation that the true character of men is revealed. He who sees himself as a potential hero in an abstract ethical debate, turns out to be ignominious. Those who were cowards in theory turn out to be heroes in reality[5]

The question of the relation between abstract thinking and reality hereby becomes one of the key questions of the novel. The novel does not provide a clear-cut answer but rather demonstrates the complexity characterizing the relationship between thought, action and context.

## THE FIFTH SEAL AS A THOUGHT EXPERIMENT

The novel discussed above is of enormous importance to our train of thought, since it gives us insight into how ethical thought experiments function and provides an example with all the distinguishing features. Moreover, the novel presents a more comprehensive definition of thought experiments than do traditional descriptions, in that it makes the central parable step outside the bounds of its narrow world. We do not merely hear the parable, but also see its context, development, impact, and consequences.

Certainly, we could treat the parable about Tomoceuszkakatiti and Gyugyu as an individual text. However, in that case it would be a simple story, not a thought experiment. It is not the parable, but the tension between the parable *and* the context, which makes the thought experiment. It makes no sense to speak about thought experiments without a context, since –like every other

experiment – they work in and say something about reality. Thought experiments impact those who listen and understand the questions they raise, influencing their thinking, their lives, and their personal or material relations.

Thus, the novel turns out to be exceptional, since it demonstrates the life of an ethical thought experiment from the telling of the parable, through the raising of its central question, to its functioning in its context. The latter is exactly what debates about thought experiments traditionally gloss over. Such discussions concentrate mostly on logical structure, the question of coherence, and epistemological status, while the role of the context of thought experiments has garnered less attention. It is precisely their contextual relevance, however, which distinguishes them from simple, descriptive texts.

Here, at the beginning of the book, I would like to call your attention to four decisive aspects of thought experiments. All four aspects clearly demonstrate how important it is to take contextual embedding into consideration. They fuel thought experiments, and without them the thought experiment would be no more than an empty car shell. The engine of the thought experiment can only be brought into motion with the help of the right context.

The first aspect is that there is a need to justify the use of thought experiments. A thought experiment is dead if it fails to induce discussion about its own rightfulness. The second is the explicitly provocative means by which the parable captivates the audience. If the inner judgment-making mechanism of the listener is not brought into motion, the thought experiment has not achieved its goal. The third is that the source of this power is the tension created by limiting the set of possible responses. The fourth aspect is the tension between theory, laid out in the form of a parable, and reality, presented in the actual case to be solved.

## "... IT'S MORE SERIOUS THAN YOU'D THINK!"

The basic question about thought experiments is whether they are worthy of consideration in the first place. Is it a worthwhile endeavour to imagine a nonexistent world such as Lucs-Lucs, and to torment ourselves with a hypothetical dilemma featuring an imaginary slave and a tyrant? Am I to expect an answer or a solution to anything at the end of the thought experiment, or am I simply engaging in intellectual hair-splitting? These questions cannot be answered by the parable itself. To find the answers we seek, we must study the effects of the parable.

It becomes clear in the novel that the parable moves the different protagonists in different ways: the crippled soldier identifies with Gyugyu and is miffed at the others for refusing to take his choice seriously; the uncorrupted carpenter tosses and turns in his bed all through the night because he does not have the courage to take the fate of Gyugyu upon himself; the bookseller escapes into the world of eroticism at his mistress's apartment. It is only the watchmaker who can symbolically pass over the parable as he goes home to the Jewish children he is hiding. It would seem that the parable about the Island of Lucs-Lucs motivates each of the protagonists to a profound extent.

There is a debate about the meaning of the thought experiment right after the watch-maker raises the question. Some consider it nonsense, others a bad joke. Still, the more they refuse to answer the challenge, the deeper it starts to penetrate their minds. They cannot slip the gimmick. It is the photographer who points out that the challenge the thought experiment poses cannot remain unanswered: "I believe that Mr. Kovács was right when he said that this was a very serious thing. I must add: it's more serious than you'd think!"[6]

But what constitutes a serious thought experiment? What makes the question unavoidable? The answer does not lie in the parable itself, but in the mesmerizing effect it has on the listener.

## EXISTENTIAL FORCE

We take thought experiments seriously because we feel that they reveal something hidden about ourselves. This particular parable helps us discern whether we are merciless tyrants or uncorrupted slaves at heart. The carpenter's wife, when hearing her husband recount the parable, gives the following answer with shocking serenity:

> I think I can make the choice since I've so much misery in my life that it's enough for three people (...) This is why I can make the choice! Rather any misery... Unfortunately, I know it well. But that Tiktak or what's his name, not him, I would rather die![7]

Her husband hesitates, however, and finds himself unable to make a choice. Kovács cannot speak his choice out loud; only in private prayer does he opt for the figure of the tyrant. They both feel that the answer would reveal something essential about them. Something the wife is able to face, but that her husband has not been able to cope with so far.

The existential force of the parable can be felt as soon as it is told, and those who have heard it cannot 'unhear' it ever again. One cannot act as if nothing has happened. The parable begins its work in those who have heard it, and even if they put off responding to the challenge it poses, they cannot remain neutral. This is why it is more than a simple theoretical challenge leaving listeners morally unscathed.

The above feature of the parable is highlighted by the following lines:

> "I'm asking you, Mr. Kovács! Imagine that you will die shortly and be raised immediately thereafter. You will become Tomoceuszkakatiti or Gyugyu according to the choice you make now."
> "Is it a game then?" asked the carpenter.
> "Very much so!" said Gyurica. "All that is real is the existence of Lucs-Lucs. Consequently both Tomoceuszkakatiti and Gyugyu exist, and you have to choose between them. Beside that everything is a game..."[8]

According to these lines a thought experiment is a game that extends well beyond the framework of a game. Of course, participants won't really enter the game and find themselves reincarnated in the person of the slave or the tyrant. Nobody expects the carpenter to really die within five minutes and rise again as Tomoceuszkakatiti or Gyugyu. The expansion beyond the framework of the game means that one cannot step away from the answer he gives to question, but remains bound to it, carrying it over into his life. Just as the winner takes his pride and gratification home from the sporting event and the losing side its shame and frustration, the person answering the challenge of a thought experiment will either be burdened or relieved after facing himself. The analogy between thought experiments and games is possible because they both work only with the help of well-defined rules.

## THE RULES

Games, especially sports, have a strong grip on people because they have set rules which all participants acknowledge and to which they must subscribe. This gives games their power. One cannot taste true victory if he won by cheating. (Of course, the exception is someone who joined the game for hidden purposes, e.g. if he was only interested in the reward, but not in the game itself.) Thus, we too must observe the rules if we wish to participate in thought experiments.

There is a particular rule of thought experiments that is described in the novel: the conscience of the slave and of the tyrant is clear. They are not burdened by a sense of guilt. There are different reasons for the absence of guilt, however: the tyrant senses no guilt because he was raised according to the morals of his time, while the conscience of the slave is unburdened because he knows that he is the one being treated unjustly and not the one oppressing others. It is for this reason that the parable presents a dilemma. From a certain aspect, the choice is to be made between two equal options where all secondary features (pleasure, wealth and power, or poverty and humiliation) do not matter. Yet these secondary features are precisely the ones that would profoundly influence one's decision under normal circumstances. The rules are set, however, and the choice must be made between the given options, as highlighted in Gyurica's warning: "Tell me: in what form do you wish to be raised, as a tyrant or a slave? Tertia non datur! (There is no third option!)"[9]

But rules do not only function as obstacles.[10] The rules of the game open up a space for creativity without which the game would fail to fulfil its purpose. A game cannot emerge out of chaos. Boundaries enable creativity. This is also true for thought experiments, since it is the well-defined rule creating the dilem-

ma situation that brings the audience into motion. However, the presence of creativity is not self-evident but depends on the participants and the context. The novel does a good job of demonstrating how protagonists are paralysed by the parable. They find themselves unable to bring a new element into the story. They fail to rethink the question in a decisive manner.

But creativity can emerge not only at the level of thinking, but also at the level of reality. We face the provocative challenge not by rethinking it, but by confronting it in reality.

## THOUGHTS AND REALITY

The novel proves to be perfectly designed to demonstrate the functioning of thought experiments, since it does more than present the parable and the *theoretical* answers given by the listeners. It is the participants' *actions*, which go beyond the theoretical debates, that are of vital importance in the novel. The pub functions as a well-defined space for theoretical debates, since nothing needs to be done around the table other than thinking and disputing. However, the protagonists have to face concrete actions and their consequences in the world outside the pub. This is where the value of all that has been discussed within the protective and comfortable environment of the pub is determined.

Still, there are two other, even more irreconcilable poles beyond that of the pub and the world outside: namely the island of Lucs-Lucs and the prison of the Arrow Cross Party. The two places are both alike and different. If we were to neglect one or the other—the parable or its context, the perfect island or the dreadful prison—the novel would lose its core message: the radical question of the relation between ideas and reality, theory and practice, daydreaming and real life.

In contrast with the simple, well-designed story, reality reveals itself to be much more complex and unpredictable, since neither the actions of the protagonists, nor the turn of historical events can be foreseen. The protagonists may make unusual decisions, sometimes in direct opposition to the answer they gave in response to the parable. Moreover, the string of events will take unexpected turns for which the protagonists may be unprepared. The afterlife of the parable is secured by its compatibility or incompatibility with reality. In this case, no perfect and sound answer can be given to the question raised.

Those who have read the novel or seen the film can certainly recall conversations they themselves have had which resemble the exchange of the five men in the pub, dialogue anatomizing the deepest questions of life. One's adolescent years, especially, are a time for profound and life-changing theoretical conversations. When recalling these, we realize that we are no strangers to the thought experiment.

## WHAT WOULD YOU DO...?

Thought experiments are present in our everyday conversations, though they mostly take a more casual form. They usually begin with the well-known formula: "What would you do if...?" The questioner is mostly seeking advice or attempting to clarify his own position with the help of the other person. Somebody asking "What would you do if your car broke down in the middle of the road?" probably wants practical advice. He wants to know how he could or should solve such an unfortunate situation. He is interested in the factual or practical knowledge of the person he is questioning. He expects answers such as "I would call the A.A. patrol", "I would ask somebody to help me push the car to the side of the road", or "I would

start the engine at full throttle". Consequently, the question and the answers received seek to solve an objective problem and aim at a sensible solution to the situation. The imaginary situation can certainly be expanded to produce more comprehensive and differentiated answers. It makes a difference if the car breaks down in the middle of the Great Victoria Desert or Ferenciek Square, the busiest junction of Budapest. The more precisely we describe the hypothetical situation, the more differentiated the solution at which we arrive.

This is also true for the ethical "What would you do...?" questions. They differ from the practical questions in one significant way: the answers they seek are not simply practical, impersonal, and objective, but are personal and concern the existential aspect of the parties in conversation. The often asked question "What would you do if you won the lottery?" can show for example what a person would do with his life in an imaginary situation, namely if his material wealth were to increase significantly. If taken seriously, such everyday thought experiments can reveal a lot about one's personality, e.g. how a person relates to money, to other people, and to his situation in life. If somebody planned to give up his profession after winning the lottery, this would reveal much about the way he relates to his current job. The answer might also shed light on the person's relationships with others by revealing which relationships he would maintain or walk away from. It is also true here that the more elaborate the description of the imaginary situation is, the more detail might be revealed about the respondent. Further questions might arise, such as "If I had more money, would my old friends stay my friends?" or "How would such a great sum change me?" Details might turn initial hope - since most people associate the scenario of winning the lottery with hope - into fear, i.e. "What if money changes me in a way that I lose my personal relationships?"

Thus the imaginary future scenario might illuminate people's intuitions about who they want to become in the future, and who they are at present.

But it is not only future scenarios which can be discussed using "What would you do if...?" questions. Similar types of questions can also help in evaluating past events. The question "What would you have done if...?" can help explore the nature of a past action, e.g. whether it was practical to call the A.A. patrol to the broken car.

But can these hypothetical questions induce empathy or encourage listeners to make moral judgments? A student who was treated unjustly at an exam expects empathy when he asks: "How would you have felt in my situation?" Or he might ask his friend to make a moral judgment when he questions him about whether it was morally acceptable not to raise a complaint against the unjust examiner. This last question, "Did I do the right thing when...?", is probably the most common everyday thought experiment. When asking this question, we mostly expect persons significant to us to give us confirmation, less often guidance, concerning the morality of our actions.

Although the thought experiments we use in our everyday conversations might be diverse, there are two features they all share. From the perspective of the listener, they describe an imaginary situation - or a real situation in the past - which the listener is not actually part of. From the point of the speaker, such thought experiments are useful for getting information about what the listener would do in that particular situation, i.e., how he would feel and how he would judge the state of affairs. In order to induce the listener to answer the question adequately, certain conditions must be met.

## THE PRAGMATICS OF THOUGHT EXPERIMENTS[11]

In order to successfully map the preconditions of a thought experiment, we must consider the role of thought experiments in communication. The aim of moral thought experiments is not only to tell a good story, but to make the audience reconsider their moral beliefs and change their system of moral judgements by hearing, or 'colliding' with, the parable.

Thus, thought experiments need to be powerful: they must induce indignation, disgust, or approval, even ovation in certain cases. Without a dramatic effect, there is no well-functioning thought experiment. The goal is to provoke target persons to arrive at a certain moral judgement. This can be accomplished by proposing an imaginary dilemma situation or by presenting the ambiguities of a seemingly clear case. The story of Tomoceuszkakatiti and Gyugyu is a good example of both. This story gains its force from neutralizing the conscience of the tyrant. The traditional criterion of evil - that it was done "knowingly and intentionally" - is thus removed, and the audience find themselves in a dilemma situation.

However, there are certain conditions that determine the effectiveness of a parable. The first is a correct understanding of the text. The precondition for understanding the parable of the runaway trolley is a basic knowledge about trolleys and rails, and the consequences of being run over by a trolley in general. If this knowledge is missing, the audience cannot understand the story, which in turn fails to bring about the expected result. However, the case we have been discussing does not just concern factual understanding, but also explicit moral understanding. The audience must move beyond a factual understanding of the story and be able to make a moral judgement concerning it. If there is somebody for whom wealth does not mean anything, that person will not be able to understand

the challenge of the imaginary scenario about winning the lottery. Similarly, someone who considers the actions of the tyrant Tomoceuszkakatiti to be good deeds - if such a person even exists - could never understand the ethical challenge posed by the parable. The story will fail to interest him or to shape his moral beliefs. Such a person would simply shrug it off, saying, "Nice story!" Thought experiments only work if the audience agrees with the speaker concerning certain anticipatory moral judgements - supposing that the speaker did not intend the experiment as a hoax, but as an exercise to be taken seriously. In our case, the audience must be in agreement that physical violence, humiliation, sexual exploitation, and slavery are fundamentally bad. Furthermore, they must agree that human beings are capable of making correct moral judgements. The audience must be of the conviction that human beings are capable of recognizing the difference between good and evil. As a next step, thought experiments also require listeners to view moral dilemmas as comparable by analogy. This is important because one of the important elements of thought experiments is the analogy between the imaginary scenario and a real situation.

In conclusion, thought experiments only work if certain ethical preconditions are fulfilled. As we will see later, most of the parables constituting thought experiments are quite improbable, many of them belonging to the category of science fiction. It is not enough to simply "decode" the parable. There is also a need to recognize the meaning and value judgment implied by the speaker. Thought experiments presume a certain common moral horizon, within which the analogy between the story and reality can be unravelled. This moral horizon secures the effect intended by the speaker in the conscience of the listener.

Thus, the success of a thought experiment is always dependent on certain pragmatic preconditions, which must be met for the speaker's intention to be fulfilled.

Chapter II

# HOW THOUGHT EXPERIMENTS MOVE US: THE SAMARITAN AND HIS NEIGHBOURS

By "pragmatics of thought experiments" I mean an examination of the power of thought experiments to activate the moral capacities of the audience. Thought experiments are known to create a unique, high-pressure situation in which a clear moral decision must be reached. Once somebody has heard (and understood) the story at the heart of the thought experiment, he cannot rid himself of its influence. Since a well-functioning thought experiment can serve as a tool for learning something new, its mechanism must be related to the faculty of cognition. The assertion that once one has come to know something, it is no longer possible not to know it is of relevance here as well.

For example, if I become acquainted with another person, then he has become one of my acquaintances, even if I find that I disapprove of our acquaintance. As long as I remember him, I cannot relegate him to the category of strangers again. Certainly, my mind might efface this person with the passing of time, and I may even forget the fact that we ever met, but I cannot actively pursue forgetting. It is impossible to forget intentionally, since the very endeavour will only serve to

ingrain the memory we wish to forget. A well-constructed thought experiment has the same power: we cannot escape its impact.

A number of conditions must be met for a thought experiment to be the proper and effective.

First, the story told in the framework of a thought experiment has to make the audience face a dilemma situation in which they are forced to make their choice between two options of (nearly identical) value. If the audience does not face a dilemma situation, the thought experiment will fail to reach its goal, which is to change the moral thinking of the audience.[1]

As a second condition, the story must "attack" certain presuppositions of the audience and compel them to question hitherto unexamined beliefs. The dilemma situation manifests itself in the dissonance between the story and the actual horizon[2] of the audience, inducing them to reconsider earlier, mostly unreasoned opinions.

In order to fulfil this requirement, however, a certain analogy must exist between the story told in the framework of the given thought experiment and the horizon of the audience. There must be a certain overall correspondence to make the difference between the previously held beliefs and the stance promoted by the story apparent.

Yet thought experiments do not only formulate expectations with regard to the story being told, but also with regard to the audience. First, they assume that the audience is able to place themselves in the story in an empathetic manner, and to understand the situation of all the relevant characters. Second, the audience is expected to make a decision intuitively, without deeper reflection, even if this decision goes against the moral position they held earlier.[3] Making a personal rational judgement and accepting the message of the thought

experiment are part of the afterlife of the thought experiment and cannot be considered its basic constituents.

In what follows I will use The Parable of the Good Samaritan to demonstrate how previously introduced conditions affect the functioning of thought experiments and their power to exert their influence on the audience.[4] I have chosen this biblical text because of the historical and cultural distance dividing us from the original setting in which the parable was told. This distance not only makes the examination of the text as such possible, but also enables us to raise questions about how the text impacted listeners in the original setting. By the end of our analysis, it will become clear to what extent thought experiments depend in their functioning on the conditions mentioned above, especially on the simultaneously harmonious and confrontational nature of presuppositions present in the text and the reader.

## THE PARABLE OF THE GOOD SAMARITAN AS A THOUGHT EXPERIMENT

After we have seen how imaginary scenarios such as the story of Tomoceuszkakatiti and Gyugyu or the hypothetical situations we present in our everyday conversations may function as thought experiments, it is worthwhile to take an example from another department to get a deeper sense of what constitutes a thought experiment. Similarly to everyday speech, religious conversation has a tendency to use extended analogies. On the one hand, those who use religious language mostly speak about things beyond the limits of the immanent world. As a consequence, analogical speech becomes necessary.[5] On the other hand, those who use religious language do so with the aim of motivating others: their speech is intended to help others

draw near to God, to undergo conversion, and to change their way of life, etc., thus they cannot become bogged down at the level of rational argument. Finally, religious language—if it is not just conversational but missionary in nature—is not very different from everyday language, since the target group for the divine message is a general audience using everyday speech. These are good reasons for taking an example from the realm of proclaiming the Gospel, since the qualities discussed above make it suitable for demonstrating the proper functioning of thought experiments outside academic discourse.

In European culture, The Parable of the Good Samaritan (Lk 10:30-37), with its explicit moral message, has always been one of the most important and influential New Testament texts. But does this parable correspond to what has been said so far about thought experiments?

Not at first sight. If we look at a non-academic definition of the genre, the primary aim of speaking in parables seems to be to deliver a message. According to the online version of the Merriam Webster Dictionary a parable is "a short story that teaches a moral or spiritual lesson; especially: one of the stories told by Jesus Christ and recorded in the Bible".[6]

If we look at the history of how the text has been interpreted, we can certainly uphold the definition above because The Parable of the Good Samaritan has often served the purpose of teaching and has been used as an illustration of moral or dogmatic content. This is confirmed by the fact that from the patristic period up to the 19th century the primary way of interpreting the Bible was allegorical.[7] Moreover, in certain cases the New Testament text itself provides allegorical interpretation of some parables. For example, The Parable of the Sower (Mk 4:3-8) is followed by a lengthy explanation (Mk 4:13-20) revealing content hidden beneath the surface which cannot be appropriated by the audience without proper

reflection. The Church Fathers, primarily under the influence of Philo of Alexandria, aimed at finding the supreme and timeless content beyond the text itself. Literal interpretations are also present among the writings of the Church Fathers – it is at least since the time of John Cassian that we can observe a differentiation between a literal, an allegorical, a moral, and an anagogical interpretation – yet allegorical readings remain dominant, due to their ability to mediate dogmatic content.[8]

There are many examples of an allegorical interpretation of The Parable of the Good Samaritan in the works of the Church Fathers. Besides Irenaeus of Lyons, Ambrose and Augustine, authors like Marcion also testify to the dominance of allegorical interpretations of the text.[9] However, it is Origen who provides the brightest example of an allegorical explanation of the parable in his Homilies on the Gospel of Luke. Here he cites a presbyter's reading of the story:

> The man who was going down is Adam, Jerusalem is paradise, Jericho the world, the robbers are the hostile powers, the priest is the law, the Levite represents the prophets, the Samaritan is Christ, the wounds represent disobedience, the beast the Lord's body, the inn should be interpreted as the church, since it accepts all that wish to come in. Furthermore, the two denarii are to be understood as the Father and the Son, the innkeeper as the chairman of the church, who is in charge of its supervision. The Samaritan's promise to return points to the second coming of the Saviour[10]

Although Origen questions certain elements of the typology applied by the presbyter, he still strongly affirms the practice of interpreting the text allegorically in his homily. As it becomes clear from the quotation, allegorical interpretations

have mainly served the purpose of illustrating dogma and unwrapping a deeper understanding of its content. However, as we will see, with the allegorical interpretation, Origen simply takes the edge off the edge of the parable. The story becomes no more than a picturesque presentation of the Church teaching about her own role in salvation history.

The basis for the dominance of allegorical interpretations of biblical texts can be found in the peculiar context of the Patristic Age in which Church Fathers sought opportunities for dialogue with the philosophical and religious movements of their time. This pursuit led almost inevitably to the application of allegorical interpretation which served as an apt way to illuminate the content of their faith to their non-Christian contemporaries. Allegorical readings appealed to the faculty of understanding, which led to a playing down of ethical content in the process of interpretation.

But what happens if we try to understand the text in its original context, within the framework of the preaching of Jesus? Adolf Jülicher was first to point out the insufficiency of an allegorical interpretation of the parables, and to reach back to the context of the preaching of the historical Jesus. He was followed by such biblical scholars as Eta Linnemann, Charles Dodd, and Joachim Jeremias. While Jülicher—in the tradition of Aristotle—approached parables as argumentative speeches, the approach of the latter researchers was more historical in nature. They claimed that we can only "reach the original sense [of a given parable] if we can reconstruct the original context in which it was told, since the meaning of a parable depends absolutely on the situation, on the hour, when it was born and on the hour it was born for".[11] According to Jeremias there is a need "to recover the original meaning of the parables of Jesus" and instead of an allegorization of the text "to place the parables in the setting of the life of Jesus".[12]

This approach enabled biblical scholars to see the text from new perspectives and thus to produce novel interpretations. It showed that earlier interpretations often overlooked the intentions of the original parable. Although in the early Church it was mission (and later teaching) which stood at the centre of interpretation, the original intent of the preaching of Jesus was to proclaim the Kingdom of God. This distinction is crucial, since the purpose of parables was different in the two contexts: while an allegorical interpretation mainly serves the purpose of doctrinal education, the aim of Jesus was to reach not just the minds, but also the hearts of the audience, bringing about conversion by confronting listeners with the story.

If we want to approach The Parable of the Good Samaritan as an ethical thought experiment, we ought to walk the way determined by Jeremias for at least two reasons. First, a historical approach has a natural disposition to look at the relationship between the text – in our case the parable – and the context in which it is told. In other words, it is interested in the contextual embedding of the story and its analogy with the real world. Second, by coming to terms with the world of the audience, we shed light on the possible reaction of the audience and on how this reaction was produced.[13]

If we define thought experiments, slightly altering the definition Gendler provides, as a "process of reasoning carried out within the context of a well-articulated imaginary scenario in order to answer a specific question about a non-imaginary situation",[14] we will see that the Parable of the Good Samaritan, in its original context, suits this definition. Another definition of parables, which shows a greater sensitivity towards historicity, puts even more emphasis on the impact of the text on the audience. Such is the definition proposed by Charles Dodd, according to which "a parable at its simplest is a metaphor or simile drawn from nature or common life,

arresting the hearer by its vividness or strangeness, and leaving the mind in sufficient doubt about its precise application to tease it into active thought".[15] A number of elements appear in this definition, which are of fundamental importance not just in the case of parables, but also of ethical thought experiments. These elements will be highlighted in what follows.

## SHAPING THE MORAL HORIZON

If we look at the question raised by Jesus at the end of the parable, "Which of these three, do you think, proved himself a neighbour to the man who fell into the bandits' hands?" (Lk 10:36), the story itself seems to deliver the obvious answer. Even the Lawyer knows the answer without any further thinking: "The one who showed pity towards him" (Lk 10:37). It is highly plausible that if asked the question today, everyone would respond as the Lawyer did.

However, the answer is not that simple. There is another difficulty, namely what we mean by the term "neighbour". The question presupposes that we have certain knowledge concerning the meaning of "neighbour". There is a need to clarify the meaning of the term only if it becomes vague in a certain situation. The parable and the answer to the challenge it poses make sense only if we know the original question behind it: the question answered for the audience by the parable itself.

Thus, to make the parable suitable to define the concept of "neighbour", or to determine whom we are to treat as a neighbour and what we owe him, we need to have certain preliminary ideas about the meaning of the term and its moral consequences. This is indicated by the text itself, since the Lawyer approaches Jesus to "test him" (Lk 10:25) and is

described as a person who – by asking the question "And who is my neighbour?" – "was anxious to justify himself" (Lk 10:29).

The text does not articulate what the Lawyer means by the term "neighbour". The question was not posed with the intention of defining the term, but with an other aim altogether. It is clear that

> the term connoted fellow-countrymen, including full proselytes, but there was disagreement about the exceptions: the Pharisees were inclined to exclude non-Pharisees; the Essenes required that a man 'should hate all the sons of darkness', a rabbinical saying ruled that heretics, informers, and renegades 'should be pushed (into the ditch) and not pulled out', and a wide-spread popular saying excepted personal enemies ('You have heard that God said: You shall love your fellow-countryman; but you need not love your enemy, Matt. 5.43).[16]

Jeremias concludes that "Jesus was not being asked for a definition of the term 'friend', but for an indication as to where, within the community, the limits of the duty of loving were to be drawn. How far does my responsibility extend? That is the meaning of the question."[17]

At this level the intent of the parable is nothing but to overwrite the presuppositions implied by the question. It does not seek an answer to the theoretical question raised within the parable but wants to tailor the range of ideas of the audience. On the one hand, the parable needs to reckon with the narrow understanding of "neighbour", which draws the line of charity to include only the Jewish people. Moreover, it must point to a true and appropriate understanding of charity, already implicitly known by the Lawyer who raised the question.

It must be emphasized that the purpose of the parable is not to provide a more correct or accurate definition of the term "neighbour", but rather to put the text to work in the minds and hearts of the audience, and to make it impossible for them to evade the changes it effects in their lives.

Paul Ricoeur points out that parables cannot be forced onto the Procrustean bed of the category of "teaching". It is the impact of parables and their ability to broaden the horizon of human life that is much more important. According to Ricoeur to

> listen to the Parables of Jesus, it seems to me, is to let one's imagination be opened to the new possibilities disclosed by the extravagance of these short dramas. If we look at the parables as a world addressed first to our imagination rather than to our will, we shall not be tempted to reduce them to mere didactic devices, to moralizing allegories. We will let their poetic power display itself within us.[18]

But how does the Parable of the Good Samaritan open a new horizon for its audience? How does it manage to shake the audience's moral certainties and awaken them to the untenability of their faith?

## THE (QUASI-)DILEMMA-SITUATION

One of the concepts with the help of which the impact of the parable can be induced is the dilemma. There are several dilemmas to consider if we take stock only of this single parable told by Jesus. To do so, however, we must approach the text with the appropriate presuppositions.

One of the first dilemmas we discover is the one faced by the priest and the Levite when confronted by the need to decide between ritual purity and mercy. The proper functioning of this dilemma depends on whether the contemporary audience was able to recognize the tension between the two options. If they had simply approved of the priest and the Levite passing without providing assistance, or if they had not recognized that the story dealt with questions of ritual purity central to their Jewish faith, the entire purpose of the parable would have been negated. This is a dilemma situated within the story itself.

In the case of the Samaritan, however, the dilemma points well beyond the story itself. The duty to assist his fellow man does not produce a dilemma situation for the character of the Samaritan. There is no indication in the text that the Samaritan had any difficulty opting to help his fellow man. The text only says that he "was moved with compassion when he saw him" (Lk 10:33b). There is no dilemma for the Samaritan, only for the audience. Whether the dilemma can be solved depends on the ability of the audience to identify with the figure of the Samaritan.

The dilemma-situation arises from the context of the story. Jesus recounted this parable not just in the context of the usual tension between the Samaritans and the Jews, but at a time when their conflict was at its height. As Josephus Flavius reported, a serious incident occurred in 8 A.D. when Samaritans violated the Temple by littering the building with corpses:

> As the Jews were celebrating the feast of unleavened bread, which we call the Passover, it was customary for the priests to open the temple-gates just after midnight. When, therefore, those gates were first opened, some of the Samaritans came privately into Jerusalem, and

> threw about dead men's bodies, in the cloisters; on which account the Jews afterward excluded them out of the temple, which they had not used to do at such festivals.[19]

If we take this antagonism between the Jews and the Samaritans seriously, we may sense the psychological and moral difficulty of the situation in which the audience finds itself upon hearing the story. They knew they must identify with one of the protagonists, preferably with the one who showed mercy to his fellow man. Those with whom they would normally identify, namely the figures belonging to the Jewish people, did not fulfil the commandment to love their neighbour since they simply walked away from their wounded fellow. The only option remaining was to identify with the character of the Samaritan who, for the reasons named earlier, was considered an "alien" and "worthy of hatred" in the eyes of the audience.

The difficulty of identifying with the Samaritan is further revealed by the Lawyer's answer to the question posed by Jesus:

> "Which of these three, do you think, proved himself a neighbour to the man who fell into the bandits' hands?"
> He replied, 'The one who showed pity towards him.'"
> (Lk 10:36-37)

The Lawyer is not even ready to say "it was the Samaritan" who proved himself a neighbour in that situation; "he avoids using the hateful term Samaritan".[20]

## ANALOGY

The dilemma situation created by Jesus when making the Samaritan the protagonist of the story does not manifest itself within the story, but rather in the relationship between the world of the story and that of Jesus' contemporaries. Here lies another precondition for the proper functioning of the parable: there must be an analogy between the world of the audience and of the story. For a dilemma situation to arise, there must be a well-identifiable difference between the two texts. This difference must not be an absolute or a very great one, since that would make all comparisons very difficult or even impossible. Some measure of difference is needed, however, to serve as a background against which similarities can become explicit and clear.

But what constitutes the analogy between the parable and the Lebenswelt[21] of the audience? According to Holyoak "Two situations are analogous if they share a common pattern of relationships among their constituent elements, even though the elements themselves differ across the two situations. Identifying such a common pattern requires comparison of the situations."[22] But which are the relevant constituent elements in story told by Jesus and the life-world of the audience?

Holyoak makes another distinction as he differentiates between the better known text, which he names source, and the lesser known text which he labels target. By means of analogy the source illuminates the target so that we gain new information about the latter. The analogy between the story and the Lebenswelt is necessary to create a dilemma situation, and to illuminate the Lebenswelt of the audience.

The framework of an analogy is determined by who is on the receiving end. Prior knowledge, beliefs and prejudices present in the audience determine which elements are identified as

similar in the story and in their own Lebenswelt. There is a reason for the dominance of simple and profane elements in the parables of Jesus. The reason is "precisely people like us: Palestinian landlords traveling and renting their fields, stewards and workers, sowers and fishers, fathers and sons; in a word, ordinary people doing ordinary things: selling and buying, letting down a net into the sea, and so on."[23] According to Ricoeur the "paradox is that the extraordinary is like the ordinary."[24] Once the audience is familiar with the ordinary, they can recognize the extraordinary elements of the parable with the help of analogical thinking.

Although a great number of similarities may be identified between the story and the reality of the audience, there are two elements concerning analogy which prove to be of crucial importance. The first is the virtue of charity and the duties stemming from it. The second relates to the difference between the story and reality: the concept of the neighbour and the non-neighbour, or alien. The Samaritan acting in accordance with the commandment of charity appears as an alien element in the thinking of the audience and thus produces a fracture in the analogy.

The story told by Jesus gains its power from this fracture. A parable is only successful if a new, full analogy is established between the story and the Lebenswelt of the audience. The imperative of Jesus to "Go, and do the same yourself" (Lk 10:37b) may be interpreted as an exhortation to the Lawyer to amend his life according to the parable and to live a life of charity unbounded.

There are three things the Lawyer might do. First, he might reject the analogy between the story and his Lebenswelt. Second, he might live his own life in accordance with the story. A third option—in which he accepts the analogy between the story and his personal life but sticks to his earlier

presuppositions in thought and deed—is hardly possible. This third option would result in a constant state of cognitive dissonance.[25] Concerning the language of the parables Ricoeur writes: "[T]he power of that language is that it abides to the end within the tension created by the images".[26] This tension, in the end, seeks to be absolved in the life of the audience.

## INTUITION

The incomplete analogy and the dilemma situation open themselves step by step to the audience. It is quite unlikely that Jesus' listeners became aware of their situation right away simply by listening to the parable. Their primary reaction must have been something different.

Upon first hearing the story, they must have felt that it was the Samaritan who did the right thing. It is hard to imagine that any of them would have considered the behaviour of the priest or the Levite exemplary, though they acted in accordance with ritual regulations. They all must have felt—even prior to rational consideration—that it was the Samaritan who acted acceptably and whose actions were just.

But what part of the human psyche does the parable bring into motion? How does the listener immediately light upon the correct answer to the question raised by the story? This is not achieved by rational consideration, since consideration needs more time and requires the use of higher faculties of the mind. It is human intuition, rather, which is activated first to provide a preliminary judgement concerning the moral question hidden in the story.

Intuition—which we conceive of in our everyday lives as a power of judgement arising from the depths of our being and preceding all rational considerations—in effect confronts

all of us with our instinctive selves. There is something uncontrollable about every intuitive judgement. It cannot be pursued, and in the short term at least, we have no influence over its quality. We may only say after the fact "this was the way I felt back then", this is what came from my heart, independent of what I consider good or bad after rational consideration.[27]

When the audience hears the parable told by Jesus, they are also put in a situation in which they have to face their real selves with the help of judgements made before any argumentation, discussion or deduction. In effect the audience must face who they could be in reality: neighbours to one another whose love reaches far beyond the barriers they once deemed insurmountable. The dilemma is thus created by the juxtaposition of intuitive judgement and previous moral assumptions. This "collision" forces the audience to reconsider the latter in the light of the intuitive judgement produced by the story.[28]

The proper functioning of the intuition requires something else beyond the scope of the intentional. This is empathy; the audience must be prepared to empathize with the situation of the major protagonists. In this case they have to be able to sense the tension inside the priest and the Levite, as they are torn between the commandments of ritual purity and charity. They need to relive the spontaneity of the Samaritan as he bends down to touch the prostrate man. Moreover, they also need to identify with the victim, who is in need of aid and is not concerned about who will come to his assistance: whether it be someone from his own people, a total stranger or someone he would earlier have considered his enemy. The activation of the empathy of the audience is necessary for the birth of intuitive judgement.

## THE MESSAGE

In the light of the previous considerations teaching through parables appears to be a very risky enterprise. There are too many conditions to be fulfilled in order for the proper message to reach the addressees. Beyond the cooperation of the speaker and the audience, a number of further requirements must be met, which all seem to be beyond the speaker's control. The speaker must construct the story in a way that the audience may not only understand it, but also be gripped by the hidden analogy and thrown off balance by the dilemma in which they find themselves.

However, one might formulate the message of the parable as a simple statement, as does Jeremias:

> In this parable Jesus tells his questioner that while the 'friend' is certainly, in the first place, his fellow-countryman, yet the meaning of the term is not limited to that. The example of the despised half-breed was intended to teach him that no human being was beyond the range of his charity. The law of love called him to be ready at any time to give his life for another's need.[29]

But this statement is not acceptable in itself; it needs to be well-founded. One may ask why this is the message of the parable and not another.

The message of the parable, if it was a proper one, reaches the addressee prior to the formulation of the message. The parable does not simply make a statement, but issues a call, which is implicitly present in the parable: the way you measure charity does not correspond to your intuitions about charity. Step beyond the limits of your earlier assumptions and dare to

think about charity, and to act accordingly, as dictated by your inner judgement.

Thus, the success of the parable cannot be measured in the spoken word, but rather in its impact on the audience: "The scribe is thinking of himself, when he asks: What is the limit of my responsibility? Jesus says to him: Think of the sufferer, put yourself in his place, consider, Who needs help from me? Then you will see that love's demand knows no limit."[30] The question is whether the Lawyer is able and willing to make this change; will he alter his point of view after recognizing the true nature of charity? This power to change the audience's perspective by activating or even provoking them to intuitive judgement is the key to understanding why Jesus spoke in parables. His purpose was not to teach, educate, or inform people, but rather to inspire conversion, to bring about a change of hearts, and to establish the Kingdom of God.

## Chapter III

# WHAT MAKES A THOUGHT EXPERIMENT?

At this point, the reader might ask whether the fictitious story of Tomoceuszkakatiti and Gyugyu, or the Parable of the Good Samaritan has anything to do with thought experiments applied under academic conditions, since the genre of the former is belles lettres, while the latter falls under the category of religious literature. In order to clarify this question, we first have to refine the definition of thought experiments. However, it is not as simple as it appears at first glance. As James Robert Brown formulates: "Thought experiments are performed in the laboratory of the mind. Beyond that bit of metaphor it's hard to say just what they are".[1]

## THE ORIGINS OF THE TERM

What are thought experiments? What is the nature of the entities we might describe in this fashion? It is hard to answer this question, since we may apply the term thought experiment to diverse phenomena. There are authors who consider even fiction as thought experiment,[2] while others restrict its domain simply to the field of science, emphasizing that experiments can

only be warranted there. Others, like Popa, are more permissive, expanding the territory from natural sciences to the whole field of academic discourse: "Thought experiments are instances of academic communicative interaction in which imaginary scenarios (i.e. stories about fictional objects and events) are employed for the purpose of testing academic claims".[3]

If we accept the latter statement, which defines academic communicative interaction as the field of thought experiments - and thus excludes, for example, John Lennon's popular song Imagine from being treated as a thought experiment - we may have to fit a great number of phenomena under one umbrella term.

A vast number of new thought experiments appeared within philosophy in the last couple of decades, and they are difficult to categorize: Putnam's Brains in a Vat, Nagel's Bat, Searle's Chinese Room, Foot's Trolley Problem, and the Experience Machine formulated by Nozick. The situation is complicated further by the fact that the natural sciences and the humanities are equally fond of utilizing both the term and the method. This multiplies the number of entities falling under the term. If we venture further and expand our investigation by another dimension, we might face an unmanageable number of thought experiments. Thought experiments have accompanied the history of Western thought since the pre-Socratics,[4] through the Middle Ages,[5] up to our times. Thought experiments can be identified in Platonic dialogues, as well as in the works of Hobbes, Locke, Descartes, or Leibniz, who were keen on using this method, despite formulating their own concerns about its application. But we can also mention the names of Galileo, Newton, Schrödinger, Einstein, or Popper from the field of physics, or Alan Turing from computer sciences. The number of thought experiments is enlarged by the often spontaneous "Imagine that..." stories from lectures, discussions, and other

events of popular science in the academic context, which emerge under less formal conditions but constitute a major source for academic thinking.

We face a further difficulty if we consider the origins of the term "thought experiment". The Danish term "Tanke-experiment" first appeared in an 1811 article written by Hans Christian Ørsted (1777-1851), who is mostly known as a physicist and chemist, but who also happened to play an important role in the history of philosophy.[6] Although this first appearance of the term did not make a major impact on the history of philosophy, it shows clearly that Ørsted established the models for the method of thought experiments in the field of natural sciences and mathematics, more specifically in the Naturlehre of Immanuel Kant.[7] This shows that the term was first formulated within the theory of science and did not change as later authors dealt with the question of thought experiments in a much broader and deeper sense.[8] Ernst Mach (1838-1916), Pierre Duhem (1861-1916), and later Carl Gustav Hempel (1905-1997) and Karl Popper (1902-1994) were interested in the scientific value of thought experiments, especially their methodological role. It was physics that kept the question of thought experiments on the agenda for a long period of time.[9] It took another couple of decades to apply thought experiments in the field of practical philosophy with similar intensity.

## EXAMPLES FROM THE FIELD OF NATURAL SCIENCES

Beyond the extraordinary situation in the history of science, which characterized the turn of the 19th and 20th centuries, it is not mere coincidence that physics made thought experiments a subject of academic discussion.

Concerning the exceptional situation in the history of science, Thomas S. Kuhn viewed thought experiments as one of the necessary components of scientific revolutions. He claims that the "thought experiment is one of the essential analytic tools which are deployed during crisis and which then help to promote basic conceptual reform".[10] Kuhn refers here to the inconsistency between the conceptual toolkit and the examined phenomena, which can only be resolved by a paradigm shift. Thought experiments may play a key role in this shift: "thought experiments perform this function by showing that there is no consistent way, in actual practice, of using accepted existing concepts. That is, the thought experiment reveals that it is not possible to apply the conceptualizations we have of phenomena and that this practical impossibility translates into a logical requirement for conceptual reform".[11] The theory of relativity and quantum mechanics have pointed out that certain dimensions of the world simply cannot be described in the language of traditional physics—there is a need for a new conceptual system to approach them.

Simultaneously, there is another reason why physics in this period preferred the tool of thought experiments. The development of physics had reached a point where an empirical examination of a vast number of phenomena had become impossible: there were simply no tools to examine, for example, gravitational waves[12] by empirical means.[13]

A basic question arises inevitably at this point: "The primary philosophical challenge of thought experiments is simple: How can we learn about reality (if we can at all), just by thinking?"[14] How is it possible to gain new knowledge on the cognitive level concerning reality?[15] Kuhn asked the very same question concerning thought experiments: "since they rely exclusively on familiar data, how can they lead to new knowledge of nature?"[16]

Luckily, there are examples from the field of physical thought experiments, which can be tested not only on the level of thinking, but also empirically. (For now, let us put aside the question of what we mean by testing, and also how far the testing of empirical data is dependent on the subject carrying out the testing.) The best known thought experiment in the history of science is probably the one formulated by Galileo Galilei (1564-1642) who famously put empirical experiments at the centre of scientific inquiry.

In his last work, *Dialogues Concerning Two New Sciences* (1638), Galileo refutes the doctrine of Aristotle on the falling of bodies not via empirical observations, but with the help of thought experiments.[17] In the book the master, Salviati – who embodies Galileo himself –, and his two students, Simplicio and Sagredo, discuss the science of matter and movement. The thought experiment is carried out within the framework of this conversation. First, the Aristotelian theory, which was generally accepted through the Middle Ages out of respect for the Philosopher, is outlined. According to the Aristotelian theory, "heavier" bodies fall faster than "lighter" ones: "Aristotle declares that bodies of different weights, in the same medium, travel (in so far as their motion depends upon gravity) with speeds which are proportional to their weights".[18]

Salviati calls attention here to those other "influences which are greatly dependent upon the medium which modifies the single effect of gravity alone".[19] He mentions as an example that gold behaves differently "when beaten out into a very thin leaf": it stops falling and "goes floating through the air".[20] Returning to the original thesis, he asks his student to prove that "the same ratio of speeds is preserved in the case of all heavy bodies, and that a stone of twenty pounds moves ten times as rapidly as one of two".[21] After the student comes up with a solution, which cannot be carried out in praxis, i.e. "Perhaps

the result would be different if the fall took place not from a few cubits but from some thousands of cubits," Salviati rejects the attempt and comes forward with his own experiment.[22] The denial of the unverifiable proposal of the student shows that Galileo considered thought experiments and practical empirical experiments as complementary methods.

The master, Salviati, defines the thesis he wants to prove as follows: "but I claim that (...) if they fall from a height of fifty or a hundred cubits, they will reach the earth at the same moment".[23] After proposing this thesis, he comes up with his own method, asserting that even without empirical experimenting "it is possible to prove clearly, by means of a short and conclusive argument" the thesis in question.[24] But how does Salviati carry out the proof?

> If then we take two bodies whose natural speeds are different, it is clear that on uniting the two, the more rapid one will be partly retarded by the slower, and the slower will be somewhat hastened by the swifter. (...) But if this is true, and if a large stone moves with a speed of, say, eight while a smaller moves with a speed of four, then when they are united, the system will move with a speed less than eight; but the two stones when tied together make a stone larger than that which before moved with a speed of eight. Hence the heavier body moves with less speed than the lighter; an effect which is contrary to your supposition. Thus you see how, from your assumption that the heavier body moves more rapidly than the lighter one, I infer that the heavier body moves more slowly.[25]

Simplicio, the defender of Aristotelian physics, inquires further: "But what if we should place the larger stone upon the smaller?"[26]

Salviatis answers: "Its weight would be increased if the larger stone moved more rapidly; but we have already concluded that when the small stone moves more slowly it retards to some extent the speed of the larger, so that the combination of the two, which is a heavier body than the larger of the two stones, would move less rapidly, a conclusion which is contrary to your hypothesis".[27] And concludes: "We infer therefore that large and small bodies move with the same speed provided they are of the same specific gravity".[28]

It is important to note that it was only several centuries after the formulation of the thought experiment that its validity could be tested empirically. One of the most famous scenes of this testing process was conducted by David R. Scott, astronaut of the Apollo 15 space mission, when he dropped an iron hammer and the feather of a falcon to the ground: the two reached the surface of the Moon simultaneously.[29] For a long time it was held that Galileo himself carried out similar empirical tests; for example, he is said to have climbed to the top of the Leaning Tower of Pisa to drop iron balls of different weight from that height. However, today we know that these empirical experiments probably did not take place. As Michael Serge notes: "In theory this story should have very little importance either for science or for its history. The experiment certainly had no impact on Galileo's thought; if it occurred it was only a public performance and Galileo would not have climbed to the top of the tower without knowing the result beforehand".[30] The emphasis was rather on experimenting in thought.

If the famous experiments at the Leaning Tower of Pisa did not occur, the source of motivation for Galileo to reconsider the Aristotelian model remains in question. Where did he get

his motivation to question the dominant concept of the falling bodies, if not from an empirical observation? Did he depend on conceptual changes in other fields of science or just plain intuition? This question remains relevant to the history of science. One answer is, however, clear from the text; Galileo questioned physical theories simply considering what was possible and justified.

But how was Galileo able to gain new knowledge about the physical world on the level of thought? By simply using a classical rhetorical method: by pointing out the contradiction in the theses of his opponents. He draws up two different imaginary cases (1), which he then applies to the thesis that he wants to deny (2). Finally, he shows that the simultaneous validity of both theses leads to a contradiction (3).

If we put his line of thought into a simple formula, we get the following structure.[31] If Aristotle was right, the "heavier" objects would fall faster to the ground than the "lighter" ones (H>L). If we stick with the objects mentioned in the example above, the hammer—since it is heavier than the feather—will reach the ground faster. But what happens if the two objects are tied together by an immaterial string? The first option is that the feather will slow down the hammer, since its "natural speed" is only a fraction of the speed of the hammer (L+H<H). But the sum of the weight of the two objects is more than that of the hammer and feather separately. Thus, they must fall faster to the ground when bound together than they do individually (L+H>L; L+H>H). Since both statements, namely that they fall faster (L+H>H) and slower (L+H<H) cannot be true at the same time, the Aristotelian thesis cannot be said to be true. There is only one way out of the aporia: to consider the possibility that they fall with the same speed.

If this thought experiment is correct, we have managed to learn something new about reality simply by using our

conceptual thinking. Despite all the criticism against Galileo's thought experiment,[32] the most important consequence for us is that he succeeded in making his partners understand something new about the physical world without relying on any new empirical data; he simply used thought.[33] His statements can be understood by using "common sense" and without extensive training in physics: we can test the heuristic function of the thought experiment without any further tools. It is of secondary importance that we could also test its validity empirically with the requisite technical background.[34]

## EXAMPLES FROM PHILOSOPHY

The question now is not just whether the conclusion of a given thought experiment can be tested through an empirical experiment, but more precisely, whether we can infer something real from the thought experiment. It is also interesting to ponder whether we find out anything essential or new by carrying out an experiment in thought. Since as Brown and Fehige formulate, "Thought experiments are devices of the imagination used to investigate the nature of things".[35]

When somebody starts to work with the philosophical thought experiments of the last decades, he or she may get the impression that this question concerning the usefulness of thought experiments has been central to contemporary philosophy. It may seem that thought experiments have contributed the most to both finding answers and inducing debates in the fields of metaphysics and epistemology. The basic question concerning thought experiments, however, is different in philosophy and in the natural sciences. Philosophers are less interested in what would happen if we did certain things

or constructed certain experimental scenarios or whether the result of a thought experiment can be tested in practice. Moreover, they are not motivated by the need to economize different, mostly material, resources.[36] Rather, they want to find out things that cannot be tested empirically. Examples of such questios are whether computers can think (Searle), whether we can see the world with the same eyes as bats (Nagel), or whether it is sufficient to know the physical description of an object in order to know everything about it (Jackson). The goal of all three thought experiments is to answer non-empirical questions with the help of conceptual thinking.

Like many other thought experiments in philosophy, Searle's "Chinese room" thought experiment tries to find the answer to a question that arises from an everyday context and can seemingly provide a solution on an empirical basis. According to Searle's personal account, listening to a lecture on artificial intelligence occasioned the formulation of his thought experiment in the late 1970s:[37]

> I was invited to lecture at the Yale Artificial Intelligence Lab, and as I knew nothing about Artificial Intelligence, I brought a book by the leaders of the Yale group, in which they purported to explain story understanding. The idea was that they could program a computer that could answer questions about a story even though the answers to the questions were not made explicit in the story. Did they think the story understanding program was sufficient for genuine understanding? It seemed to me obvious that it was in no way sufficient for story understanding, because using the programs that they designed, I could easily imagine myself answering questions about stories in

> Chinese without understanding any Chinese. Their story understanding program manipulated symbols according to rules but it had no understanding. It had a syntax but not a semantics.[38]

This story shows that Searle's thought experiment also has its own context: the golden age of the development of computer sciences when a particular question became more and more pressing, namely whether it was possible to create machines that can think similarly, or even the same way, that humans do. Certainly, the question has a long history, since the relationship of machines and thinking had been raised earlier by a number of thinkers, including Gottfried Wilhelm Leibniz (1646-1716) and Alan Turing (1912-1954).

Leibniz tries to prove the thesis according to which "perception can't be explained by mechanical principles, that is by shapes and motions, and thus that nothing that depends on perception can be explained in that way either".[39]He constructs a thought experiment in order to support this thesis. He asks the reader to imagine a machine "whose structure produced thought, feeling, and perception"[40] and to enlarge it to the size of a mill. Then he asks what we would see if we walked inside this structure: "all we would find there are cogs and levers and so on pushing one another, and never anything to account for a perception".[41] Thus, perception must be something different—according to Leibniz, perception "must be sought in simple substances, not in composite things like machines" – than a sum of determined causal processes.[42]

Alan Turing also sensed that everyday language suggested that thinking was something beyond a bare mechanical process. He proposed reformulating the question "Can machines think?"[43] and made up a game called the "imitation game":

> It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart front the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'.[44]

Turing asks what would happen if one participant - according to his proposal the man - was substituted by a machine. Would we get similar results concerning the questions posed by the interrogator about the identity of the two other participants in conversation? Can the interrogator tell whether the partner on the other side of the wall is a human being or a computer? Turing thus puts the question of rationality to a practical test: if the answers of the partners in conversation fit our everyday discourses, we have no reasons not to view them as intelligent beings.

It is clear from these two examples that Searle did not raise a brand new question in his essay, but investigated a problem with a long history in philosophy. However, his thought experiments might be considered to be much better than any previous attempts, regarding both their intelligibility and their elaboration. The question of understanding ("What is understanding?") is answered by an imaginary scenario that refutes the idea of thinking about understanding as a kind of computer-program. His article "Minds, brains and Programs" was formulated explicitly against a strong understanding of artificial intelligence. According to representatives of strong artificial intelligence, "the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given

the right programs can be literally said to understand and have other cognitive states".[45]

Searle answers this claim with the now-famous Chinese room thought experiment.[46] He asks the reader to imagine being locked up in a room with a large pile of Chinese writings. Since we do not speak Chinese, all these texts are just a great number of "meaningless squiggles" for us. We also get a guidebook written in English, which contains all the necessary rules about how to pair up the Chinese signs in the first pile with those in the second: "The rules are in English, and I can understand these rules as well as any other native speaker of English".[47] However, Chinese writing happens to appear not as a text, but as a line of "formal symbols" for us. We can identify the symbols only by their shape, but not by their meaning.

Imagine that Chinese people outside the room send us questions through the holes on its walls, and we can write, or rather draw, our answers after identifying the symbols, and by using the English language guide. The Chinese partners in the street will think that someone is communicating with them from inside the room. Moreover, they will suppose that their partner in communication can understand their questions and is able to provide a sensible answer to them. But reality is different: in the scenario described, we can only understand the instructions listed in the English manual; we still cannot understand the Chinese language and Chinese script. All that occurs is a manipulation of symbols.

According to Searle, the thought experiment brings to light two important things. First, it is not easy to differentiate between the two "communications", namely the one based on understanding and the other, which is purely mechanical: "I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, but I still understand nothing".[48] The

other is that the simple ability to manipulate symbols does not provide a sufficient explanation for understanding: "the computer and its program do not provide sufficient conditions of understanding since the computer and the program are functioning, and there is no understanding".[49]

Searle's thought experiment received criticism right after its publication and has induced a large number of reply articles ever since, among which we find numerous counter thought experiments.[50] One of the most challenging criticism points out that "Searle's argument depends for its force on intuitions that certain entities do not think".[51] This is another important comment concerning our argumentation about the nature of thought experiments. The thought experiment at hand does not function as a logically constructed system of arguments, but rather as a story that calls attention to one of the most important intuitively conceived supposition of our everyday horizon, namely that understanding cannot be described as the result of a causal program. We identify other people in the process of communication as "other minds" and treat them as such. This is true independent of whether we correctly identified the other as an actual agent of consciousness or were simply mislead by apparent behavior.

This is the situation in the case of the two other thought experiments mentioned above. In his 1974 article, Thomas Nagel seeks to answer the question of whether we can know "What Is It Like to Be a Bat?" The writing formulates criticism against the reductionist approach of physicalism, namely that "mental states are states of the body; mental events are physical events".[52] According to Nagel, despite our ability to describe the physical process of perception in the case of bats, we still do not know, and will never know, what it is like to be a bat. This is an intuitive insight claiming that we can never know exactly what it is like to be someone other than who

we actually are. This is one reason that we are constrained as we try to conceive the consciousness of a bat, since, as Nagel formulates, "if I try to imagine this, I am restricted to the resources of my own mind, and those resources are inadequate to the task".[53] The other reason is the subjective nature and the individual quality of cognition, which remain unrepeatable despite all our efforts. The difference between the subjective experience of consciousness and its physical description becomes clear to the reader not primarily through arguments, but rather through intuitive insight into the difference.

Similarly, the intuition activated by the thought experiment is the point of departure for Frank Jackson as he formulates his arguments concerning physicalism. Jackson asks his readers to imagine a genius scientist, Mary, who from her birth "is confined to a black-and-white room. She is educated through black-and-white books and on lectures relayed on black-and-white television".[54] Mary is raised and taught by scientists, so she knows all the physical qualities of the world: "She knows all the physical facts about us and our environment, in a wide sense of 'physical', which includes everything in completed physics, chemistry and neuropsychology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles".[55] After this, Jackson makes a both rhetorically and argumentatively important statement: "Is physicalism is true, she knows all there is to know".[56] Then he objects to the statement with the help of the intuitive judgement of the reader.

Mary does not know all there is to know, because if she gets out of the black-and white room in which she was earlier confined and comes in contact with something colourful, she will know something new that she was unable to learn simply from the physical description of things. Jackson grounds his argumentation in the intuitive insight that bare physical

description of phenomena is transcended by concrete and immediate experience of these phenomena. This cannot be conceived simply through argumentation; there is also a need for intuitive insight to make the distinction.

It is a characteristic feature of scientific and philosophical thought experiments that their authors attempt to say something about reality simply with the help of thinking. In the case of philosophical thought experiments, however, there is no possible way to check the correctness of the conclusion drawn with the help of empirical experiments. The only way to check the correctness of the thought experiments is to further investigate the intuitive insight produced by the story and the question posed.

## TYPES OF THOUGHT EXPERIMENTS

Thought experiments might be categorized in numerous ways. This is demonstrated by the vast amount of literature providing different catalogues of thought experiments depending on a particular approach.

For example, in his essay On the Use and Misuse of Imaginary Experiments, Especially in Quantum Theory, Karl Popper distinguished between the critical, the heuristic, and the apologetic uses of thought experiments.[57] Popper considers their critical application as the most valuable, and finds its "perfect model" in Galileo's falling bodies thought experiment.[58] Its value manifests itself in the achievement of showing the absurdity of an earlier theory. He also values the heuristic application of thought experiments, for example their application as "the heuristic basis of atomism".[59] However, he issues "a warning against what may be called the apologetic use of imaginary experiments", which are formed in defence of a particular.[60]

Similarly to Popper, Brown and Fehige distinguish between destructive and constructive thought experiments. Destructive thought experiments may "draw out a contradiction in a theory, thereby refuting it" (e.g. Galileo's Falling Bodies); show "that the theory in question is in conflict with other beliefs that we hold" (e.g. Schrödinger's Cat); undermine "a central assumption or premise of the thought experiment itself" (e.g. Thomson's Violinist); or they might function explicitly as a "counter thought experiment".[61] Constructive thought experiments, however, may function as "a kind of illustration" and provide "heuristic aid" in service of a theory and may even provide the "aha effect" that is "so typical of thought experiments".[62]

Daniel Cohnitz also uses three categories to identify different types of thought experiments. He views thought experiments as formulated in connection with a particular theory and examines their contributions to these particular theories. The so-called "clarifying thought experiments" (klärende Gedankenexperimente) do not add anything to the theory, since they "serve only to explain the content of certain natural laws by the use of an example".[63] The same holds for "functional thought experiments" (funktionale Gedankenexperimente), which "take a functional role within the theory".[64] A good example for the latter is "the use of certain counterfactual assumptions in applying test theories to correct statistical data for errors".[65] It is common for both clarifying and functional thought experiments not to provide new information about reality and not to aim at changing the readers' presuppositions.[66] This latter aim is the central to the third type of thought experiments, which Cohnitz labels as "thought experiments to change convictions" (Gedankenexperimente zur Überzeugungsänderung).[67]

Though these distinctions may be of great importance, the focus of our recent investigation is different. For our purposes,

it is more proper to apply categories that help us to distinguish ethical thought experiments from other, non-ethical types. Toward this aim, the taxonomy laid down by Tamar Szabó Gendler draws the line in a simple and clear manner. She claims that it is the fundamental "tripartite structure", which is common for all thought experiments:[68] They (1) describe an imaginary scenario. Then (2) an "argument is offered that attempts to establish the correct evaluation of the scenario," and (3) the "evaluation of the imagined scenario is (...) taken to reveal something about cases beyond the scenario".[69]

Although the core of every thought experiment is constituted by the imaginary scenario, argumentation and evaluation are also essential parts of most thought experiments. In some cases, argumentation and evaluation happen to be non-explicit, as we have already seen in the case of the tyrant and the slave, and also in the Parable of the Good Samaritan. But, as Gendler claims, this "tri-partite structure" can be discovered in all types of thought experiments.

Gendler defines the different types according to the questions they raise concerning the imaginary scenario. In contrast to the categorizations mentioned above, she does not investigate their role within the scientific argumentation, but instead, studies the nature of knowledge at which they aim. According to Gendler, thought experiments may ask three basic questions:

(1) What would happen?

(2) How, given (1), should we describe what would happen?

(3) How, given (2), should we evaluate what would happen?[70]

She labels thought experiments asking the first question as factive, since they seek facts about a future situation. The second type is called conceptual, since they focus on the concepts used in a particular description. Finally, thought experiments that ask questions about the ethical or the aesthetic aspects of a

situation are categorized as valuational. According to Gendler, we can differentiate between scientific, metaphysical or epistemological, and ethical or aesthetic thought experiments using the distinctions between these three basic questions.

The examples mentioned earlier can easily be sorted into these three categories. It is obvious that Galileo's thought experiment concerning falling bodies would belong to the category of factive thought experiments, since it seeks the answer to the question (somewhat simplified) of what would happen if a heavier and a lighter stone were bound together and dropped from a certain height. Searle's Chinese Room and Jackson's Mary's World thought experiments would certainly be classified as conceptual thought experiments, since their major concern is the conceptual description of a certain scenario. The questions they raise are certainly conceptual: May we label what happens inside the Chinese room as understanding? Would we describe as new knowledge what the physically omniscient Mary acquires when she sees the color red for the first time in her life when stepping out of her black and white world?

Gendler claims that there is a "distinct philosophical puzzle" concerning all three types of thought experiments. She formulates these puzzles in forms of questions. Concerning factive thought experiments, she raises the following question: "how is it that thinking about something in a new way can lead us to recognize something new about the physical world?".[71] For factive thought experiments, the key question is of an epistemological nature and concerns the possibility of the identity between the thought and realist corresponding reality. Concerning conceptual and evaluative thought experiments, she holds the following to be the most important questions: "What do we expect to learn about our concepts or values by trying to make sense of this imagined case? Why should

thinking about a case that has not occurred or is not going to occur help us understand how we (should) evaluate actual cases?"[72]

In what follows, we shall approach these puzzling questions from a new angle, focusing on ethical thought experiments. Proceeding from the practice of ethical thought experiments, we will seek an answer to the question of what it is that we can learn by performing an ethical thought experiment.

## THE FEATURES OF ETHICAL THOUGHT EXPERIMENTS

Gendler's question about evaluative thought experiments – "How (...) should we evaluate what would happen?" - sounds different when an actual ethical thought experiment is being performed. In our earlier examples, we heard imperative questions like: "So tell me: how would you like to rise from the dead, as a tyrant or as a slave? Tertia non datur!"[73] and "Which of these three, do you think, proved himself a neighbour to the man who fell into the bandits' hands?" (Lk 10:36)

These imperative questions are all formulated in the second person singular. It is highly probable that thought experiments would bear similar features in the classroom. They would almost certainly end with a question asking what we would do in a particular situation. More precisely, we would consider what the "right thing" to do would be under the given circumstances.

This is a very important claim, since it shows that in the case of ethical thought experiments we do not aim at the acquisition of knowledge independent from ourselves, namely the subject of cognition, judgement, and action. We would like to acquire objective knowledge about the subject. The answer to Gendler's question, namely how we should evaluate ethically

what would happen in a certain situation, greatly depends on who we actually are.

This shows the need for a new definition of thought experiments in ethics, highlighting the process of actions when a thought experiment is being performed:

Ethical thought experiments are (1) imaginary scenarios (2) referring to selected morally relevant aspects of reality and (3) aiming at testing moral beliefs, theses or theories (4) by activating the moral intuitions of the audience.

Although this definition strongly resembles that of Gendler, especially in its fundamental "tripartite structure"[74], there is a major difference: thought experiments might be parts of arguments, and can be used as such. They do not only operate on the rational level of arguments but also on the spontaneous, intuitive level, which is often used as a reference point (if only implicitly) in ethical argumentation. This is the reason why thought experiments can be so perturbing and persuasive at the same time; they appeal to more than our faculty of reason: they challenge our (moral) identity.

Since this text seeks to provide a functional definition, the definitional elements of thought experiments will be shown and explained from a functional perspective, inquiring which need to be fulfilled in the case of a well-functioning thought experiment. The process of communication will play a central role in our discussion, because its proper functioning determines whether a thought experiment will reach its goal. In other words, we will ask: how should an ethical thought experiment be designed so that the imaginary scenario seizes the morally relevant aspects of reality and activates the moral intuitive machinery of the audience?

## IMAGINARY SCENARIOS

One of the most serious criticisms against thought experiments is that they often appeal to what an everyday speaker would say. This is mostly true for thought experiments from the fields of ethics and philosophy. Those from the natural sciences presuppose certain knowledge, which cannot be acquired without a deeper familiarity of the subject. However, most thought experiments in philosophy do not require any knowledge beyond a familiarity with the world around us and how it manifests itself in our everyday language. Since this language with which we are all familiar does carry certain elements of ethical judgement, the appeal to what an everyday speaker would say presupposes an intuitive judgement that every audience member possessed of common sense would make after listening to the story.

But a reference to common sense judgement is often misleading. For example, in the philosophy of language, which often makes use of conceptual thought experiments, a reference to everyday language occurs quite frequently. Linguistic phenomenologists (Austin, Kripke) often appeal "to what we say and, in doing so, they use ordinary language as a self-evident stand from which they proceed to demonstrate the obvious answer to the problem that worries them. Our shared, commonsensical, spontaneous linguistic responses guarantee the truth of their conclusions".[75] This naive appeal to ordinary language is criticised since it "overlooks the difference between alternative natural languages, special languages or dialects. Moreover, it neglects the fact that many factors interfere with our linguistic performance, a lot of which have nothing to do with linguistic competence. (...) So, a biologist's response on whether we would call a bean alive will probably vary from a chef's response, for example, although both would qualify as

average or competent speakers".[76] Thus ordinary language - what that might be, and if it exists as such at all - is just one of the languages which might be referenced.

In the case of ethical thought experiments, referring to ordinary language and to what one might say can be justified. Most of all, ethical thought experiments presuppose that the audience can identify the moral problem described the story. This can certainly depend on previous knowledge of the audience and also on their particular perspective. It is possible to formulate the story in a way that only a small circle of people can understand it. The ordinary character of language does not reside in its accessibility to everyone, independent of culture and education. It is rather that the story presupposes the language spoken by the audience in which the moral problem is formulated. There is no need to create an artificial ethical language to make the audience understand the moral question; it is sufficient to use the language they already speak.[77] Thus ordinary here simply means that which is ordinary for the audience.

The use of ordinary language still does not mean that the imaginary scenarios proposed in thought experiments would not go beyond what is known as "ordinary" in our everyday world. It is possible to formulate counterfactual scenarios without going beyond the limits of ordinary language.

The so called Super-Kittens thought experiment, formulated by Michael Tooley in his much-discussed paper on "Abortion and Infanticide", is a good example.[78] Tooley coined the thought experiment to answer one of the basic questions in the abortion-debate: "What properties must something have in order to be a person, i.e., to have a serious right to life?"[79]

He also provides an explicit formulation of his position concerning the question: "An organism possesses a serious right to life only if it possesses the concept of a self as a

continuing subject of experiences and other mental states, and believes that it is itself such a continuing entity".[80] Tooley formulates a reasonably bizarre imaginary scenario concerning the potentiality argument regarding the moral status of embryos. He asks readers to imagine cats "having all the psychological capabilities characteristic of adult humans"; these animals "would be able to think, to use language, and so on".[81] This can be achieved by giving kittens, bearing none of the mentioned features, a certain injection, which would generate the development of the mentioned facilities. Then he draws a parallel between a fetus and its potential to develop these features, and the now real potential of kittens to acquire these as well. He concludes that "if it is not seriously wrong to destroy an injected kitten which will naturally develop the properties that bestow a right to life, neither can it be seriously wrong to destroy a member of Homo sapiens which lacks such properties, but will naturally come to have them".[82] Without criticizing the analogy drawn between the fetus and the kittens, it is obvious that most people who know the difference between a cat and a human in our everyday thinking, know what thinking and speaking means, and understand that certain medical substances can enhance diverse faculties of animals can imagine and understand the imaginary scenario. One need not go into detail about how this method of enhancement might technically work in the future. What matters is our everyday understanding of what it means to be an intelligent human, how we differ from other animals, and the moral conclusions we draw from that distinction.

By imaginary scenarios, we do not necessarily mean unreal settings. Creators of thought experiments may also recall an actual situation which happened earlier as an imaginary scenario, as long as it poses a certain moral challenge to the audience. The Parable of the Good Samaritan is a scenario that

could have actually happened. Whether it actually happened does not add to or detract from the value of the story as a thought experiment.

## SELECTED ASPECTS OF REALITY

It is not a decisive factor concerning thought experiments whether they outline an imaginary and considerably unlikely scenario or a real situation that actually took place earlier. If the story is coherent and the audience can follow it, it does not really matter whether it is an actual, registered case or just science fiction. It is essential that the imaginary scenario refer to the morally relevant aspects of reality. The audience has to be able to identify these morally relevant elements, so there must be a certain level of identity between the imaginary scenario created by the speaker and the real world of the addressees. As concerning the Super Kittens thought experiment, if we understand the differences between a fetus and a fully developed man, we will also understand the analogy between the imaginary scenario of the super kittens and our conception of the different stages of human development, as well as the moral consequences we draw from them.

The power of imaginary stories is to be found in the existing difference between the story and the (moral) Lebenswelt of the audience. The analogy cannot be a complete one, since this would mean an identification of the story with the Lebenswelt, and the fiasco of the thought experiment as a whole. It is not only that we do not come to know anything new by hearing the story, but also that the moral machinery of the audience remains inactive. At a certain point the analogy needs to differ from the presuppositions of the audience concerning their own world. This difference is what makes the story function.

Concerning earlier examples, if the audience of Jesus, which was of Jewish origin, had thought that ritual purity was more important than the duty of charity, or had not viewed Samaritans as enemies, but as people capable of doing what is good, the story would have lost its substantive message. If the audience failed to identify the morally relevant difference between cats and infants, the story of the super kittens would not work either.

Analogical thinking is not just present in counterfactual thought experiments, but also when the speaker chooses the description of an actual, real situation as an imaginary scenario, or if he asks the partner in conversation to provide a solution to an everyday "what if..." situation. A certain process of selection occurs in both cases. As Gaspara notes: "selectivity is always the case. Even if we are in a real life situation, we concentrate on certain aspects of it. And in any kind of reasoning - especially in scientific studies - one has to neglect some aspects of the phenomena studied in order to draw interesting and valid conclusions".[83] Analogical thinking is a precondition for all thinking: without the act of comparison there is no explanation or new understanding.[84] Moreover, there is no moral judgement without analogy, as is shown clearly by the use of precedent cases in some legal systems.[85]

By altering significant elements, we become able to rethink and restructure the imaginary scenario in thought experiments. This can be carried out in order to resuscitate a particular thought experiment. For example, if the Parable of the Good Samaritan was told in modern Israel, in front of a Jewish audience, the Samaritan might be substituted by a Palestinian man.

But we can also change significant elements in the thought experiment in order to activate the moral intuition of the audience, and thus to rethink our initial conclusion. The

thought experiment about the super kittens might be changed fundamentally if we refer not just to a cat in general, but to our own pet to whom we are emotionally attached. Furthermore, when the story is told in front of Hindu people, we might substitute a cow for the kitten because Hindus consider cows to be holy. The same is the case with other thought experiments: for example, a well known, and historically important figure, such as Desmond Tutu[86], or one of our close relatives, to whom we are emotionally bound, could be substituted for Thomson's Famous Violinist.

This tendency of thought experiments to undergo constant modifications is another characteristic feature. Thought experiments challenge our moral beliefs and theories and stimulate our intuitive machinery and capacity for judgement. As Souder notes, "philosophical thought experiments often (...) evolve through a series of revisions as they pass back and forth between interlocutors".[87] This is even true in cases in which certain specific rules limit the possible scope of modifications. For example, if it were permissible for Tomoceuszkakatiti to know that he was in the wrong, and if he were suffering pangs of conscience, the thought experiment would simply fail. The same goes for Tooley's Super Kittens experiment. If we modify the scenario by asking the reader to imagine a world in which kittens and humans have the same dignity, without any regard for the differences between the two "animals", the purpose of injecting the enhancement serum would simply vanish. The blend of possible modifications and their limitations constitute a playground for testing our moral capacities and beliefs.

## TESTING MORAL BELIEFS

It can certainly happen that a thought experiment confirms our moral beliefs. Although thought experiments tend to end with an open question, one may find a tendency hidden in the well-designed imaginary scenario that shepherds the reader in one direction or the other. Thought experiments might support and even confirm our previously held beliefs, but if that is the case, they fail to mobilize our intuitive and reflective moral machinery. If we have a single, straightforward answer to the thought experiment, it cannot be regarded as functional.

Imaginary scenarios are often used in our everyday conversation to reveal the weak points of our moral beliefs. They may be turned against us by conversation partners who put us in the middle of an imaginary scenario where we might feel uncertain about a particular belief. For example, opponents of the death penalty are often tasked with adopting the perspective of victims and are asked whether they would still hold to their position if a certain crime were perpetrated against them or their loved ones. We may also scrutinize the correctness of our past decisions by placing our interlocutor into the situation we were in and asking his opinion of what the right decision would be in the given circumstances.

Thought experiments about ethical questions in an academic setting also function as tools of testing. Instead of affirming our moral beliefs, functional ethical thought experiments have a tendency to challenge them. For instance, we might believe that we would never act like a tyrant who tortures his slave and makes everyone miserable purely for his own pleasure. After listening to the story of Tomoceuszkakatiti and Gyugyu, however, our certainty is shaken. The same is the case with the often-considered scenario called the Plank of Carneades in which two shipwrecked sailors vie for a plank on the open sea.

Should the first sailor who after reaching the one-man plank kicks his fellow into the open sea be convicted of murder? Someone who holds that people's lives are worthless will probably answer this question with a plain "no". He will not find anything challenging in the thought experiment. But for most of us, things are different. We conceive of this example as a true dilemma.

Thought experiments are often explicitly used to test various theories of ethics. For example, Philippa Foot in her much-cited essay, The Problem of Abortion and the Doctrine of Double Effect, combines several imaginary scenarios to both criticise and rehabilitate the doctrine of double effect.[88] Foot asks us to imagine a group of potholers trying to exit a cave, when one of them, a fat man, gets stuck and obstructs the way out. Foot makes the scene even more dramatic by adding that floodwaters have erupted in the cave and are starting to rise, thereby endangering the potholers stuck inside. The potholers have dynamite, however, which could help them to escape. The only problem is that the fat man will die of the explosion. Foot asks us: "may they use the dynamite or not?" This is a typical thought experiment, putting the audience in a dilemma – it's either him or us! – and activating their intuitive machinery. Foot's use of the thought experiement is also typical. She claims that her imaginary cases served "to show how ridiculous one version of the doctrine of double effect would be".[89] But what does she find ridiculous? It is possible to answer the challenge within the imaginary scenario in the following way: "For suppose that the trapped explorers were to argue that the death of the fat man might be taken as a merely foreseen consequence of the act of blowing him up. ('We didn't want to kill him…only to blow him into small pieces' or even '…only to blast him out of the cave.')"[90] Thus Foot shows us how ridiculous certain interpretations of the principle of

double effect may be, and what impossible conclusions they may yield. Interestingly, she does not provide reasons for considering this interpretation of the principle ridiculous. She just uses it as an illustration, claiming that "those who use the doctrine of the double effect would rightly reject such a suggestion" as well.[91] In this example we can see that even if thought experiments are not explicit arguments, they can still guide our thoughts with the help of their illustrative power.

Another good example of testing a moral theory with the help of a thought experiment is Bernard Williams' Jim and the Indians.[92] The imaginary scenario is described as follows: Jim loses his way on his botanical field trip somewhere in South America and ends up in a small town. He finds twenty Indians awaiting execution by militants. Jim learns that these people will be put to death to deter other citizens from protesting against the government. At this point, Jim becomes the central figure of the story. Pedro, the leader of the militants, offers him a "guest's privilege of killing one of the Indians himself".[93] If he accepts the offer, the other Indians may walk away freely. Since it is impossible to overpower the militants, there are no other options left. "The men against the wall, and the other villagers, understand the situation, and are obviously begging him to accept. What should he do?"[94] Williams claims that the answer to this dilemma seems to be obvious, and Jim should accept the offer: "if the situations are essentially as described and there are no further special factors, it regards them, it seems to me, as obviously the right answers".[95] However, Williams uses this exact imaginary scenario to point out the deficiency of utilitarianism, namely the reflection of how we feel about certain acts, the sense of "what we cannot live with", our moral identity and integrity.[96] His appeal to these dimensions of morality is based on our experience of being trapped in the dilemma described by the imaginary scenario.

His critique of utilitarianism proceeds from our experience of our own revulsion against the seemingly straightforward course of action and highlights our intuitive response against imaginary scenarios being resolved via shortcuts.

## ACTIVATING MORAL INTUITIONS

Intuitions are often ignored in ethical theories. As McBain notes: "On the face of it, (…) [intuitions] would seem to have no real value. But, when we ask whether a particular theory is true, we usually turn to our intuitions. This is nowhere more prevalent than in moral theorizing. When we attempt to show that a particular moral theory is mistaken, we usually present cases that yield counterintuitive results for the theory".[97]

The references to intuitive judgements often serve as cornerstones for ethical theories. Although many would consider these references to be indiscriminate and uncritical, this is not the case. Intuitive judgements concerning how things really are might be unreliable, but they still constitute the first step on the way to a well-founded ethical judgement. The basic moral orientation of the subject becomes apparent when making an intuitive judgement. As Béla Weissmahr claims, "when it comes to the life of the innocent, or the prevention of great injustice, or just the avoidance of being ungrateful, everyone knows that it matters how (s)he acts".[98] It is questionable whether everyone is explicitly conscious of the moral importance of our actions in all such situations, but the statement makes a clear point: this conviction "that it matters how we act" constitutes the major cornerstone of every ethical judgement. This belief is expressed in the spontaneous intuitive judgements about imaginary scenarios in thought experiments.

If we are to test certain ethical beliefs or theories with the help of imaginary cases, we have to face accusations of tautological thinking: "since our goal is to have a moral theory that coincides with our intuitions about cases, if our intuitions fit with the theory, then we have prima facie evidence for the theory. If our intuitions do not fit, then we have prima facie evidence against the theory. Thus, it is our intuitions that carry the evidential burden".[99] This seems to be a real problem: if we take a closer look at the nature of the theory being tested, we find that it is not just an explicit theory, e.g. consequentialism, that is being tested here, but also our implicit theories, which manifest themselves in our intuitive judgements. This is why affirmative thought experiments do not work and are vulnerable to accusations of tautology: they simply confirm our intuitions. Functional thought experiments pose a challenge to our intuitions and question them rather than merely illustrating their correctness.

Others, like Daniel Dennet, claim that thought experiments simply shorten long lines of argument and help us to understand problems and find solutions faster.[100] Dennet's concern is that "the highly imaginative scenarios of some thought experiments can distract from a thorough examination and critical reflection of thought experiments".[101] He mounts the critique that "by appealing to our intuitions, thought experiments can lead us to a quick and uncritical jump to a conclusion that is not really warranted".[102] But taking intuitions seriously does not mean accepting them as solutions to problems in an uncritical manner: rather, thought experiments help us to express our implicit ethical theories in the form of intuitions. This is a prerequisite for their critical analysis. The appeal to our intuitions about an imaginary case is a pivotal step in testing ethical beliefs and theories: it helps us to see the incompleteness of these beliefs and theories. This

use is also recognized by Dennett when he writes that intuition pumps are "not supposed to clothe strict arguments that prove conclusions from premises" but to "entrain a family of imaginative reflections in the reader that ultimately yields not a formal conclusion but a dictate of 'intuition'".[103] As a tool for pumping intuition, he also values its variability as he describes it as "a tool with many settings" that allows you to "turn all the knobs" to see if the same intuitions still get pumped when you consider variations".[104]

The reference to our intuitions also enables ethicists fond of thought experiments to escape another critical claim, namely that the answer we give to the moral challenge of a particular story does not neccesarily overlap with what we actually would say or do. Some even claim - rightly - that we can never be sure of how we would act in a particular future situation.[105] This also means that we cannot know for certain how we would evaluate a certain case in the future. These claims are supported by empirical studies, which show that people giving a particular answer to a particular story do not act the same way in an analogous situation.

According to the well-known Good Samaritan Study of Darley and Bartson, the encounter with a story does not neccesarily entail its influence on our future actions or behavior.[106] In the experiment, students at Princeton Theological Seminary were given the task of preparing a speech based on the Parable of the Good Samaritan, while the other students were to write speeches on different topics. On their way to the place where they were supposed to deliver the speech, they encountered a "victim (...) sitting slumped in a doorway, head down, eyes closed, not moving".[107] The experiment showed no real difference between between those who had the Parable of the Good Samaritan in mind, and those whose speeches dealt with other subjects, which affirmed the hypothesis that a

"person going to speak on the parable of the Good Samaritan is not significantly more likely to stop to help a person by the side of the road than is a person going to talk about possible occupations for seminary graduates".[108] It showed that other factors—mostly "the degree of hurry a person is in" – were the variables that "determine his helping behaviour".[109] Still, the message of the parable and its actuality are unwittingly confirmed by the experiment: "A person not in a hurry may stop and offer help to a person in distress. A person in a hurry is likely to keep going. Ironically, he is likely to keep going even if he is hurrying to speak on the Parable of the Good Samaritan, thus inadvertently confirming the point of the parable."[110]

This does not change the value of the parable as a thought experiment. First, since it was used in an unrevised fashion, the expected effect could not be the same as in its original setting. Second, thought experiments in ethics are not intended to predict what one will actually do, but aim rather at activating and laying bare the moral intuitions of the audience. It is beyond the scope of a thought experiment to convert intuitive judgements into real actions. (When taken a step further, however, they may unveil the dissonance between our thinking and actions.)

Another example comes from the field of psychology. It points to a need for "educating" our intuitions. The experiment applies the so-called Ticking Time Bomb Scenario, one of the dilemmas which cannot be labeled as imaginary on any account, as it is much discussed in the current political and ethical debate about the fight against terrorism. The underlying story goes as follows:

> Officials have recently captured a suspect with information regarding the whereabouts of an explosive device set to detonate in an urban area.

> The suspect is unwilling to cooperate. It is known that the explosive device will detonate within the next six hours, making evacuation of the area impossible. It is also known that torture will be effective on this person, and that it will be effective in time to diffuse the bomb and save thousands of lives. No other means of interrogation can be assured of equal success.[111]

The research indicates that our intuitive answer about the acceptability of torture in this situation can be manipulated in at least two ways. If the audience is characterized by a high level of outgroup prejudice, the use of names from different cultures can alter the intuitions which arise. The research demonstrates clearly that "in general, the higher one's level of outgroup prejudice, the more likely one would be to condone torture".[112] Thus, if elements causing fear in the audience – such as racist stereotype elements – are built into the story, the outcome of the thought experiment may change. This underlines the outlined theory about possible responses of the audience to the Parable of the Good Samaritan in its original setting, but also provides a basis for criticism against using thought experiments as arguments. The influence of outgroup prejudice on the outcome of the ticking time-bomb thought experiment "is particularly troubling for any defense of torture that relies heavily on ticking-bomb methodology—and the vast majority of those who defend torture begin with the ticking-bomb".[113]

But the experiment shows an even more ethically significant factor influencing the intuitive judgement following the story: "the more general and abstract the thought-experiment, the less likely the results of the thought experiment can be generalized".[114] Concerning our example, the results show that "the more abstract the victim in a thought experiment

involving extreme violence, the less universal the results of the thought experiment".[115] This shows again that we do not really know how we would act when looking in the eyes of the captured suspect.

Thought experiments cannot be written off as useless because of the concerns mentioned. They provide the chance for participants to consider the particular situation described in the story, and also to reflect on the particular intuitive judgement it evokes. Despite the aforementioned criticism, thought experiments, even the ticking time-bomb thought experiment, fulfill their particular purpose of making intuitions visible. In this sense

> intuition and thought are not neccessarily opoosed to eachother. (...) Intuition can be a very helpful, even indispenable, guide to us in many situations. Nevertheless, our intuitions in one situation can be improved greatly if we have thought problem through more carefully in previous situations. A good test of any approach to moral decision-making is whether it prepares us to make better intuitive judgements.[116]

This is also the reason why intuition and conscience cannot be considered identical entities. As Daniel Sulmasy puts it, "conscience is not a little voice whispering to each of us infallibly about what we should do".[117] If it were an infallible jugdgement, we could merely rely on our intuitions in every possible situation, without the risk of acting wrongfully. A responsible person would simply be one who acts according to his intuitions. If one suggests that inuitions are infallible or considers them to be final judgements beyond which it is impossible to go, he may have to face the actual pluralism of intuitive judgements. This is what makes Sulmasy also

"skeptical about any form of act intuitionism as a theory of ethics".[118] He points at the fact that

> our intuitions about particular cases will almost certainly differ. If they do, as they seem to in the troubling cases that confront us, such as abortion and physician-assisted suicide, then all we would be able to do would be to recognize that our intuitions differ. According to a theory of moral intuitionism, these differences could neither be explained nor challenged. This leaves open too many possibilities. My intuitions about what is right and what is wrong differ from those of the Janjaweed militia in Darfur. I want to reserve the right to challenge their intuitions.[119]

Still, intuitions are strongly connected to what we call conscience. First, they do rely on the principle found on its highest level, synderesis, namely that good must be done and bad avoided. Second, on the level of conscientia, intuitive judgements must be tested rationally, so that judgements of conscience may use intuitive judgements as a matter of examination. Thus, intuitions are both served and tested by conscience. Without examination, they are just facts, showing instinctive features of the human good as it manifests itself in a particular historical situation.

## INTUITION AND EXISTENTIAL FORCE

The question of whether we can use thought experiments to shore up certain practices or ethical theories must be answered negatively. Even if thought experiments are designed to induce specific intuitions and judgements and thus to "lead to

specific results",[120] they cannot function as arguments. McBain mentions two arguments that support this idea. First, there is always "the possibility of constructing new cases that pump people's intuitions to the other side of a moral debate".[121] Second, there is a "systematic and patterned disagreement as to what the correct intuitive response for a given case is".[122] But thought experiments do not always stop at revealing our particular intuitions provided as a response to a particular story. They may also change the subject on the existential level. Induced intutitions may challenge earlier beliefs, and also moral identity, and this may pressure the audience to reconsider earlier moral constructions.

An example of such an existential impact is presented by Gendler in an example from outside the academic discourse. The well-known biblical story of David and Bathsheba (2Sam 11) lucidly shows how a well-constructed imaginary scenario might change one's moral horizon, even is one happens to be a tyrant.[123] The story is about King David who sets eyes on a woman of great beauty, Bathseba. He invites her to the palace and sleeps with her, although he is aware that she is married to one of his soldiers, Uriah the Hittite. Bathseba conceives a child and is unable to conceal her pregnancy. David decides to send Uriah to the front line of the war so that he may be killed. The text does not inform the reader about what David thought of his actions. The only thing we know is that after Uriah's death, and following the time of mourning, he moved Bathseba to his palace where she delivered a baby boy. This, in broad strokes, outlines the biblical description of what happened. The logic behind King David's actions is not much different than that of Tomoceuszkakatiti: he, too, acts like an omnipotent tyrant.

David gets a visit from the prophet Nathan, whoever, who was sent by God and presents him with the following imaginary scenario:

> In the same town were two men, one rich, the other poor. The rich man had flocks and herds in great abundance; the poor man had nothing but a ewe lamb, only a single little one which he had bought. He fostered it and it grew up with him and his children, eating his bread, drinking from his cup, sleeping in his arms; it was like a daughter to him. When a traveller came to stay, the rich man would not take anything from his own flock or herd to provide for the wayfarer who had come to him. Instead, he stole the poor man's lamb and prepared that for his guest. (2Sam 12:1-4)

David reacts wrathfully to the scenario described, thinking it was an actual report: "'As Yahweh lives,' he said to Nathan 'the man who did this deserves to die. For doing such a thing and for having shown no pity, he shall make fourfold restitution for the lamb.'" (2Sam 12:5-6)

It is only at this point that Nathan reveals the true nature of the story, pointing out the analogy between the imaginary scenario and the deeds of the king:

> I anointed you king of Israel, I saved you from Saul's clutches, I gave you your master's household and your master's wives into your arms, I gave you the House of Israel and the House of Judah; and, if this is still too little, I shall give you other things as well. Why did you show contempt for Yahweh, by doing what displeases him? You put Uriah the Hittite to the

> sword, you took his wife to be your wife, causing his death by the sword of the Ammonites. (2Sam 12:7-9)

After hearing this interpretation of the imaginary scenario, David not only confirms that he has understood the message, but he also makes a confession, which shows that the story did not only work on the cognitive level, but also affected David existentially. He confesses: "I have sinned against Yahweh." (2Sam 12:13) By this confession, he confirms that he now sees his earlier deeds with different eyes.

We may wonder why Nathan did not relay the message of Yahweh literally, in a straightforward style, instead of wrapping it up in a fictional story. What was the communicative purpose of the imaginary scenario?

If we recall the discussion above outlining how thought experiments work, it becomes clear that the story did not end up in the text by accident: it serves a specific purpose. The description of the imaginary scenario activated the ethical faculties of David and enabled him to stay ethically neutral, to react in a pragmatic manner, and to save his power and his reputation as king. But after hearing the ficticious report and reacting to it in an indignant manner, he expressed a moral judgement that he cannot not revoke. He has to hold onto this moral judgement after realizing the analogy between the ficticious report and his own life: "Nathan has enabled David to acknowledge a moral commitment that he holds in principle, but has failed to apply in this particular case."[124] The fictional story did even more by shaping David not just on the cognitive, but also on the existential level. The story is not only effective because "it reshapes his cognitive frame, and brings him to view his own previous actions in its light", but also because it changes the person himself.[125] His confession, (i.e."I have sinned against Yahweh."[126]), cannot simply be

interpreted as a fleeting insight, which had no real impact on his life, but as a realization inducing fundamental changes on the existential level. The Prophet Nathan's parable induced in David moral self-transcendence,[127] leading the king to go beyond the limits both of his cognitive, and of his existential horizon. This example shows that thought experiments may - and well-functioning ethical thought experiments always do - go beyond the cognitive level and affect the existential subject.[128] Affective and moral self-transcendence thus yield existential self-transcendence.

The story of David and Nathan is a good example of a thought experiment as defined above: Nathan's imaginary scenario (1) referring to selected morally relevant aspects of reality (2), namely the deeds of King David, was able to test (3) the king's moral beliefs about his actions (4) by activating his intuitive moral capacities. The king was faced with a dilemma, caught between his earlier judgment of the fictitious scenario and his hesitation to condemn his own evil deeds. The only way for him to escape the dilemma was to allow his own story to be judged by the same intuition.

Gooding's claim that "personal participation is essential" for thought experiments is underlined by the story of David and Nathan, but also by any other ethical thought experiment we would care to name.[129] Personal participation calls forth the singularity and uniqueness of every concrete thought experiment as it is performed. It also becomes clear that an ethical thought experiment does not take place "out there", in the physical world external to us, but inhabits the realm of interaction between the story and the participating audience.

## Chapter IV

# THOUGHT EXPERIMENTS IN PRACTICAL PHILOSOPHY AND BIOETHICS

Thought experiments have been a widely used tool in philosophy and ethics ever since these fields of inquiry came into existence. Their popularity peaked in the second half of the twentieth century in Anglo-Saxon philosophical circles, especially in the United States. At a time when practical philosophy seemed relegated to the sidelines, the thought experiment was one of the instruments that helped ethics to become an important player in the court of philosophers once again. A special historical situation led to the rediscovery of the thought experiment and brought practical philosophy back into the academic game. In his essay on "Singer and the Practical Ethics Movement" Dale Jamieson provides an accurate description of the situation of ethics in 1960s America.[1] He points out the division between the problems in the purview of public interest and the topics discussed in academic discourse: "In the United States there was a clear 'disconnect' between what was going on in the university and what was happening in the streets".[2] Questions central to public discourse in the early second half of the twentieth century were the black liberation movement, feminism, and the Vietnam War. Students at prominent uni-

versities were concerned with these events rather than the abstract questions of academic philosophy. Jamieson recalls a case when students disturbed Searle's lecture at Berkeley: "Searle wanted to lecture on deriving an 'ought' from an 'is', while students wanted to discuss the war in Vietnam".[3] This gap between public and academic questions yielded tensions between the two worlds but also expanded the horizons of contemporary philosophy. Although "Ethics in the classroom was W. D. Ross and P. H. Nowell-Smith", the conversation outside the classroom was mostly dominated by "Martin Luther King, Che Guevara, and the Black Panthers".[4] A new generation of philosophers emerged who were less interested in the discourse on the philosophy of language and keen to explore questions of practical interest. Abbot notes that it is "difficult to think of a major policy or ethical dispute in American politics that has not been subjected to the scrutiny of philosophical analysis – capital punishment, affirmative action, income distribution, civil disobedience, conscientious objection, IQ measurement, vivisection, sexism, pacifism, racism among them".[5]

The first volume of the journal *Philosophy & Public Affairs* (established in 1971) is a case in point. The authors of *Philosophy & Public Affairs* were determined to utilize thought experiments in their work. The first article of the first volume begins with a sentence that clearly shows how imaginary or actual cases were central to the lines of argumentation. Michael Walzer commences his article "World War II: Why Was This War Different?" with the following sentence: "The war against Nazi Germany is an extreme case, but not – one meets young men and women who need to be told – an imaginary case."[6] In his article, Walzer offers an ethical analysis of the reasons for Great Britain entering the war, and attempts to answer the question whether it is right to kill a few in order to save many – an argument which often surfaces when the necessity

of bombing military and civil targets is assessed. This very first article clearly shows the preference of authors for presenting actual cases in their argumentation over abstract reasoning.

Questions of public interest, many of which would now qualify as bioethical issues, were also treated in the journal. Already in the first volume Judith Jarvis Thomson published her hotly debated article "A Defense of Abortion",[7] while Michael Tooley's "Abortion and Infanticide"[8] appeared in the 1972 Autumn edition. Besides abortion,[9] suicide,[10] questions of war,[11] and conscientious objection[12] were also much discussed topics. It was not only the authors of *Philosophy & Public Affairs* who were keen to discuss bioethical topics with the help of thought experiments, however. Philippa Foot published her essay "The Problem of Abortion and the Doctrine of the Double Effect"[13] in 1967, James Rachels his "Active and Passive Euthanasia"[14] and John Harris his "The survival lottery"[15] in 1975, each a landmark essay in the history of bioethics.[16]

## THOUGHT EXPERIMENTS AND BIOETHICS

Why did thought experiments become so attractive in the 1970s and why have they retained their popularity to this day? Why were they utilized with such enthusiasm in discussions of a wide range of bioethical topics? The answers to these questions are inextricably bound together.

The first reason for the popularity of thought experiments is that they are able to connect questions of theory with practical concerns. Since Bioethics encompasses both the theoretical and the practical, there is a special need for a tool which can serve as a channel of communication between them. Bioethicists, coming both from the theoretical context of the academic world and the everyday praxis of the health care

system or environmental policy, found thought experiments to be the ideal method of communication. Second, with the help of tapering scenarios ethicists were able to reach out not only to their fellow academics, but to a much wider audience. Complex philosophical issues could now be brought before the public using a simple story and a well-formulated question. These were important concerns in 1970s America, and are still good reasons for bioethicists to use thought experiments today. James Wilson also points out that "public ethical discourse relies much more on narratives and systems of analogies than on rigorous normative arguments".[17] If so, thought experiments are perfect tools to mediate not just, as Wilson claims, between "normative theory" and "real world cases", but also between the academic world and the public.[18] And, since they "are designed to simplify a philosophical problem along a number of dimensions", they do not only render "the problem more philosophically tractable" but also make it available to a wider audience.[19]

László Nemes also considers the permeation of bioethics with thought experiments an essential part in the evolution of the discipline.[20] He underlines that philosophical bioethics has enhanced the sensibility of philosophers towards practical matters. Thought experiments have also made it possible to "clarify moral intuitions, to reflect on them and to test their validity", since the process of abstraction from a real situation through an "imaginary scenario (…) may affect our thinking in a liberating fashion".[21] Finally, Nemes makes it clear that there was a need for an alternative way to discuss ethical questions, namely one different from "the methods of natural sciences".[22] The sensibility towards practical matters, as well as the necessity of a new approach to these questions on the academic level, were both factors leading to the rediscovery of thought experiments in ethics.

## FAMINE AND DIRTY JEANS

The turn towards practical social issues, as well as the striving to reach the widest audience possible, is aptly demonstrated by Peter Singer's essay "Famine, Affluence and Morality", which appeared in the first volume of *Philosophy & Public Affairs*.[23] Singer combines reality with imaginary scenarios, eliciting both public interest and an appetite for argumentation. The article begins with a description of actual events happening at the time of writing: "As I write this, in November 1971, people are dying in East Bengal from lack of food, shelter, and medical care."[24] He continues with a moral claim packaged in a form that stirs the conscience of Western readers, explaining that "the suffering and death that are occurring there now are not inevitable, not unavoidable in any fatalistic sense of the term".[25] Before analyzing the moral content of the situation, he goes into further detail about the situation of the countries hit by the disasters, and reports on the support provided by Western countries.

He uses the actual situation to present his thesis, namely that "the way people in relatively affluent countries react to a situation like that in Bengal cannot be justified", and argues for a change of "our moral conceptual scheme".[26] Then he uses an imaginary case to call forth the readers' moral instincts, and employs the analogy between the actual and the imaginary case to develop his argumentation. The imaginary case is now known as The Drowning Child scenario: "If I am walking past a shallow pond and see a child drowning in it, I ought to wade in and pull the child out. This will mean getting my clothes muddy, but this is insignificant, while the death of the child would presumably be a very bad thing".[27]

This imaginary scenario is designed to demonstrate and support the thesis that "if it is in our power to prevent something bad

from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it", and creates a framework for criticizing the morals and actual practice of the West when providing aid to other, less affluent regions of the world.[28] By drawing a parallel between the drowning child and the people in the disaster area, Singer manages to create a discourse in which his ethical argumentation can be expounded.

The use of two scenarios, an imaginary and a real one, made it possible for the article to reach two different goals. First, it appealed to a general audience well beyond the scope of the academic circles. Second, it enabled this audience to join the discussion concerning our duty to help less privileged countries. Most importantly, it managed to keep the question of international justice on the agenda.[29]

## CRITICISM OF THOUGHT EXPERIMENTS IN GENERAL

Although the renaissance of thought experiments can be observed not just in Anglo-Saxon and analytical philosophy but in all of philosophy, critical appraisals of their use emerged simultaneously in practical philosophy. Goodin points to a tendentious use of thought experiments in *Philosophy & Public Affairs*, and makes the following critical remark:

> The methodological hallmark of *Philosophy & Public Affairs* is the 'thought experiment.' First we are invited to reflect on a few hypothetical examples - the more preposterous, the better, apparently. Then, with very little further argument or analysis, general moral principles are quickly inferred from our intuitive responses to these 'crazy cases.' (...) Whatever their

> role in settling deeper philosophical issues, bizarre hypotheticals are of little help in resolving real dilemmas of public policy.[30]

Then, Goodin's critique goes deeper. He lists a number of objections to the use of hypothetical cases.

First, Goodin claims that "contrived cases are gratuitous" and add nothing to a given argument. He refers to Onora O'Neill's essay "Lifeboat Earth"[31], where she compares our planet to a lifeboat with first class cabins, while others are deprived not just of enjoying its luxuries, but even of a chance of survival. "But talk of lifeboats adds nothing here. Our objections to some luxuriating while others starve apply equally to the mother ship (first-class passengers feasting while hundreds die below decks) or to the real world directly (Americans overeating while Somalis starve),"[32] comments Goodin. He also highlights that intuitions about actual cases might be much stronger than about certain imaginary scenarios.

Second, they are often "too stripped down", too simplistic or vague to provide "real policy guidance".[33] For example Nozick's Wilt Chamberlain example falls into this category.[34] Nozick uses it to criticize Rawls' idea of the difference principle as a principle of just distribution. Goodin objects to the imaginary scenario because of its vagueness: "We find nothing wrong with Chamberlain's new-found wealth merely because we have not considered all the things that he might do with it. What if he could use that stack of quarters to acquire some of the spectators as his slaves? Or buy their houses out from under them, or all the food off their tables?"[35] These factors are not included in the thought experiment, yet they might very well influence our judgment regarding the regulation of distribution.

The vagueness of imaginary scenarios evokes another objection, namely that "clean cases obscure those interactions

between several moral considerations that are so typical of the complex cases in the real world".[36] Goodin brings the question of punishment as an example. Utilitarians who claim that punishment should serve the purpose of deterrence and are consistent in their thinking should agree with "hanging the murderer's wife" if we could thereby prevent additional murders.[37] Retributivists, by the same token, should agree to the punishments of a criminal even if the punishment fails to deter further criminals. Goodin points out that when hearing the two examples, we might feel that "neither deterrence nor retribution is adequate justification for punishment".[38] The two intuitive answers need to be conjoined, since justified punishment "depends on the interaction of the two", namely "guilt and deterrence".[39] Thus our gut reaction to a specialized imaginary scenario "naturally leads us to mistaken conclusions".[40]

Fourth, Goodin claims that "clean cases fail to tell us how to trade off one moral consideration for another".[41] The question of torture might be assessed one way if someone is being tortured who "has nothing left to offer", and another way altogether in a ticking bomb scenario.[42] Clean cases cannot provide answers to the challenge of balancing different moral values, and also fail to draw the line between the morally justifiable and unjustifiable in actual practical matters.

The fifth objection against thought experiments is that they contain "too much unrealistic or inappropriate detail".[43] This often results in a misguided application of the given imaginary scenario to real affairs. Goodin points here to the inadequacy of the analogy Judith Jarvis Thomson[44] made between the dining club providing special treatment to guests "who had to be excluded at the last sitting", and universities hiring staff from underrepresented groups, such as women or Afro-Americans. He points out that "the real argument revolves around hiring less qualified minority candidates in preference to those white

males who are, on meritocratic grounds, more entitled to the position".[45] The unnoticed difference between the two situations results in an inappropriate application of the imaginary scenario to a real problem.

Finally, "crazy cases contain so much preposterous detail that they stretch our intuitions too far".[46] Here he points to the "desert island case" formulated by W.D. Ross,[47] which asks us whether we must keep our promise if we were one of two old men on the verge of dying on a desert island and our actions would have no effect on the concept of promise as a social institution. Goodin claims that the intuitive answer that it is wrong not to keep the promise even under these conditions arises "just because my intuitions about promising were not shaped on a desert island with a dying companion".[48] He writes that "crazy cases" are problematic since our intuitions, "having been shaped by different circumstances" than those described in the imaginary scenario, would help us just as much as "walking" served us "when on skis".[49]

Goodin's criticism, however, does not stamp out thought experiments from academic discourse. Instead, it sheds clear light on their strengths and weaknesses. Goodin's insights prove especially useful in determining the appropriate application of thought experiments in ethics.

First, even if imaginary cases add nothing to a given argument, they may serve as cornerstones to the edifice of thought. Arguments are not castles in the air, but are founded on the horizon of their inventors. Since they are usually intended for communication, they rest also upon the horizon of the audience. The image of a ship with a few people reveling in luxury while others barely survive does not only put intuition into action and scream for justice, but may reveal the essence of the problem being discussed. The ethical problem of the North-South divide is much harder to catch hold of when discussed

in concrete details than when it is described with the image of a ship with differing levels of comfort.

Second, it might be hard to convert clean-cut imaginary scenarios to actual normative guidance. This is so because the narrative is much simplified and restricted to a few essential aspects. At the same time, we must remember that there is always a concrete, tangible case at the core of a thought experiment. Narratives of actual occurrences can be amended and transformed to highlight substantive elements over what is merely accidental. In the same way imaginary cases might be amended and transformed, providing new accents to the narrative and opening new perspectives for ethical analysis. Just like the Wilt Chamberlain example, every case in ethics is open to further discussion and argumentation.

Third, it is true that the vagueness of imaginary scenarios may lead to simplistic answers which do not live up to the complex nature of the real world and may even obscure the complex nature of ethical phenomena. Although this claim might be true in some cases, thought experiments generally do not hide but reveal the compound nature of ethical cases. By challenging the intuitions of the audience, and by virtue of their ever-changing character, they are actually inimical to simplistic answers. The Parable of the Good Samaritan challenged the conventional answer of contemporary Jews, and transformed their ethical judgment by rearranging their complex system of preferences.

This brings us to the fourth characteristic of imaginary scenarios, namely that they tend to challenge seeming moral absolutes. They do not always provide clear guidance on how to balance different moral values but may point out inconsistencies in our fossilized moral system. The ticking time bomb scenario clearly shows how our moral intuitions change when one or more factors are altered.

Goodin's fifth critical objection that imaginary scenarios often contain "too much unrealistic or inappropriate detail"[50] is not an argument against their use but an appeal for their proper application. The tension between the real situation and the imaginary scenario is open to criticism, and critical insight may both lead to a better understanding of the real situation and inspire further elaboration of the imaginary scenario, finally resulting in a deeper understanding of an ethical problem.

Although the previous points of Goodin's critique are legitimate, the final critique, according to which "crazy cases contain so much preposterous detail that they stretch our intuitions too far"[51], cannot be justified. First, the term "crazy" is too vague to define a certain type of imaginary scenario. Second, it is also questionable whether intuition can be stretched too far. Intuitions may be active or inactive, but there is no limit beyond which they might be labeled illegitimate. Third, and this is the most important of all objections, imaginary scenarios can never be absolutely alien to us or our frame of reference. They are always rooted in our previous experiences and knowledge of the world. If the understanding of the moral nature of a particular situation depended solely on the intuitions shaped by the culture surrounding us, and if these intuitions were disqualified when applied to a situation in another culture, the basis for a mutual understanding would be lost. The real world shaping our intuitions and the world of the imaginary scenario are analogous, and this means that they are simultaneously similar and different. Crazy cases, even if they stretch the boundaries of what we consider reality, may be the perfect tools to fire up our intuitive arsenal.

## TESTING THEORETICAL BIAS: THE HEINZ DILEMMA

It is a common dogma in philosophy that neither intuition nor empirical research can serve as a sufficient foundation for an ethical theory. However, both the natural sciences and the humanities do actually inspire philosophy, including its practical branches. One example of this inspiration is the rise of care ethics which evolved following the debate over Carol Gilligan's 1982 book *In a Different Voice: Psychological Theory and Women's Development*.[52]

Gilligan was a colleague of Lawrence Kohlberg, but eventually became one of his most strident critics. Kohlberg described the ethical development of the individual as a growth process in the direction of justice and autonomy, similarly to Kantian ethics. According to Kohlberg, development reaches its peak at what has been dubbed the post-conventional moral level, where the "social contract" and respect for universal moral law motivate an individual's actions. Kohlberg came to the conclusion that most women were unable to reach this level and to act according to universal moral principles. This was the point in Kohlberg's theory that elicited Gilligan's criticism. She challenged him not only on an empirical basis, but also by questioning the ethical theory behind Kohlberg's findings.

Kohlberg used short imaginary scenarios to identify the principles people use in their ethical decisions. Since most of the responses by women did not reflect the post-conventional level, Kohlberg concluded that most women were unable to attain moral maturity. Gilligan "observed that after reaching the post-conventional level, women are not motivated by some universal and abstract law, but by the care for others. Their ethical decisions are determined primarily by their relationship to others, by the perspective of the relationship between persons. According to Gilligan it is not that women were underdevel-

oped in their moral judgments, but rather that there is another perspective beyond justice and rights, namely that of care."[53] This insight serves as the foundation of the normative theory of "care ethics", which tries to emphasize genuine caring relationships over justice-oriented ethical theories. As Gilligan claims: "Adding a new line of interpretation, based on the imagery of the girl's thought, makes it possible not only to see development where previously development was not discerned but also to consider differences in the understanding of relationships without scaling these differences from better to worse."[54]

Gilligan uses an imaginary story, originally designed by Kohlberg, to illustrate her thesis:

> In Europe, a woman was near death from cancer. One drug might save her, a form of radium that druggist in the same town had recently discovered. The druggist was charging $2000, ten times what the drug cost him to make. The sick woman's husband, Heinz, went to everyone he knew to borrow the money, but he could only get together about half of what it cost. He told the druggist that his wife was dying and asked him to sell it cheaper or let him pay later. But the druggist said, 'No.' The husband got desperate and broke into the man's store to steal the drug for his wife. Should the husband have done that? Why?[55]

The story was presented to two eleven year old children, Jake and Amy. To avoid any gender bias, the two children chosen "resisted easy categories of sex-role stereotyping, since Amy aspired to become a scientist while Jake preferred English to math".[56] However, their answers to the moral question clearly showed a difference in their ethical thinking.

Jake insisted that Heinz should steal the drug. He discovered a conflict in the described situation between right to life and right to possession.[57] He claimed that

> For one thing, a human life is worth more than money, and if the druggist only makes $1,000, he is still going to live, but if Heinz doesn't steal the drug, his wife is going to die. (Why is life worth more than money?) Because the druggist can get a thousand dollars later from rich people with cancer, but Heinz can't get his wife again. (Why not?) Because people are all different and so you couldn't get Heinz's wife again.[58]

Jake solved the ethical problem as if it were a mathematical equation. Since the right to life is more fundamental than the right to possession, Jake may steal the drug his wife needs for recovery: "Considering the moral dilemma to be 'sort of like a math problem with humans,' he sets it up as an equation and proceeds to work out the solution".[59] As in the case of every mathematical solution, Jake considers it to be universal, assuming that any other reasonable person would come to the very same conclusion. However, Amy is one person who responds in a different way. She resists solving the ethical dilemma as if it were a math problem. She responds to the question at the end of the story in a creative manner: "Well, I don't think so. I think there might be other ways besides stealing it, like if he could borrow the money or make a loan or something, but he really shouldn't steal the drug – but his wife shouldn't die either."[60]

She does not let to get herself trapped in the dilemma of the story, but tries to arrive at creative solutions, like talking once more to the chemist or asking relatives for financial help. She thinks in terms of relationships, not abstract laws: "For her hu-

man relationships matter most, and she trusts the powers of communication and mutual understanding. It is not abstract principles, such as legal rights or justice which take precedence in her mind, but the responsibility for each other and our relationships."[61]

Using the example of Amy and Jake Gilligan shows that there are two distinctive moral orientations, one centered on justice and the other focused on care. Although these might be described as masculine and feminine respectively, they are in truth accessible to both genders. The philosophical tradition – due to the simple fact that the core curriculum in philosophy includes almost exclusively male thinkers – tends to ignore the ethics of care and focuses instead on the ethics of justice.[62]

The role of relationships when making ethical judgments is even more obvious if we reconfigure the original scenario. What responses would we get if it was a man who was dying and a complete stranger tried to help him by breaking into the pharmacy?

> In Europe, a man was near death from cancer. One drug might save him, a form of radium that druggist in the same town, a woman named Hilda, had recently discovered. The druggist was charging $2000, ten times what the drug cost her to make. A perfect stranger, a woman named Heidi, chanced to read about the sick man's plight in the local newspaper. She was moved to act. She went to everyone he knew to borrow the money, but she could only get together about half of what it cost. She asked the druggist to sell the drug more cheaply or let her, Heidi, pay for it later. But Hilda, the druggist, said, 'No.' Heidi broke into the woman's store to steal the drug for a man she did not know. Should Heidi have done that? Why?[63]

The new story shows that socially well-defined relationships such as marriage matter when we evaluate the moral dilemma of breaking into the drugstore. Therefore, even justice-centered ethical theories cannot escape the question of caring. (This of course depends on whether one's intuitive judgment suggests that it was less right for the stranger to break in than for the wife or the husband.)

What can we learn about thought experiments from the Kohlberg-Gilligan debate? First of all, we conclude that imaginary scenarios can help us to form ethical theories. When interpreted with the help of an open theory, the responses can direct our attention to possibilities we may have neglected in the past. Gilligan has shown that Kohlberg was wrong to conclude that women's different response to the dilemma could be ascribed to their ethical immaturity. Kohlberg's original thesis needs to be amended with the help of a different moral approach centered on care for others. Gilligan's work may serve as a warning against simply dismissing unexpected responses to a thought experiment. People, and their responses to ethical dilemmas, are far too complex to fit snugly into any particular theoretical box. Second, we learn that imaginary scenarios may also help us to develop an anthropology by revealing what matters to the respondents. In this case the original theory glossed over the fact that care is a central component of the human good and that it could serve as the basis of an ethical theory. The claim for justice is in no way stronger than the simple fact that every human being is in need of care. Thus, even if neither an ethical theory nor a philosophical anthropology can be derived from the responses given to the imaginary scenario, the results may still serve to point out important aspects left out of the original theory. Third, the need to alter original imaginary scenarios is clear. Well-constructed alterations can lead to new insights and shed new light on earlier results.

## THOUGHT EXPERIMENTS, SOCIAL CHANGE, AND BIOETHICS

Since imaginary scenarios are hard to see as solid elements of an argumentative structure, they are often considered superfluous to ethical reasoning. Their power lies in their potential to connect the audience with the moral problem and the ethical argumentation. Ethical concerns about disastrous conditions in the Third World versus lavish Western lifestyles are better understood through the image of a drowning child than by means of argument. Philosophical arguments may result in further philosophical arguments presented in a detached manner at the level of the mind, but picturesque scenarios move the audience in a way that makes it impossible for them to ignore the ethical problem.

The authors of *Philosophy & Public Affairs* used thought experiments since their goal was not only to create better ethical theories but to effect social change. In their writings the question of justice hardly ever appeared as an unchanging and ahistorical universal concept, but rather as a practical principle guiding distribution or warfare. It was understood as something that affects all readers, their society, and the world at large. Their message reached a much wider audience and induced a far greater number of discussions than if they had simply presented their arguments in an abstract form. Imaginary scenarios provided a medium which connected the situation of people in Bangladesh with the Lebenswelt of an American readership. They served as handholds to help people grasp not just the inequalities between poor and rich countries, but also the ethical question of the individual and social action needed to reduce suffering.

Such mediation is also crucial in bioethics. Autonomy, justice, and the sanctity of life are only abstract principles

until one experiences their meaning first-hand or is provided with an example compatible with his own life experience. Questions of birth and death, health and suffering may be phenomena common to all of humanity, but mediation is needed to show individuals their complexity and to open up perspectives beyond one's current state of being.

The thought experiments described in the following chapters play an important role in bioethical discussions. They often serve as starting-points for debates and are part of the classical bioethical canon. Their analysis shows the variety of uses – argumentative, rhetorical, and educational – to which bioethicists have put them. For this reason, any examination of these thought experiments must go beyond mere considerations of their use in theoretical arguments to explore their practical application.

## Chapter V

# THE EXPERIENCE MACHINE

Saint Anselm's *Unum Argumentum*, better known as the ontological argument, is one of the most discussed texts in the history of philosophy.[1] One of the reasons for its popularity is that the argument, which is the odd one out among the numerous arguments for the existence of God, constitutes a puzzle for believers and sceptics alike. Its ongoing presence in philosophical and theological discourse, and in fields as unlikely as physics, is due to its exceptional status as a third category of argument besides the cosmological and the anthropological. The same holds for the Experience Machine Thought Experiment which belongs clearly to the realm of ethics yet also differs from other thought experiments in the category. Most ethical thought experiments begin with an imaginary scenario where a particular life episode calls for an ethical solution. But this is not the case with the Experience Machine. It does not ask whether a certain action was right or wrong, but forces us to consider what really matters when it comes to the question of a good life.

## THE EXPERIMENT

The first version of the Experience Machine Thought Experiment was formulated in Robert Nozick's *Anarchy, State and Utopia* in 1974.[2] It goes as follows:

> Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life's experiences? If you are worried about missing out on desirable experiences, we can suppose that business enterprises have researched thoroughly the lives of many others. You can pick and choose from their large library or smorgasbord of such experiences, selecting your life's experiences for, say, the next two years. After two years have passed, you will have ten minutes or ten hours out of the tank, to select the experiences of your next two years. Of course, while in the tank you won't know that you're there; you'll think it's all actually happening. Others can also plug in to have experiences they want, so there's no need to stay unplugged to serve them. (Ignore problems such as who will service the machines if everyone plugs in.) Would you plug in?[3]

The imaginary scenario does not sound as extreme at the beginning of the twenty-first century as it may have sounded

at a time when computers had only begun to spread in the US and the possibility of a virtual life still seemed both amazing and distant. Our lives today often seem to be mediated, which makes the Experience Machine a highly probable scenario. We can slowly forget about carrying cash in our wallets as we usually purchase items using the virtual money on our credit cards. When we drive our cars, we spend more time watching the screen of the GPS-device than the actual road, and we wistfully read the news on Google glasses or on the screens of future vacuum trains, which provide a lively representation of the world outside. No matter how these examples resemble Nozick's machine, they differ in one crucial aspect: these technologies are aimed to connect us to reality in new ways. The credit card represents the actual 50 Euros we can spend; the properly functioning GPS shows us the way home or to our friend's house along an existing road; Google glasses may provide us information about the cuisine of a real restaurant, and the screens in the future vacuum train will show us the view between Budapest and Vienna that we would actually see if the tunnel was transparent. These media serve to connect us with reality, making them radically different from Nozick's proposal.

## THE CONTEXT

The passage quoted above is not just unusual when compared to other ethical thought experiments, but it also deviates from other parts of the book. There are other thought experiments throughout the work, such as the famous Wilt Chamberlain thought experiment, but those are well-embedded in the text and play a major role in the line of reasoning. By contrast the Experience Machine is a detached passage, loosely connected

to the previous and the subsequent passages. The passage preceding it evaluates the slogan "utilitarianism for animals, Kantianism for people", providing an ironic critique of utilitarianism by discussing the moral status of animals, while the passage following the thought experiment again asks whether "beings from another galaxy [could] stand to us as it is usually thought we do to animals, and if so, would they be justified in treating us as means a la utilitarianism".[4]

Its extraordinary position in the book is certainly one reason for the popularity of the though experiment. As Feldman notes, "the passage is a bit of a mystery - perhaps it functions as a kind of Rorschach test for the readers".[5] It certainly attracts the attention of the readers more than any other part of the book. Readers might feel that the passage does not fit neatly inside the line of reasoning, yet it certainly supports the anthropology behind the ideas of libertarianism and the concept of the minimal state[6] Nozick advocates in his book. Readers tend to gloss over the two remaining thought experiments in the subchapter, namely the "transformation machine" and the "result machine". The former 'machine' "transforms us into whatever sort of person we'd like to be", while the latter "produces in the world any result you would produce and injects your vector input into any joint activity".[7] Nozick finds these imaginary machines upsetting since they live "our lives for us".[8] The idea of living one's own life is actually a central tenet of libertarian thinking, as is shown by the Experience Machine Thought Experiment. Nozick may well have included it to emphasize our desire to live our own lives.

This view seems justified by the sentences framing the thought experiment. Nozick begins the passage with the following statement: "There are also substantial puzzles when we ask what matters other than how people's experiences feel 'from the inside.'"[9] The key expression is repeated in the form

of a question after the outlining of the imaginary scenario: "What else can matter to us, other than how our lives feel from the inside?"[10] The answer to this question arrives at the end of the passage where Nozick highlights the centrality of the pursuit to live our own lives.

## THE EXPERIENCE MACHINE AS AN ARGUMENT

Even if this statement is true and follows obviously from the text itself, both the question (What matters to people more than how they feel from the inside?) and the answer (living their own lives) are still too vague. This vagueness opens the passage to a great number of interpretations. As Feldman notes, it "may seem that the various interpretations tell us more about the readers than about the argument that Nozick actually presented".[11] Interpretations of the thought experiment in fact range from the very obvious to the radically sophisticated. Most comments either interpret the text as a case against utilitarianism or as an argument challenging hedonism. This uncertainty results from the dissonance between the text and its context. Those who see the Experience Machine as an argument against utilitarianism usually point to the passages preceding the thought experiment. These passages focus on the concept of utilitarianism, which has "led some commentators to think that Nozick was presenting an argument against utilitarianism in the experience machine passage".[12] The problem with this interpretation is that "there is no evidence in the passage itself that would support this interpretation".[13] Feldman provides two reasons why this is true. First, Nozick never explicitly mentions in the passage that it was written against the utilitarian doctrine, whereas he is very clear about his objections to utilitarianism in passages not connected to the

Experience Machine in any way.[14] Second, Feldman claims that "the proposed argument would be relevant only to versions of utilitarianism that incorporate a hedonistic axiology".[15] There is one more reason to think that utilitarianism is not the real target of the thought experiment: there is no calculus involved. It seems obvious that the experience machine could be used to maximize pleasure for the individual; however, it says nothing about maximising the net amount of pleasure in a society or in the world. We may go further and ask whether we could maximize pleasure if the machine were to function forever, automatically, without it breaking down and with everyone plugged in, but these factors are simply not part of the original thought experiment, in fact, they are not even hinted at. The lack of a hedonic calculus undergirds the claim that the passage was not intended as a challenge to utilitarianism. Thus, as Feldman concludes, the Experience Machine Thought Experiment should rather be "understood as an argument against ethical hedonism".[16]

The ambiguous use of language casts doubt on this interpretation as well. While the thought experiment is often labelled as "the pleasure machine", the word pleasure actually does not appear in the passage. Experience is automatically interpreted as pleasurable experience, implying that no one would wish unpleasant experiences for himself, and would therefore avoid programming them into the machine. This claim might be true, but it needs to be adduced by arguments. Yet, as Silverstein points out, "many of the most prominent philosophers of value (...) take this thought experiment to be the definitive response to hedonism and, more broadly, to all mental state theories of well being".[17] Furthermore, "in anthologies of moral philosophy, Nozick's experience machine is often the only argument offered in response to classical hedonism."[18] There is nothing in the text against understanding

experiences as pleasurable experiences, and it is logical to do so even if Nozick failed to elaborate his understanding explicitly. His examples of possible experiences, like "writing a great novel, or making a friend, or reading an interesting book"[19] are all pleasurable activities. He also writes about "desirable experiences", but the question still remains whether only pleasurable experiences are desirable. Even if we do not specify the experiences as pleasurable, the main message of the thought experiment cannot be missed.

## THE PRAGMATICS OF THE EXPERIENCE MACHINE

Before turning our attention to the message of the Experience Machine, it is important to assess it as a thought experiment. Does the pragmatic definition established in the previous chapter fit the Experience Machine? According to the definition

*ethical thought experiments are (1) imaginary scenarios (2) referring to selected morally relevant aspects of reality and (3) aimed at testing moral beliefs, theses or theories (4) by activating the moral intuitions of the audience.*

The Experience Machine can be categorized as a sci-fi thought experiment, since it uses images taken from the realm of science such as neuropsychologists stimulating the brain and a tank with electrodes attached to the brain, and the machine described is a fictitious object. Certain conditions must be met before the imaginary scenario can be presented to the audience. (1) First of all, the audience must have some knowledge of who neuropsychologists are and why electrodes are used to stimulate the brain. Although the experience machine does not exist, they need to have a vague understanding of its functioning. (2) Second, they need to understand the difference between reality and the virtual world, and the existential

relevance of this distinction. If someone thinks that saving a child from a pond in reality and on the television screen are of similar importance he will not grasp the point of the Experience Machine Thought Experiment. Since this moral distinction is at the heart of the thought experiment, it is impossible to participate in it without a fundamental understanding of the distinction. (3) Third, they need to be aware of the claim that it is only pleasurable things which are valuable to us, whether or not they personally agree with the statement. Without this common ground the thought experiment will fail to yield new insight. (4) Finally, the audience must react with intuitive disgust (or enthusiasm) to the idea of being plugged into the proposed machine until the end of their lives. If all these conditions are met we can expect the thought experiment to function properly. The power of the Experience Machine comes from its ability to put the audience on the horns of a dilemma, namely choosing between the experience of all they desire and their actual lives. While Nozick does not explicitly claim that one would choose not to plug in, he does rely heavily on the intuition of the audience regarding the distinction between their actual life and the possibility of a life spent plugged into the experience machine.

Thus by referring to the existentially relevant difference between reality and the virtual world, by testing the idea that only pleasurable things are of value to us, and by making the audience react with intuitive disgust or enthusiasm to the idea of being plugged into the proposed machine until the end of their lives, Nozick succeeds "in isolating the fact that we care about more than our experiences."[20]

## NOZICK'S CONCLUSIONS

The Experience Machine Thought Experiment certainly has an existential hold on the audience and fundamentally tailors their horizon. The intuitive insight that there is something more important to us than merely our (pleasurable) experiences is elaborated further as Nozick defines the content of what matters. He asks "What does matter to us in addition to our experiences?"[21] Or, to put it another way, what are the reasons "for not plugging in"?[22] He mentions three. "First, we want to do certain things, and not just have the experience of doing them."[23] This does not simply mean that we want our actions to have a certain effect in reality. This is refuted by the thought experiment about the "result machine, which produces in the world any result you would produce and injects your vector input into any joint activity".[24] It does so once again on an intuitive basis, by pointing at our fundamental thirst to perform our actions as real actions, and not merely to experience them or to enjoy their fruits. Nozick is aware that the reason mentioned is less argumentative and more intuitive. This is underscored by his question: "But *why* do we want to do the activities rather than merely to experience them?". He describes an intuition but cites no argument.

This approach holds for the second reason Nozick mentions: "We want to be a certain way, to be a certain sort of person."[25] He again uses an additional thought experiment, the "transformation machine which transforms us into whatever sort of person we'd like to be", claiming that we would not make use of it either.[26] Intuitive rejection is explained by the disgust induced by the image of a "floating" body "in a tank" described as an "indeterminate blob", as well as the author's use of the word "suicide"[27] to describe this state. Nozick provides a

deeper explanation when he points out that "what we desire is to live (an active verb) ourselves, in contact with reality".[28]

The third reason opens up the nature of this reality. If "plugging into an experience machine limits us to a man-made reality, to a world no deeper or more important than that which people can construct", then the real world, which goes beyond the limits of what man can actually construct, is something more valuable to us. Although the first two insights are clearly intuitive, this statement relies rather on the assumption that reality is always more than what we can construct. This statement is rather derivative than intuitive, which is confirmed by Nozick's reference to "psychoactive drugs", which some "view as avenues to a deeper reality".[29] The experience machine may seem desirable to someone who accepts this argument. To one who believes that reality is something more than what our brains or minds can produce, the experience machine will seem preposterous.

## CRITICAL VOICES

Although it took only a short time for the Experience Machine Thought Experiment to become a standard part of the argumentation against hedonism, criticism was formulated at much the same pace. Some critiques target the role of thought experiments in ethics in general, while others focus on certain points of the passage.

One of the general claims is that we simply don't know how people would react if faced with an agent offering them a lifetime of pleasurable experiences in a machine. Not only do we not know how they would respond, but we are also ignorant of their reasons for responding in a particular way. Perhaps they would say no not because they were disgusted

by the idea of spending the rest of their time in the experience machine but because they mistrusted the agent who made this unlikely offer. At this point the difference between actual events and imaginary scenarios becomes visible. How people make decisions under given circumstances is an empirical question and the subject of psychology and other empirical disciplines.

The Experience Machine Thought Experiment does not ask what one would actually do when faced with the dilemma of entering the super-duper machine, but looks for the intuitive answer the imaginary scenario induces in the audience. As the passage shows, Nozick firmly believes that everyone would answer "no" when faced with the offer and gives his reasons for his belief. Still, there is the possibility that some people may answer with a "yes" and would also be able to justify their choice.[30] All we know by listening to the scenario is our intuitive response to it. This can be extended when we reflect on the reasons for our choice or explain these reasons to others. In the end what we stand to gain from considering the scenario is intuitive insight.

Another objection is that the distaste for the idea of entering the experience machine can be explained with reasons other than those mentioned by Nozick. We have certain experiences of reality which determine our intuitive responses. Silverstein names two of the experiences which might be responsible for the intuitions concerning the experience machine: detachment and the unenforceability of happiness. We all harbor a certain aversion towards detachment, since it is usually accompanied by pain. "We develop a desire to track reality in a world in which detachment from reality is painful. The thought experiment marks a radical departure from this world and the circumstances under which this desire was formed."[31] The second is the experience that "when we pursue

our own happiness, it invariably eludes us".[32] Silverstein claims that "the experience machine scenario is an exception to this teaching; it is one of the rare situations in which if we think of only our own happiness, that is exactly what we will attain."[33] Our negative experiences with detachment and the pursuit of happiness prevent us from accepting the offer: "We are unprepared, however, to respond to the machine in this way. We have been programmed, as it were, to recoil in horror from such a departure from reality, and we have been conditioned to aim for ends other than our own happiness."[34] Silverstein concludes that the "fact that we all intuitively reject the experience machine is merely a sign that our intuitions are functioning properly, that we are prepared to find happiness in the real world, where the failure to track reality inevitably has painful consequences."[35]

Others claim that we are not able to set all our worries aside when it comes to the decision about plugging into the experience machine. We can't get over our suspicion that the scientists who invented and are running a machine may be unreliable, just as we cannot fully trust its proper functioning. But in order to make a decision according to the immanent rules of the thought experiment we have to suppose that both the machine and the scientists are sound and trustworthy. L.W. Sumner describes precisely what concerns might arise when someone imagines himself as a body floating in a tank:

> Once we are floating in the tank, we will have relinquished all control over how things subsequently go for us; we will be in no position to change our minds or demand a refund if the goods are not as promised. We immediately begin to imagine the ways in which things could go horribly wrong. How do we know that the technology is foolproof? What happens

> if there is a power failure? Suppose the operators of the machine are really sadistic thrill-seekers, or the premises are overrun by fundamentalist zealots? In order to isolate the philosophical point which the experience machine is meant to illustrate, we have to suppose that all of these risks have somehow been neutralized. But this is very difficult to do, since we know that in real life we cannot control malfunctions. For the experience machine to yield any philosophically interesting results we must imagine ourselves in a world very different from our own - so different that any choices we make in that world might tell us very little about how we think our lives should go in the real world.[36]

But do these fears necessarily arise? Are we really unable to put them aside when responding intuitively to the thought experiment? In reality we often rely on technology without a deep understanding of its mechanisms. We drive our cars without having examined the condition of the brakes and engine, and allow doctors to implant our pacemakers without testing the device ourselves. Fears and concerns do arise automatically, but they do not seem to have total control over our decisions. Our fears of the uncontrollable may influence our decision, therefore, but we also have the opportunity to refine our responses based on the assurance provided by the thought experiment that the scientists and the machine are reliable.

Another fear that might arise is a fear of changing our lifestyles. Kolber illustrates this with an imaginary scenario:

> Imagine an investment banker with no relatives, working for twenty-five years with little or no job satisfaction. Her only pleasure in life is to come

> home after a twelve-hour work day and read passages from Zen Buddhist philosophers. In fact, she's come to believe that her life would be much better if she used her considerable wealth to move to Asia and study Zen Buddhism. Though she could have reason to believe that such a life would be better (...), she does not necessarily feel comfortable with such a drastic life change. It is natural to fear a drastic change in life-style. The experience machine offers us a drastically different life-style and, in this case, our fears are far greater than in the case of the investment banker. Not only is the experience machine life-style unusual und uncommon, it has never been undertaken before.[37]

But is plugging onto the experience machine simply a change of lifestyle? Is it just another change of lifestyle, like when an investment banker from the USA moves to Asia, only a more drastic one? Isn't there a fundamental difference between the two, namely that the person who moves to Asia opts for another part of the same reality, while the one entering the experience machine rejects reality itself? Certainly, there is a qualitative difference between the two. Another problem is that if it were only the fear of changing lifestyles that held us back from plugging in, vagabonds and adventurers who change their lifestyles frequently would be happy to enter the experience machine. This is not the case. They would most probably prefer their own lives.

Felipe De Brigard similarly refers to status quo bias as the main reason why people would not change their current lives for the one offered by the experience machine. He claims that "what mobilizes people's intuitive reaction against disconnecting is not solely a reflection on the nature of reality, nor their hedonistic

preference for pleasure, but also a psychological bias toward maintaining their status quo."[38] Status quo bias can be defined as "an inappropriate (irrational) preference for an option because it preserves the status quo".[39] De Brigard illustrates it with the following example:

> Marcia, a philosopher friend, acquired a 1932 edition of Kant's Critique of Pure Reason a while ago, at a time in which it cost no more than $20. Last week, at a party at her house, one of her book-lover acquaintances told her that such a particular edition of Kant's work had significantly appreciated in value, and that any book collector would surely be willing to pay between $100 and $120 for a good copy. Marcia knows she could get three or maybe even four decent newer copies of the same work for that amount of money. However, she is not in the least interested in selling her current copy - even though she would never pay $100 for the same book, if she didn't already have it.[40]

The story is a good illustration of status quo bias. Still, Marcia's decision cannot be justly compared with the decision of those offered the opportunity to plug into the experience machine. The question here is: what do people confronted with the experience machine already "own"? What is their *Critique of Pure Reason* that they won't let go of? For example, are their relationships in real life and the potential relationships in the experience machine analogous to the 1932 edition and the current edition of Kant respectively? The difference between the relationships in real life and relationships in the experience machine is far more profound than the difference between the two editions of Kant. Relationships in people's current lives are real relationships with real people, and as such are

never fully under one's control, while relationships in the experience machine are pre-programmed, fully controllable, and simulated.

But there is another consideration which justifies both decisions, that of not selling the book and also that of not plugging into the experience machine. The book has probably been part of Marcia's biography, even if she wasn't aware of its monetary value. In just the same way relationships are a part of everyone's biography, and even imperfect relationships are highly valuable. Now the question is whether Marcia would have sold the book if she had been offered a life in the experience machine with experiences identical to her real life experiences, book and all? Probably not, since such a life would not be real. The status quo bias interpretation misses the whole point of the thought experiment and the audience response. It's not only about changing one's lifestyle, but about exchanging a real life for a fake one.

Another psychological objection to such an exchange is the "roller coaster effect":

> On our first trip to the amusement park, we are awestruck by roller coasters - huge masses of twisting and turning beams. Though we are used to driving up and down hills and taking elevators, a ride on a roller coaster is well beyond our previous range of experiences. Furthermore, the intense pleasure that others tell us we will experience on the roller coaster does little to assuage our fears. But after one's first experience, some become zealous roller coaster fanatics. (...) This suggests that people who have actually connected to an experience machine might feel differently about connecting than would a virgin experience machine subject.[41]

But do people really feel differently, even if they understand that they would have the same experiences, albeit in an artificially simulated form? Once a person understands that the experiences will be identical, the decision can be made at a whole new level. The question is no longer solely "what it feels like" but also "whether it is real". Furthermore, the experience machine is not about trying new experiences but about altering all of our experiences for all time. Those facing the experience machine have the option of programming their current state of affairs into the computer and thereby preserve the current status quo until the end of their lives. They could keep having the experiences they have become accustomed to. But the question is whether they would be willing to enter the experience machine just to have the experiences of their current lives finalized.

A parallel is often drawn between the experience machine and drug use.[42] However, there are major differences between the two. The experience machine is physically harmless, does not result in addiction, and is not illegal. All these negative effects can be eliminated from the imaginary scenario. It is a fact that some people choose the world of drugs over their real lives. Now if people in the real world opt for a world influenced by drugs, with the attendant physical harm, addiction, and crime, wouldn't they happily enter the experience machine where these forms of physical and social damage could be avoided? It is questionable whether anyone, including drug users, would consider such a life a good life. But what is the situation when the use of drugs is justified? Let us consider the example of terminally ill patients experiencing constant severe pain. It is highly understandable that such patients strive for the elimination of pain, and are as ready to take painkillers or morphine as they would be to enter the experience machine in such a situation. The machine would

eliminate their real pain in just the same way as it would eliminate the fake pain of a fool who feels like there are nails piercing his feet when in reality it is only his mind playing tricks. The example of terminally ill patients with severe pain who are driven to eliminate that pain shows that the thought experiment might function differently with people in diverse situations.

This is sharpened by the so called "couch potato argument", which claims that those who "would maintain connections to an experience machine (...) are people who watch a lot of television and play video games all day and lead rather unfulfilling lives".[43] It is questionable whether couch potatoes would plug into the experience machine, but if they did, this would show that the Experience Machine Thought Experiment is not without its own set of preconditions. Someone who thinks that watching television is more important than doing the job he is responsible for or spending time with his family may be considered an addict just like the junkie who is not interested in anything but his daily fix. Yet the question remains whether those living the life of a couch potato would consider their lives to be fulfilling or whether they would opt for a different life if they could?

One last objection concerns the original wording of the thought experiment. Images of "neuropsychologists" stimulating the brain, the body "floating in a tank", and "electrodes attached" to the brain may scare the audience away from saying yes to the offer. But what if the conditions were altered so that a client had only to take a pill and lie down in his bed a home or he could be sure that the experience machine would perfectly preserve his body? Even with these new options, the experience machine does not seem any more appealing. What is the use of a perfect body if we cannot live in it?

## REFORMULATIONS

Like other thought experiments which are permanent subjects of ethics discourse, Nozick's Experience Machine Thought Experiment has also undergone several reformulations. Imaginary scenarios may be altered for at least two reasons. First, to disguise a well-known thought experiment for the purpose of testing its story and its functioning. This is especially needed when thought experiments are carried out among those familiar with the genre. Second, imaginary scenarios may be altered and amended if the underlying thesis of the original is put to the test. This is the way to bring to light the hidden implications of audience responses. The first method is typical for empirically testing thought experiments, while the latter serves their critical analysis.

## TESTING THE MACHINE

One of the reasons for altering the original imaginary scenario was to test it empirically. Although empirical examination seems alien to philosophy, experimental philosophers are keen to test thought experiments, challenging the intuitive answers provided by their authors.[44] Their aim is show that the intuitive responses are neither self-evident nor general, and are therefore unsuited to undergird an argument.

## A TRIP TO REALITY

One way of testing our preference for the real world opposed to the fake albeit pleasurable one is to see what we would choose if we had to make the decision from the other side,

namely from the world within the experience machine. One such scenario was formulated by Weijers:

> Imagine that you leave your family for a weekend to attend a conference on the Experience Machine thought experiment. While you are there, someone informs you that you are actually in an experience machine. She offers you a red and a blue pill. She explains that taking the blue pill will take you back to reality and taking the red pill will return you to the machine and totally wipe any memories of having being in reality. Being a curious philosopher you swallow the blue pill. It turns out that reality is fairly similar to the world you have been experiencing inside the machine, except that your experiences are a little mundane and do not feel quite as enjoyable as before. Some things are different, of course. You discover that nearly all of your friends and family are either in experience machines or do not exist in reality! Your father is there, so you spend time with him. But, a few conversations reveal that he is not really the person you know as 'Dad'. It is time to make the choice. Will you take the red pill so that you can go back to your life, family and friends with no idea that it is not in fact real? Or will you throw the red pill away and try to make the best life you can in the more real, but less enjoyable, surrounds of reality?[45]

Weijers claims that in his "experience of presenting the two scenarios, dramatically more people choose a life in an experience machine when considering the Trip to Reality thought experiment than when considering the Experience Machine thought experiment".[46] Certainly, a life with people

who are just as much alive, active, and interesting as we are, opposed to a life where most people are plugged into experience machines or where people are simply very boring, seems to be the attractive option. Still, one may ask, what would happen if the real life scenario were more attractive, offering a life "better" than our current one. Isn't Weijers' thought experiment just a form of escapism, an attempt to bypass the pain and suffering that are a part of real life?

Felipe DeBrigard tested a similar "trip to reality" scenario and proposed two different solutions:[47]

> It is Saturday morning and you are planning to stay in bed for at least another hour when all of the sudden you hear the doorbell. Grudgingly, you step out of bed to go open the door. At the other side there is a tall man, with a black jacket and sunglasses, who introduces himself as Mr. Smith. He claims to have vital information that concerns you directly. Mildly troubled but still curious, you let him in. 'I am afraid I have to some disturbing news to communicate to you,' says Mr. Smith. 'There has been a terrible mistake. Your brain has been plugged by error into an experience machine created by superduper neurophysiologists. All the experiences you have had so far are nothing but the product of a computer program designed to provide you with pleasurable experiences. All the unpleasantness you may have felt during your life is just an experiential preface conducive toward a greater pleasure (e.g. like when you had to wait in that long line to get tickets for that concert, remember?). Unfortunately, we just realized that we made a mistake. You were not supposed to be connected; someone else was. We apologize. That's

> why we'd like to give you a choice: you can either remain connected to this machine (and we'll remove the memories of this conversation taking place) or you can go back to your real life.'[48]

He also provided participants with two alternative endings. The negative variant depicted the following considerably repulsive scenario: "By the way, you may want to know that your real life is not at all as your simulated life. In reality you are a prisoner in a maximum security prison in West Virginia."[49] The positive variant, however, outlined a significantly more attractive life: "By the way, you may want to know that your real life is not at all as your simulated life. In reality you are a multimillionaire artist living in Monaco."[50]

De Brigard's research yielded results[51] which at first sight are not in accordance with the intuitions called forth by the original experience machine thought experiment. The negative vignette was clearly rejected, with only 13% opting for a life in prison, while 87% wished to stay connected. However, both the positive and the neutral vignette resulted in outcomes unpleasant for both sides in the debate. The "real life" in Monaco, and the current life stimulated by the machine were favoured by an equal number of respondents, while in the case of the original scenario without any amendment 54% opted for reality, opposed to 46% staying connected.

Although the results can be explained in various ways[52], they still seem to be embarrassing for both the hedonists and the reality party. De Brigard points out that "it would be a mistake to think that, since the quality of life affected the folk's decision, their choice was in effect dictated by a hedonistic preference. If that were the case, one would expect to see the opposite effect with the Positive scenario than with the Negative scenario, viz. a strong preference for reality".[53] However,

those who claim that it was reality that mattered to people most also find themselves in a difficult position, seeing as too many respondents opted for staying plugged into the machine. It seems that quality of life and reality were equally important factors motivating the respondents. Status quo bias must also be considered, since the results of the neutral vignette cannot be explained otherwise.[54] De Brigard concludes that "what these results suggests (...) is that, although people seem to value, at least to some extent, both contact with reality as well as pleasure, it is also true that, given the right circumstances, they are willing to give up either of them".[55]

One of the obvious conclusions concerning these empirical tests may go well beyond the reality vs. pleasure dilemma. If we consider why so many people opted for staying in the projected virtual reality, it becomes apparent that the life they actually live matters to them. Thus, one's actual life may often simply be assumed to be real life. There are several reasons supporting this identification between actual and real life. First, it seems very unlikely that even philosophers working on Descartes' Evil Demon problem would believe Mr. Smith and make a serious choice. Second, we assume that the people around us are real and not just bundles of information generated by a computer. It is impossible for a healthy mind to view other people as mere robots or holograms and to treat them as such. We cannot easily assume that everything around us is the result of blind causal processes, and certainly cannot build our lives around such an assumption. Finally, most people simply assume their actual lives are real and act with the resulting sense of responsibility.

Adherence to one's actual life might be a strong, though not irrevocable, motivation when making the decision. Transformed experience machine scenarios often ask us to imagine "an extremely happy" life producing "more

happiness for others", while asking us - mostly implicitly - to leave "home, family and friends, and never see them again".[56] But the fact that "most people would refuse such an offer" is not the equivalent of turning down the offer to plug onto the experience machine.[57] Many people are willing to change their lives for a "real" one, or would like to make the change, but do not have the means (or courage) to do so. This shows that there is something more important in our lives than preserving the status quo, even though we have a strong inclination to do so.

Although these tests shed light on psychological biases which influence our response to the outlined scenarios, they fail to address an essential question, namely the question of what we consider to be a good life. Respondents are trapped in dilemmas, forced to choose between remaining in their actual fake life or becoming inmates in a real high security prison, and their fear inclines them towards the first option. However, such a scenario does not show whether they would consider their actual life, if it were found to be fake, a good one. It only shows psychological preferences, but not what we understand by "good life".

## THE FAKE LIFE OF A BUSINESSMAN

In order to define what we mean by "good life" we must return to Nozick's original question concerning what really matters to us. To do so we can break with the assumption of the two worlds - the real one as lived and the fake one as projected by the experience machine - and examine what we consider real in our current lives. To put it another way, we are challenged to determine what matters in our life.

Shelly Kagan formulates a scenario which highlights what we consider as real and as fake in our actual lives:

> Imagine a man who dies contented, thinking he has achieved everything he wanted in life: his wife and family love him, he is a respected member of the community, and he has founded a successful business. Or so he thinks. In reality, however, he has been completely deceived: his wife cheated on him, his daughter and son were only nice to him so that they would be able to borrow the car, the other members of the community only pretended to respect him for the sake of the charitable contributions he sometimes made, and his business partner has been embezzling funds from the company, which will soon go bankrupt. In thinking about this man's life, it is difficult to believe that it is all a life could be, that this life has gone about as well as a life could go. Yet this seems to be the very conclusion mental state theories must reach![58]

Kagan designed the thought experiment to challenge mental state theory. If we compare the misled businessman with someone who was really loved by his wife and his family and respected by the community and whose business was truly successful, we may find that their mental states were identical. The same holds for the hedonistic calculus, at least if we focus on the businessman. The mental states of the two families - namely those faking and the ones living genuine lives - should be as different as their results on the hedonistic calculus. It seems that we want something more than merely to experience certain mental states or particular pleasures. We want the things in our life to be real, even if we don't have the means to ascertain their reality. Kagan's insight concerning the scenario is that a "natural response to these examples - the deceived businessman, the experience machine - is that these

people don't really have what they want. They think they do, but they don't. (...) The point could be put this way: what we want out of life is to have what we want out of life, and what we want is always far more than merely having certain kinds of experiences."[59]

## CONFUSING THE REASONS FOR NOT PLUGGING IN

Although it is usually the scenario of the experience machine which undergoes reformulations, the reasons Nozick gives for not plugging in may also be tested this way. Nozick claims that the first reason why we would refuse to plug in is that "we want to do certain things, and not just have the experience of doing them".[60] Kolber asks us to imagine ourselves having a "reason to believe that we were hooked up" to an experience machine and asks us point-blank: "Would you then care less about your parents or friends (that is, the people you call your 'parents' and 'friends')?"[61] The problem with this idea is that if we really knew that all our friends and family were simulations, this would influence not only how we care about them but also our fundamental understanding of relationships. One could opt to play along as if he were playing a video game which offers a life and identity in a virtual world. Making them no more than a game, however, changes the nature of our personal relationships. If one *had certain knowledge* that he was hooked up to an experience machine and that the world around him was fake, everything in that world would continue to seem complex and serious, yet it would all just be a video game.

Kolber presents a similar inverse thought experiment about personal identity. People may have different identities in the real world and in the world created by the machine. For example "a farmer in Oklahoma may choose to experience the

life of Mahatma Gandhi", and while on the machine he would clearly identify with this person.[62] But we can also imagine that we are currently plugged into the machine and that our identity in the real world is different from our current one:

> Since you are to imagine that you are currently on the machine, you may suppose that you are really John Doe when off the machine. Though you currently feel like you are yourself and you currently care about yourself in fundamental ways, your identity is composed only of neural stimulations in the brain of John Doe. In reality, John Doe may be much taller or older or of the opposite sex than you are. (...) When deciding whether or not to get off the machine, assume that you are told in great detail about the life of John Doe. But which identity would matter more to you, the identity that you have always associated with or the identity you are told that you 'really' have?[63]

We are certainly attached to our current identity, and most of us are probably against changing it. (This again depends on how a particular individual would respond to the question.) But we can formulate the challenge in another way: wouldn't it matter to us who we really are? Wouldn't we be concerned if our real character were a negative one, like a prisoner or a terrorist? I think that most of us would be. And the farmer who thinks he is Albert Szent-Györgyi, wouldn't he feel that his Nobel Prize Medal was a mere trinket without any real achievement behind it?[64] Giving up his identity as Szent-Györgyi would certainly be a great loss to him, but it would be a loss occurring at a moment when he was someone else in reality.

## WHAT DO EXPERIENCE MACHINE COUNTER-SCENARIOS TEACH US?

The experience machine counter-scenarios raise concerns similar to those generally formulated by critics. They mostly affirm (or simply disregard) the claim that we want to have more in life than mere experiences. Counter-scenarios do not disprove this simple claim but highlight additional concerns. The first of these is that our reasons for not entering the experience machine are different from those mentioned by Nozick. A desire to uphold our actual state of affairs (status quo bias), aversion, and fear of the unknown, as well as secondary doubts may all be possible explanations for our decision. Yet these claims do not go to the heart of the thought experiment, namely that the life that the experience machine offers is not just different, it is fake. Second, they show that under specific circumstances we may choose a different option and would perhaps enter the experience machine immediately. The latter is a more serious concern than the former, because if it is true, the account of the decision may rebut the arguments for not entering the experience machine.

## WHEN LIFE IS JUST PAIN

Imagine that all there is in life is just pain and suffering. There is no escape from it, not even for a second. There is nothing more than pain overwhelming everything. If you were offered to enter the experience machine in this particular situation, would you say yes to the offer?

This extreme imaginary scenario shows that we can imagine a situation when, I assume, everyone would opt to enter the machine to escape this world of suffering. One may

even draw an analogy between terminally ill patients in almost unbearable pain and no hope of relief outside of sedation. Most people would agree to doctors giving them large doses of narcotics in an effort to relieve their pain. But can we really compare what the experience machine can offer to the mere alleviation of pain? Moreover, is it possible to view terminal sedation or euthanasia as forms of experience machines?

First of all, testing the thought experiment on patients who are in severe pain also highlights one of the methodological problems with the experience machine: people experiencing different circumstances will answer differently. Basil Smith asks us to imagine what would happen if we carried out the thought experiment with "Christian religious leaders, with Japanese internet gamers, or with World War Two veterans" and to think about the possible answers we would get.[65] In much the same way, we might ask whether we would arrive at the same results in a lecture hall at the University of Vienna among young and healthy students and in a hospital ward for chronic patients who are in severe pain. Our intuitions say that the answers given under such different circumstances would diverge radically.

Second, the parallel between the experience machine and palliative care is only justified if we speak of irreversible anaesthesia or, under specific conditions, euthanasia. Palliative treatment offers a wide range of options which may connect the patient back to reality, while irreversible anaesthesia and euthanasia constitute an irrevocable break with their conscious being.

## SEDATION AND DEATH AS EXPERIENCE MACHINES

Michael Barilan draws a clear parallel between the experience machine and palliative care. He claims that "for many patients terminal sedation (i.e. irreversible anesthesia) and death (i.e. euthanasia, physician assisted suicide, suicide etc.) constitute a sort of Experience Machine".[66] He stresses that "in actuality many (perhaps most) terminal patients who suffer terribly do not ask for either sedation or euthanasia", and points at research which shows that "patients' wishes for hastened death" are not connected "with their severity of symptoms", but rather with their "stage of disease and hopelessness", which he sees as "empirical support to Nozick's argument".[67]

This position is confirmed by the distinction between pain and suffering.[68] Although these two phenomena usually go hand in hand, the connection is not inevitable. Mild pain might result in suffering, sometimes even immense suffering, while patients "may tolerate even extremely severe pain if they know what it is, know that it can be relieved, or know that it will soon end".[69] Thus an important factor in the nexus of pain and suffering is whether the patient is aware that the pain can be relieved and that it will end within the foreseeable future. Another important observation is the way patients handle palliative drugs. If they are aware that their pain can be relieved and have first-hand experience of this possibility, they are less likely to request palliative drugs, even if their pain comes back. "Once assured that relief is possible, suffering often subsides although the pain remains. It is difficult to relieve the suffering of patients who are frightened without also relieving their fear".[70] Moreover, fear of pain can also be a source of suffering.[71] A migraine, for example, may impact the whole life of the patient, even if it only appears irregularly for brief intervals: "They [i.e. migraine sufferers] suffer when

they do have actual pain and also when they do not."[72] This clearly shows that suffering is a much broader concept than pain. Suffering is always connected to patients' sense of the future and whether they are convinced that they can live their lives, that their pain can be relieved, and that they can handle their pain.

Although the above facts may support the central thesis of the experience machine thought experiment, namely that people tend to cling to reality, one should not forget the situations in which a patient simply wants to be free of pain. It is no wonder that Nozick uses the term "suicide"[73] in connection with the experience machine, since the only arguable reason to enter it can be to be freed of suffering overwhelming pain, which is supposedly the motivation behind most cases of voluntary euthanasia. The extreme situation of suffering from pain that overrides every other aspect of one's life is certainly a situation where a "yes" to the experience machine would seem reasonable.

However, this heightened scenario glosses over the point of the thought experiment which asks participants whether anything matters to them besides pleasurable experiences. In other words, the thought experiment asks readers to refine their intuitions about what constitutes a good life. A state of suffering from overwhelming pain is certainly not a state that anybody would consider "good". Moreover, those asking for terminal sedation or even euthanasia are not seeking pleasure but an escape from a reality they cannot bear. Terminal patients suffering from pain are not hedonists but find themselves far from the option of what one might call a good life: "as long as the person is suffering, it seems not to matter at all whether its source lies in reality or fantasy – he merely wishes for relief".[74]

This is underlined by the fact that the desire for different forms of euthanasia presents itself where patients cannot be

provided with sufficient support. Palliative centers and hospice groups mostly confirm the correspondence between care in suffering and the choice for or against active euthanasia: "They unanimously report the remarkable result that the initial desire to bring about death in a targeted manner falls silent as soon as we give the patient the opportunity to personally accept his death through effective pain control and human care".[75]

## THE EXPERIENCE MACHINE AND THE MEANING OF LIFE

When one draws a parallel between the experience machine and end-of-life care, there is a need to justify why entering the device can be compared to narcosis or even to death. Since the purpose of these measures, be they palliative care or euthanasia, is not the maximization of pleasurable experiences but control of or an end to suffering, the comparison is flawed. Even if narcosis offers the patient a similar state as the one to be found in the experience machine, death is not usually visualized as the door to a world of pleasure. But why do we still feel that the experience machine is somehow comparable to death and that Nozick's labelling it as a kind of "suicide" is more or less justified?

The answer is hidden in Nozick's original question concerning what really matters to us. In his book *Moralische Grundbegriffe (Basic Moral Concepts)*, Robert Spaemann uses the very same question to help place the experience machine in its proper anthropological context.[76] His approach differs from those mentioned above in two key ways. First, Spaemann uses the thought experiment to induce intuitions, thereby making his arguments about the "pleasure principle and reality principle" accessible to a wider audience. Second, he places the thought experiment in a broader context, i.e. he does not use

it as "the argument" but as a single element within his train of thought. Spaemann shows how the experience machine works when taken out of isolation and placed in a proper context.

In the second chapter of his book, and as the title already suggests, Spaemann makes an attempt to tease out what matters to us more, pleasure or reality: "Education or the pleasure principle and reality principle". However, his aim is not to argue for one system of ethics over another, but to provide a description of the human good. For Spaemann the basic purpose of ethics is not the question of what we "ought" to do, but "what we actually and basically want".[77] The latter constitutes the concept of a good life. It is clear from the starting point of the argumentation that both Spaemann's and Nozick's texts were written with the intention to clarify one basic question, namely what it is that we really want or, in other words, what matters to us.

Spaemann then attempts to outline the principle according to which we can distinguish everything that matters to us from everything that does not. He challenges hedonism as the "quickest" and "most common" answer. He points out that satisfaction is not the sole goal of our actions, which are often motivated by far loftier purposes.[78] Starting from the earliest stages of human life the pursuit of pleasure is accompanied by the pursuit of self-preservation. Both are also present in animals in the form of instinct, yet human beings are not necessarily bound by their instincts: "The world does not confront us in a way that has already been prepared by instinct for a species' specific environment, but as an open realm of infinite possibilities for satisfaction and also of infinite threats – because we cannot fulfill every one of our desires unpunished".[79]

Here Spaemann alludes to Freud who differentiates between the "pleasure principle" and the "reality principle"

when pointing out the two often contradictory drives in children's development. A child's libido comes face to face with its limitations as her understanding of reality expands. A child learns that "Reality doesn't comply with us. We have to comply with it. So we have to give up some parts of our desires in order to be able to fulfil other parts and uphold our existence".[80] Spaemann wants to provide support for Freud's discovery, and this is the point where he enlists the help of the experience machine.

He uses the thought experiment uncritically as a link in his train of thought, and assumes that his readers would turn down the offer to enter the experience machine. The reason he provides is the same as Nozick's, namely that someone choosing to enter the experience machine would "find himself outside reality".[81] The role of pain is also described as important, since it binds us to reality and alerts us to those things which threaten our self-preservation. Spaemann claims that the exposure of our instinct for self-preservation to danger is the primary source of our sense of being alive, even if we know that we will die someday.

At this point Spaemann uses another thought experiment to highlight the significance of our knowledge that we will die: "Imagine if we found out right now that we would never die."[82] In other words, we are to imagine that we will continue living forever "just as we are right now, (...) painlessly and without aging".[83] He then asks the readers whether this would be a desirable situation and claims that "anyone who has enough imagination to see what that would mean will quickly understand that it would be a catastrophe".[84] The reason is that everything would lose its significance, our personal relationships just as much as our actions. Such a state of being would deprive every moment of our life of its uniqueness whose

source is its never-recurring character. Spaemann concludes that "There can be no fulfilled existence without concern and care for the life endangered by death".[85]

By combining the two thought experiments Spaemann successfully arrives at a partial description of the human good. He shows that pleasure and self-preservation per se are not the sole criteria of what we consider to be a good life. We want something more than just pleasure or personal survival. Spaemann points at a "Dialectic of preservation and fulfillment", which is present not only at the individual, but also at the social and political level.[86] The successful balancing of these two principles is the key to a good life, since both extremes miss something very essential.[87] They both miss what we call the sense of life, which cannot be found in reducing it to pleasure or self-preservation, no matter how important they are in our lives.

The anthropological purpose of the text is clear, especially if one reads it in the context of the whole book. But what is the role of the two thought experiments in the line of argumentation? First of all, Spaemann uses them as central elements in his reasoning. The text itself was originally written as series of short lectures for the Bavarian Broadcasting Company in 1981, and was therefore designed to appeal to a broad audience. As Spaemann claims "it was [his] wish to come a little closer to the frequent family conversations of which Plato speaks" and to "attempt (…) to think about these [ethical] terms without terminological complexity and without learned presuppositions".[88] The two thought experiments played an important role in his argumentation since they highlight intuitions which are central to the concept of the good life. When talking to a general audience presumably consisting of non-experts there is a need to establish common insights which can be used as springboards for the discussion. There is always a risk that some listeners may not share these intu-

itions, or may simply disagree with the conclusions deduced from them. But without taking this risk and establishing such intuitive insights, philosophy would have to function in a vacuum. Second, Spaemann uses thought experiments as teaching tools, thereby introducing non-experts to the practice of philosophical reflection. Although the cases presented seem extreme, they are both accessible to a general audience, providing them with concrete and simple material for further reflection. Spaemann follows the basic steps of philosophy – beginning with everyday experience expressed in everyday language, followed by astonishment and doubt – and makes use of thought experiments to reach his goal.

Spaemann's example is proof that the experience machine thought experiment, similarly to other thought experiments, is not only a useful tool in the hands of analytic philosophers but is also suited to become part of a Continental philosophy aiming at a synthesis.

## Chapter VI

# THE LAST MAN ARGUMENT

The next thought experiment discussed in this book appears to be the odd one out for at least two reasons. First of all, it comes from the field of environmental ethics, which can be called the purview of bioethicists only if the boundaries of the discipline are generously drawn. Although environmental ethics is growing in importance due to the radical changes in the natural systems of our planet, medical ethics remains front and center in bioethical discourse. A thought experiment in environmental ethics is therefore unique and will differ substantially from experiments in other branches of ethics. This leads us to the second distinctive feature of the Last Man Thought Experiment, namely that it does not concern relationships between persons - at least if we stick to the original version of the thought experiment -, but tries to explore the question of the value of nature in itself. The Experience Machine, the Trolley Problem, and the Famous Violinist Scenario all focus more or less directly on interpersonal relationship, while the Last Man thought experiment seems to dismiss this question completely.

The Last Man Thought Experiment was originally formulated by Richard Routley[1] in his essay "Is There a Need for a New, an Environmental, Ethic?" published in 1973.[2] As the title

indicates, Routley was criticizing prevailing traditions of Western ethics in an attempt to show that there was a need for a new ethical approach to environmental questions. Although the original text contains four separate thought experiments, it was the first one, i.e. the Last Man Argument, which went on to become a fundamental part of environmental discourse. The thought experiment goes as follows:

> The last man (or person) surviving the collapse of the world system lays about him, eliminating, as far as he can, every living thing, animal or plant (but painlessly if you like, as at the best abattoirs). What he does is quite permissible according to basic chauvinism, but on environmental grounds what he does is wrong. Moreover, one does not have to be committed to esoteric values to regard Mr. Last Man as behaving badly (the reason being perhaps that radical thinking and values have shifted in an environmental direction in advance of corresponding shifts in the formulation of fundamental evaluative principles).[3]

## THE CONTEXT

Neither the choice of topic nor the means of discussion is a coincidence. Both Richard Routley and his wife Val Routley were important actors on the stage of environmental activism from the 1960s on. They found themselves in the middle of events which proved groundbreaking for the environmental movement in Australia:

> Throughout the 1960s and 1970s huge environmental struggles were erupting throughout Australia. Spec-

> tacular campaigns were fought for the Great Barrier Reef, the Colong Caves in the Blue Mountains, Fraser Island and Lake Pedder. Meanwhile, along the eastern coast of the continent the native forests, threatened with wholesale wood-chipping by the Forestry Commission, were providing a training ground for young environmental activists.[4]

Routley's engagement with the environmental movement suggests that his work in environmental ethics did not only serve theoretical purposes but aimed at moving things forward. The shift from the fields of logic and metaphysics to practical philosophy also changed his style of writing, inspiring him to rhetorically polish his ethical texts. These writings were designed to propagate a new view of nature and its value, and thereby to create a different attitude towards the environment. The paper in which the Last Man Argument was first formulated was presented at the XV$^{th}$ World Congress of Philosophy in 1973, where Routley had the opportunity to impress his fellow colleagues by presenting his ideas within the framework of a rhetorical masterwork.

Environmental activism was a typical phenomenon at the time and included intellectual endeavours to provide a conceptual framework for the environmental crisis. An important milestone was the publication of the report *The Limits to Growth. A Report for the Club of Rome's Project on the Predicament of Mankind* in 1972, which called attention to the burden that accelerating population- and economic growth placed on the natural environment.[5] Growing awareness of the ecological crisis also brought with it critical reflection concerning not only the current state of affairs but also the very foundations of Western culture and ethics. In 1967 Lynn White Jr. published his essay on "The Historical Roots of Our Ecologic Crisis" in which

he drew a sketch of how Western attitudes towards nature had evolved through history, pointing out the role of Christianity within that development.[6] He criticised not just "democratic culture" for producing the ecological crisis,[7] but also Christianity for putting man at the center of the universe:

> Since both science and technology are blessed words in our contemporary vocabulary, some may be happy at the notions, first, that, viewed historically, modern science is an extrapolation of natural theology and, second, that modern technology is at least partly to be explained as an Occidental, voluntarist realization of the Christian dogma of man's transcendence of, and rightful mastery over, nature. But, as we now recognize, somewhat over a century ago science and technology - hitherto quite separate activities - joined to give mankind powers which, to judge by many of the ecologic effects, are out of control. If so, Christianity bears a huge burden of guilt[8]

White's article produced a volley of responses not only from theologians but from thinkers throughout a Western world reflecting on its fundamental myth that man was created to extend his dominion over the world (cf. Gen 1:26-30). White claimed that the attitude of Western culture, where man is still at the center of the universe, had to be changed: "Despite Copernicus, all the cosmos rotates around our little globe. Despite Darwin, we are not, in our hearts, part of the natural process. We are superior to nature, contemptuous of it, willing to use it for our slightest whim."[9] His heavy criticism against this anthropocentrism leads him to search for an alternative path in the history of Christianity, which he finds in the person and legacy of Saint Francis of Assisi whom he later proposed

"as a patron saint for ecologists" because he "tried to depose man from his monarchy over creation and set up a democracy of all God's creatures".[10] The key to Francis' relationship to nature is the "virtue of humility", which is understood not just as individual virtue, but as a character trait vital to "man as a species".[11] White's ideas clearly point in the direction of what Routley argued for in his essay a year later.

However, it was not only White, but a whole trend against anthropocentrism in environmental ethics which may have set the scene for the formulation of the Last Man Argument. One of the first opponents of anthropocentrism was Aldo Leopold who is mentioned by name in Routley's essay. He invented what is called the Land Ethic which sees human beings as simple members of a larger community. This view "changes the role of Homo sapiens from conqueror of the land-community to plain member and citizen of it" who ought to respect "his fellow-members, and also respect... the community as such".[12] As Freyfogle resumes, "Leopold's land ethic rests on an understanding that humans exist within an integrated community of life that also includes other animals, plants, rocks, soils, and waters".[13] He invented the term "biotic community" to describe this relationship between man and other members of nature. Human actions may be evaluated ethically according to their relationship to the biotic community. They may be deemed ethical or unethical based on whether they contribute to its balance and sustenance or have a destructive effect instead. This last step clearly makes Leopold's argument a forerunner of Routley's Last Man. The development of ethics in this direction is seen by Leopold as "an evolutional possibility and an ecological necessity".[14]

The resituation of mankind did not only occur within ethical theories, however, but went far beyond them. The looming menace of severe crises, with the prospect of nuclear war

among them, induced visions of the end of mankind. Franklin J. Schaffner's film *Planet of the Apes*, first shown in cinemas in 1968, depicts a vision of an earth long after nuclear war where humans are an insignificant and inferior species dominated by intelligent apes. The film was followed by a sequel, Ted Post's *Beneath the Planet of the Apes*, which ends in an apocalyptic scene. Taylor, the astronaut who arrives on earth after making a long space journey in a state of hibernation, pushes the button of a doomsday device, thereby destroying all living beings, including superior apes and inferior humans. Norva Y.S. Lo and Andrew Brennan draw a parallel between the film's final scene and Routley's vision of the last man, claiming that the latter was the "philosophical version of Taylor's final act".[15] Although there is no direct evidence that it was the film which led Routley to the formulation of the Last Man Argument, the film certainly echoes the concerns of his contemporaries.

Routley does not mention Taylor in his essay but names another fictional character, Robinson Crusoe. This mention of Defoe is also used to point out the deficits of modern ethics, namely its preoccupation with man. Modern dogmas such as social contract theories are questioned by the loneliness of the castaway: "Crusoe comes from a society with a social contract in force. He is shipwrecked and thereby returned to a state of nature."[16] Prior the arrival of Friday with whom he has the opportunity to establish social norms once again he finds himself in a situation which is fairly unfamiliar to most people. Robinson is isolated from every fellow human and therefore seems to be beyond all familiar ethical or legal systems. As Melden notes, from the perspective of legal theory "it cannot be maintained that Robinson either has or does not have rights to freedom and well-being, for the question does not arise".[17] After the arrival of Friday "the situation changes and Robinson now can claim rights, and perhaps he also must

claim them"[18]. Yet the question remains whether rights or moral duties persist when Crusoe is in total isolation from other human beings but connected to the island, the ocean, and to nature as a whole. Although this problem is not explored by Routley, the solitary figure of Robinson facing nature certainly contributed to the development of the thought experiment.[19]

All these impulses led Routley to formulate a new environmental ethics, which is more than just "an extension of traditional ethics" tailored for man. It is a new ethics putting nature at its center.

## THE PRAGMATIC STRUCTURE OF THE LAST MAN ARGUMENT

Although the Last Man Argument comes from the field of environmental ethics, it shares the pragmatic structure of other practical thought experiments. It is rhetorically designed to induce a certain effect in the audience using the method behind the Dying Violinist or the Runaway Trolley Example.

First, it describes an imaginary scenario, namely a solitary last man who has the power to destroy nature or to sustain it even after his death (i.e. the extinction of all sapient beings from the world). Although the situation is extreme, it is more realistic than many other imaginary scenarios used in ethical thought experiments. It is a real possibility - though a thankfully unlikely one - that someone in the future will find himself in such a situation. The description of Mr. Last Man's situation is laconic. The audience is not given any information about his character or detail about his personal situation.

Second, the description of the Last Man facing the choice of destroying or sustaining nature refers to select aspects of reality which are morally important. At the heart of the matter

is the question of the value of nature and whether the existence of man is needed to confer this value. Do we owe nature something irrespective of its benefits for us? This can be translated into concrete situations, such as the everyday experience of those living in a technologically modern, capitalist world which views nature only as the raw material of its actions.

Third, the intuition induced by the image of the last man faced with the prospect of destroying the world is supposed to be a clear rejection of such a possibility. The expectation is that most people hearing the thought experiment would conclude that it was not right to destroy nature, even if one's actions did not affect a single human being.

Fourth, the experience of a system centered on mankind is contrasted with the intuition condemning the pointless destruction of nature. This makes it possible to test certain moral beliefs, theses or theories. Routley openly describes these presuppositions, namely the prevalent Western ethical view which he calls "chauvinism". Today we would call this position anthropocentrism since "it affirms that only human interests and concerns feature in moral deliberation and choice".[20] Anthropocentrism is based on the liberal principle according to which "one should be able to do what he wishes, providing (1) that he does not harm others and (2) that he is not likely to harm himself irreparably".[21] The intuition which considers nature valuable independent of human interest challenges prevailing Western ethics. The basic idea is that if nature has intrinsic value - and the audience arrives at this conclusion via intuitive insight - the basic principle of Western ethics, anthropocentrism, fails to apply.

Accordingly, Routley provides some guidance for the proper understanding and application of the thought experiment: "... what is permissible holds in some ideal situation, what is obligatory in every ideal situation, and what is wrong is excluded

in every ideal situation".[22] Routley challenges the universality of the Western liberal principle. He makes use of the intuition induced by the thought experiment in a deductive manner, putting intuitive judgement first and showing its validity for the particular imaginary scenario. (This is also the step that sets the scene for a bout of circular reasoning.) Deductive argumentation eliminates all escape routes from the ideological trap of the thought experiment: the audience is forced to juxtapose their moral theory (i.e. Western chauvinism) with their intuition.

## THE INTRINSIC VALUE OF NATURE

Routley does not conceal the importance of intuition to his argumentation. At the end of his essay, he criticises chauvinistic ethical theories which "try to offer some rationale for their basic principles" in contrast with "intuitionistic theories".[23] He seems to be well aware that his arguments rely heavily on the intuitive response of the audience to the imaginary scenario. The expected intuition, namely the rejection of the last man's intention to destroy nature, was supposed to support another thesis concerning the intrinsic value of nature.

This thesis is not unusual among environmentalist ethicists. As Carter notes "it should come as no surprise that numerous environmental ethicists seek to establish that certain nonhuman natural entities possess intrinsic value, for their possession of intrinsic value would most likely provide the strongest plausible reason for preserving them when they might otherwise be destroyed for their instrumental value as, for example, economic resources".[24] He names Routley's thought experiment as "the most cogent of the available arguments that might be put to such use".[25] Indeed, if nature has intrinsic

value, and this idea becomes widely accepted in society, representatives of non-anthropocentric positions will gain moral ground and actions potentially harmful to nature will need to be justified in a non-anthropocentric manner.

But what does "intrinsic value" mean? It is a key concept in philosophy and especially in ethics, but its meaning is often too opaque due to overuse. This ambiguity is confirmed by Zimmerman's attempt at a preliminary definition: "The intrinsic value of something is said to be the value that that thing has 'in itself,' or 'for its own sake,' or 'as such,' or 'in its own right'".[26] Martin Peterson and Per Sandin point out the same ambiguity concerning the use of intrinsic value in the Last Man Argument. They distinguish four distinct types of value: instrumental value (something "is valuable as a means to some end"), final value (something is "valuable for its own sake, rather than as a means to something else"), extrinsic value ("the source of the value lies outside the object itself"), and intrinsic value. They claim that "the question the Last Man Argument seems to settle is not whether the value of nature supervenes on properties that are internal to nature itself. The question is whether nature has final value, i.e., is valuable not just as a means to an end".[27]

Yet with this attempt to attest to the final value of nature, Routley contests one of the basic dogmas of Western civilization, which is the definite distinction between persons and objects. As Kant writes in his *Groundwork for the Metaphysics of Morals*, "[t]he human being, however, is not a thing, hence not something that can be used merely as a means but must in all his actions always be considered as an end in itself."[28] Yet if nature has final value and cannot "be used merely as a means", it is placed in a category thus far reserved for persons alone.

This interpretation is plausible, but there are signs that the concept of intrinsic value is actually more ambiguous. Al-

though it might be true that Routley's aim was to go beyond the view of nature as simple means to human goals and to demonstrate its final value, the thought experiment clearly opens itself to an alternative interpretation. After the death of the last man there is no one to recognize and determine the value of nature. No one can make a judgement about any sort of value of any existing object anymore. This goes beyond the predicate of value realism claiming that "value claims (such as friendship is good and burning baby's feet for fun is bad) can be literally true or false; that some such claims are indeed true; that their truth is not simply a matter of any individual's subjective attitudes or even of the attitudes of some larger collective; and that facts about value enjoy a certain metaphysical independence from other matters of fact."[29] The question whether mankind ought to shape his actions according to the value content of reality or just follow its subjective constructions of value has already been decided. The point is that the scenario presented in the thought experiment extinguishes all subjects who might make subjective value judgments, perform actions or have pleasurable feelings concerning nature. It is contradictory to ask whether nature has any value after the extinction of all rational beings, since to understand this post-human situation we would have to suspend our subjectivity, which is certainly impossible. It is not possible to imagine anything without acting as a subject.

Thus, if Routley considers the Last Man Argument as proof that nature has value even after the extinction of the last human being, he is inconsistent in his argument. Routley appeals to intuitive judgement to support a statement about a fact absolutely independent of this judgement. His argument suggests that the intuition of the audience rejecting the destruction of nature by Mr. Last Man proves the value of nature which is, at the same time, independent of the audience's judgement.

## THE RED BUTTON

Now, since it is impossible to decide the value ascribed to nature based on Routley's article, there is a need for a better description of the intuition induced by the imaginary scenario. The imaginary scenario needs to be transformed in order to isolate all aspects of the intuition it evokes. The Last Man Argument describes an action whose consequences are only realized in the future, after the death of the agent. Neither he nor anyone else will witness the destruction caused by his deeds. Thus, there is a clear barrier between the world he lives in and the world to dawn after his death. But does it make a difference if we transform the imaginary scenario by changing the linear succession of the two worlds to a parallel existence? This can be demonstrated by the Red Button Scenario.

> Imagine two enormous territories. In Territory A there are no beings capable of suffering or rational thinking. Territory B is inhabited by humans. The two territories are divided by a huge fence which makes it impossible for the humans to have any experience of the world behind the fence. However, the two territories are connected by a mysterious wire which ends in a red button right at the centre of Territory B. The inhabitants know about the button, and also know that once they push the button it destroys something in Territory B. They do not know what is destroyed - it can be the ugliest little worm or the most beautiful bird -, but are aware that they will never experience this destruction or its aftermath in any way. In other words, neither they nor any other being capable of suffering will be affected in any way by the pushing of the button. Is it wrong for the people of Garden B to push the button?[30]

Let us assume that the answer to this question is a resounding "yes". Most readers are probably inclined to reject the pushing of the button, thereby attributing a moral quality to the action. But what lies at the heart of this intuition? It is not the quality which perishes due to the pushing of the red button, but the essence of destruction itself. Since it is impossible to know what lies on the other side of the fence, and whether anything there has intrinsic or final value, the intuition can certainly refer to objects with such values. It does not make any sense to push the button under these conditions. The same holds for the Last Man Argument. The audience fails to find any sense in destroying nature (or anything) without a rational purpose. It is not the assumed value of the objects on the other side or the post-human future which induces this intuition, but the protest against pointless destruction.

## ALTERNATE VERSIONS OF THE LAST MAN ARGUMENT

Among the numerous reformulations of the Last Man Argument there are three which were formulated within the original article. Although it is the Last Man Argument which attracted the most publicity, these reformulations would stand their ground as individual thought experiments. Routley, however, uses them to expand his argumentation.

## THE LAST PEOPLE

In the Last People Example Routley describes a group of people who know that they are the last of their kind.[31] They are unable to reproduce themselves due to the damage caused by some sort of radiation. There is no chance that rational beings

will ever take their place, thus a succession is ruled out this way too. The Last People decide to engage in activities through which they exploit all natural resources on earth: "They humanely exterminate every wild animal and they eliminate the fish of the seas, they put all arable land under intensive cultivation, and all remaining forests disappear in favour of quarries or plantations, and so on."[32] However, in contrast to Mr. Last Man, they are able to justify their actions: "they believe it is the way to salvation or to perfection, or they are simply satisfying reasonable needs, or even that it is needed to keep the last people employed or occupied so that they do not worry too much about their impending extinctions."[33]

Routley finds their actions and the reasoning that "they do not wilfully destroy natural resources (...) environmentally inadequate".[34] This shows that Routley did not mean to use this second version as part of the thought experiment, but merely as an example for how we might be misled if we see environmental ethics only as an extension of Western chauvinistic ethics.

Routley is right in claiming that the Last Man Example does not serve his purposes. It does not elicit the intuition that the Last People's behaviour is ethically wrong, for their actions are performed with good aims, e.g., to sustain their lives or to prevent their suffering. Most people would probably not condemn their behaviour. A very precious piece of art could with good reason be used and even destroyed under certain conditions - at least when human lives are at stake - and the same is true for the destruction of natural objects.[35] Routley admits that the intuition induced by the argument does not fit with his idea of "an environmental ethic" according to which "the last people have behaved badly; they have simplified and largely destroyed all the natural ecosystems, and with their demise the world will soon be an ugly and largely wrecked place."[36]

## THE GREAT ENTREPRENEUR EXAMPLE AND THE VANISHING SPECIES EXAMPLE

The reasons given in case of the Last People for the exploitation of nature are intuitively justified, especially because they happen to be mostly humane purposes. Accordingly, Routley transforms the imaginary scenario to point out purposes justified by Western ethics, which the audience might intuitively reject. He calls attention to the logic of industrialist societies and their relationship with nature to showcase the failings of ethical chauvinism.

> The last man is an industrialist; he runs a giant complex of automated factories and farms which he proceeds to extend. He produces automobiles among other things, from renewable and recyclable resources of course, only he dumps and recycles these shortly after manufacture and sale to a dummy buyer instead of putting them on the road for a short time as we do. Of course he has the best of reasons for his activity, e.g. he is increasing gross world product, or he is improving output to fulfil some plan, and he will be increasing his own and general welfare since he much prefers increased output and productivity. The entrepreneur's behaviour is on the Western ethic quite permissible; indeed his conduct is commonly thought to be quite fine and may even meet Pareto optimality requirements given prevailing notions of being 'better off'.[37]

The behaviour of the industrialist Mr. Last Man is probably intuitively rejected by most people. Reasons such as "increasing gross world product", "improving output to fulfil some plan",

or "increasing his own and general welfare" are seen solely as the means to an end and thus have only instrumental value. The intuitive response, which is one of repulsion, suggests that the integrity of nature is more valuable than the human goals mentioned. Routley claims that "the entrepreneur's behaviour is on the Western ethic quite permissible; indeed, his conduct is commonly thought to be quite fine and may even meet Pareto optimality requirements given prevailing notions of being 'better off'."[38] Interestingly the Great Entrepreneur Example lacks anthropocentrism. It is not man who is at the center of Mr. Last Man's actions, but only the optimization of the industrial process and the expansion of the industrial system. Thus, industrialism is falsely identified with anthropocentrism. Despite this failing, the Great Entrepreneur Example is a clear and legitimate critique of the contemporary industrial system and its blindness to all natural systems.

Routley does not stop at the analysis of the logic of industrialism, but also targets the flipside of the coin: consumerism. He uses the actual example of the hunting of the blue whale, which had brought the population to the verge of extinction. Routley describes the blue whale as a "mixed good" which has both public and private value.[39] He focuses on the latter aspect, however, namely use of the whale "as a source of valuable oil and meat".[40] In the example the possible harm to individuals or to society is neutralized so that whale hunting appears to be almost neutral with regard to human individuals or communities: it "does not harm the whalers; it does not harm or physically interfere with others in any good sense".[41] Moreover, whalers do not stand in the need of hunting, since those who might be upset by whale hunting are "prepared to compensate the whalers if they desist".[42] Thus it is safe to say that the hunting and extinction of the blue whale do not harm anyone. (Although Routley cannot eliminate the suspicion that it might

still harm others, the most obvious harm to man is neutralized. This is due to the fact that the Vanishing Species Example is an actual example with already existing implications and presuppositions in the audience.) Routley claims that "the behaviour of the whalers in eliminating this magnificent species of whale is accordingly quite permissible - at least according to basic chauvinism. But on an environmental ethic it is not".[43] The point is that chauvinism, which is the underlying moral framework of consumer society and the logic of the free market, is simply blind to the ethical problem of impoverishing the natural world by hunting.

Both the Great Entrepreneur Example and the Vanishing Species Example are much closer to the Lebenswelt of the audience than is the Last Man Example. Industrialism and consumer society have been the fundamental experience of the Western world since the 1970s. Using these examples Routley manages to point out the blindness of industrialist and consumer mentality towards nature and its value. His examples induce obvious intuitions to reject the deeds of industrialist Mr. Last Man and the whale hunters, and he also succeed in pointing out the evil of the senseless destruction of nature. Neither production nor consumption appears to have the final value that would justify the destruction of nature.

Intuitions would certainly change if the purpose of the actions was altered, for example if the whalers were hunting for the last representative of the species to escape starvation or industrialist Mr. Last Man were attempting to design an extra safe and fast car to escape from the wild beasts threatening his life. These examples fail as criticism of anthropocentrism since neither industrialism nor consumerism is anthropocentric in that neither ideology has final value. Numerous practical examples may be cited in support of this claim. The artificial environment in centers of industrial production is often

harmful to human health, as is the overconsumption of food in certain rich Western countries. These factors are certainly opposed to the anthropocentric principle.

## OTHER REFORMULATIONS

In his original article, Routley proposed three reformulations of the Last Man Example. However, he was not the only one to make use of various versions of the thought experiment to modify or refute the original theory which criticized anthropocentrism and established the intrinsic value of nature.

## THE GREAT CHAIN OF BEING

The original version of the Last Man Example does not identify any particular part of nature that is destroyed by Mr. Last Man's actions. It simply speaks of "every living thing, animal or plant", thereby drawing the line between the animate and inanimate. But do our intuitions change if this line is drawn between plants and animals?

Mary Anne Warren proposes a thought experiment which both challenges and refines the intuitive judgement evoked by the original example. She asks us to imagine

> that a virulent virus, developed by some unwise researcher, has escaped into the environment and will inevitably extinguish all animal life (ourselves included) within a few weeks. Suppose further that this or some other scientist has developed another virus which, if released, would destroy all plant life as well, but more slowly, such that the effects of the second virus would not be felt until after the last

> animal was gone. If the second virus were released secretly, its release would do no further damage to the well-being of any sentient creature; no one would suffer, even from the knowledge that the plant kingdom is as doomed as we are. Finally, suppose that it is known with certainty that sentient life forms would never re-evolve on earth (this time from plants), and that no sentient aliens will ever visit the planet. The question is would it be morally preferable, in such a case, not to release the second virus, even secretly?[44]

Warren sees the intuition prohibiting the release of the virus as proof that "we do not really believe that it is only sentient - let alone only human - beings which have intrinsic value."[45] The problem again is that it simply does not make any sense to release the second virus. It is nothing more than - to use Robin Attfield's term - vandalism.[46] The author does not provide any further hints about what purpose the release of the plant-killing virus would serve. The thought experiment would evoke different, but very significant intuitions, if the second virus would not only "destroy all plant life as well, but more slowly", but also extend the life of animals both in time and quality. Another effect would be that no one would suffer from the knowledge of the future extinction of animals or plants. Now the price of the wellbeing of all animals would be the desolation of the kingdom of plants following the extinction of all sentient beings.

The audience can answer this challenge in two possible ways. Respondents can either support or reject the release of the second virus, and both decisions carry value judgements concerning the place of animals and plants in the value-hierarchy. One who rejects the release considers both plants and animals of equal value, or ranks plants even higher. With the

support of the release, one discloses his value preference in favour of animals. However, this altered narrative can also result in a dilemma situation for some, since the value of the limited life of particular animals is contrasted here with the seemingly eternal flourishing of the flora. The possibly mixed intuitive answer to the release of the modified virus affirms the value of both plants and animals - since presumably no one would see the destruction of any of these kingdoms as desirable or good -, but also shows a hierarchy in which animals, especially human beings, are valued more highly than plants. Two problems remain, however. First, how can we make a judgement about an imaginary situation in which no subjects remain? And second, how does the point of the thought experiment change when the audience learns of the reasons for Mr. Last Man's actions?

## MR. LAST MAN AND VIRTUE

Peterson and Sandin point out that the original version of the Last Man Example lacks an explanation of why Mr. Last Man would destroy everything around him: "...in the original version of the argument, one does not learn anything about Last Man's motives or character traits. What kind of person is he or she? A frugal, humble and earthy person? Or a gluttonous, arrogant and greedy one? And what are Last Man's motives for destroying living things?"[47] And how would our perception of the original thought experiment change if we knew his motive?

In his article "The Good of Trees" Robin Attfield formulates a slightly modified version of the Last Man Example which - instead of proposing the destruction of nature in general - focuses on the destruction of particular objects:

> The last man knows, in my version of this example, that all life on this planet is about to be terminated by multilateral nuclear warfare. He is indeed himself the last surviving sentient organism, and knows that he too will die within a few minutes; but he also happens himself to be possessed of a workable missile capable of destroying all the planet's remaining resources of diamond. The gesture of doing so would certainly be futile, but for himself it has a symbolical significance; and the question with which he is faced, and which we can ask about his projected act, is whether it would do any harm or destroy anything of intrinsic value.[48]

What is special in this argument is that Attfield uses diamonds (non-organic objects) in his example instead of species of plants or animals. He does so in order to underline the unity of nature and the contribution made by all existing objects to its diversity. The use of diamonds also alludes to monetary value, since Western culture regards diamonds as one of the most precious objects created by nature. However, the question from the original version of the Last Man Example remains, namely whether it is morally wrong to destroy these diamonds if there will be no sentient being to value them in the future.

What is also unique in the modified example is that Attfield gives the reason for Mr. Last Man's action. It is a "gesture" with "a symbolic significance".[49] One can only guess what this symbolic significance might be: are diamonds the symbol of his anger about the passing of the world or of his hatred towards human greed? Through his reformulation, Attfield subtly pushes the question towards the realm of virtue ethics. This is underlined by a thought experiment he previously proposed in the article, in which he asked the reader

"to imagine ... a world in which there are no conscious experiences and no activities" and to ask whether complexity or its loss would change the value of this world.[50] He claims that "such a thought-experiment may be barely possible until we imagine the agency which might carry out the deprivation".[51] Now the question is: why would one want to make the world a less diverse and less complex place? Attfield sees "no reason for preferring a slightly more diverse inanimate world to a slightly less diverse one, unless its constituents are objects of someone's or something's experience", and arrives at the conclusion that "it is rather the wrongness of vandalism which accounts for our objections to the elimination of species".[52]

If the thought experiment were a simple description of Mr. Last Man's final action - pushing the button and destroying all diamonds, or the flora or fauna of the earth, - one would certainly refrain from describing either the action or the agent as virtuous: "... it seems safe to say that wanton destruction of parts of nature, as Last Man indulges in, is not something a virtuous agent would, characteristically, carry out".[53] At this point, however, a further problem arises: what actions are to be considered virtuous in the situation of Mr. Last Man which is radically different from ours? Not only is all social context missing from these scenarios,[54] but the protagonist is vested with powers and placed in circumstances which are unimaginable to ordinary people.

Peterson and Sandin propose an imaginary scenario with three alternate versions focusing on the motivations of Mr. Last Man:

> Last Man manages to escape in his spaceship just before the Earth crashes into the Sun, and he is now circling a distant planet. The on-board super computer informs him that there is some Earth-like but

> non-sentient life on the new planet, which will never evolve into sentient life forms since the new planet will crash into its sun within a year (which will, of course, destroy all living organisms). However, Last Man can delay this process for five years by firing a nuclear missile that will alter the planet's orbit. There are no other morally relevant aspects to consider.[55]

In the first alternate version Mr. Last Man aborts the launch of the nuclear missile "because he feels that the noise caused by the missile launch would make him feel distracted for a short moment while reading Death on the Nile by Agatha Christie", while in the second his reason for launching the missile "is that during his entire adult life, he has had a yearning to blow off a really big nuclear missile" and "when he fired the missile, he knew that he would take great pleasure in setting it off".[56] In the third version he launches the missile because "he believes – incorrectly, as it happens, but based on the best available scientific evidence – that plants and other organisms in this part of the universe benefit from being exposed to radioactivity".[57]

The biggest problem in deciding about which version of Mr. Last Man was most virtuous stems from the extreme nature of his situation. Under normal conditions there is no doubt that saving the entire flora of a planet for five years longer is more important than avoiding brief distraction from a good read. But Mr. Last Man in his space cabin might deem Agatha Christie's world to be more important than the life of the plants on a distant planet. Similarly, although under normal conditions "to enjoy performing very violent acts is simply wrong", launching a missile for fun and, as a side effect, prolonging the life of the flora on a distant planet does not evoke clear intuitions. Although Peterson and Sandin hold that "the motives from which one acts has a larger influence on Last

Man examples than intuitions about the value of nature", the vagueness of intuitions evoked by these diverse motives does more than create further doubts about the solidity of the Last Man Example. The extreme conditions make it (almost) impossible to arrive at a solid intuitive judgement concerning the ethical value of the agent's motives.

## THE VALUE OF LAST MAN THOUGHT EXPERIMENTS

As we have seen, the Last Man Example claims an important place in environmental ethics to the point "that enough scholars have had strong enough intuitions about it to not question the intuition".[58] As demonstrated by the previous examples, however, neither the intuition nor the judgement resulting from it are as solid and clear as they might appear at first glance. Minor alterations of the original example might yield alternative intuitions, and one's judgement about the intrinsic value of nature does not seem to result directly from the intuition either.

In his *Principia Ethica* published in 1903 G.E. Moore formulated one of the predecessors of the Last Man Examples. It goes as follows:

> Let us imagine one world exceedingly beautiful. Imagine it as beautiful as you can; put into it whatever on this earth you most admire—mountains, rivers, the sea; trees, and sunsets, stars and moon. Imagine these all combined in the most exquisite proportions, so that no one thing jars against another, but each contributes to the beauty of the whole. And then imagine the ugliest world you can possibly conceive. Imagine it simply one heap of filth, containing ev-

> erything that is most disgusting to us, for whatever reason, and the whole, as far as may be, without one redeeming feature. (...) The only thing we are not entitled to imagine is that any human being ever has or ever, by any possibility, can, live in either, can ever see and enjoy the beauty of the one or hate the foulness of the other. Well, even so, supposing them quite apart from any possible contemplation by human beings; still, is it irrational to hold that it is better that the beautiful world should exist than the one which is ugly? Would it not be well, in any case, to do what we could to produce it rather than the other? Certainly I cannot help thinking that it would; and I hope that some may agree with me in this extreme instance.[59]

According to Moore's isolation test objects have intrinsic value if they are considered valuable even when isolated from everything else. "He advises us to consider what things are such that, if they existed by themselves 'in absolute isolation,' we would judge their existence to be good".[60] With the help of this thought experiment, we can discover what confers value upon entities in our world. Moore isolates beauty in this example, which is objectively good per se. Routley undertakes the same endeavour in his Last Man Example; by skipping forward in time after the death of the last man he isolates nature from human subjectivity. The only thing he misses is that subjectivity cannot be eliminated, not even by the death of the sole remaining subject. Even after the death of Mr. Last Man the audience of the thought experience are present as subjects imagining and enjoying the beauty, diversity, and liveliness of nature, and evaluating it as something good.

However, when we isolate nature or beauty from everything else, a slice of our intuitive structure manifests itself.

William Grey is right to point out in his critique that "value intuitions depend crucially on the nature of evaluators".[61] In other words, these intuitions do not inform us about some objective quality outside our human constitution but reveal something essential concerning the human good. He ironically notes that "it is far from clear that our preference would be shared by, say, a dung beetle or a blowfly".[62] What Moore pointed out is "the existence of deep-seated aesthetic intuitions widely shared among humans", while Routley revealed "a widespread, though sadly not universal, biophilia - an affinity for rich, diverse, complex and beautiful biological systems".[63]

It is also important to clarify that the Last Man Example was not constructed for Mr. Last Man, but for a contemporary audience. They are the ones who have to imagine the situation Mr. Last Man finds himself in and to reflect on what is morally right or wrong in such an extreme case. Why do we intuitively think that the moral rules born under "normal" conditions - i.e. conditions radically different from the situation of the lonely last man - which have always stood us in good stead, should be observed by Mr. Last Man? Carter elaborates further on the reasons for these intuitions. First, "projected moral properties only retain their social utility while we remain within their grip".[64] But what happens when the social system is dismantled? Are we set "free" from these moral properties? Carter suggests that "we are likely to remain in their grip even when the social pressure that plays a part in their formation has ceased. Consequently, it would not be surprising if the last person still saw the world as if it possessed real and objective moral properties, and thus felt the gratuitous destruction of natural entities to be morally wrong, even when no sentient being would ever suffer from their loss."[65] Second, Carter points out that it is the current audience and its intuition which is challenged by the Last Man Argument, not the last man himself: "the Last Person

Argument is not actually addressed to the last sentient being who is conceived as living at some future time. It is addressed to *us*. And it is addressed to us *now*. The Last Person Argument is deployed by environmental ethicists in order to help us clarify in our own minds the values we currently hold."[66] Third, Carter draws our attention to the exceptional nature of Mr. Last Man's circumstances: "the situation in which the last person finds himself or herself is a highly unusual one - one that everyday moral thinking would be unlikely to have evolved to take into account."[67] But is it correct to apply general, socially accepted principles - or rather every-day, ordinary moral rules - to an extraordinary situation in which these very rules might have lost their function?

Even if, for the reasons above, we cannot determine with ethical certainty how Mr. Last Man should act, this does not lessen the value of the Last Man Example. The thought experiment achieves what it was originally designed to do: it calls attention to the value of nature, respect for which is part of our intuitive constitution. It seems that even Routley was cautious of creating an ethical theory based on intuition and using the Last Man Example. In the original article and following his criticism of Western ethics, he underlines that "these ethical and economic theories are not alone in their species chauvinism; much the same applies to most going meta-ethical theories which, unlike intuitionistic theories, try to offer some rationale for their basic principles."[68] He bases his theory on intuitions against any senseless destruction of nature. Although he does not prove the intrinsic or final value of nature, he does successfully pinpoint the deep human revulsion against the destruction of nature.

The Last Man Example was not only successful in academic discussions but yielded practical results as well. It contributed to the development of the discourse about the intrinsic value

of nature in academics, politics, and policy-making: "For example, the United Nations (U.N.), governments, and nongovernmental organizations hold that nature has intrinsic value."[69]

## A THEOLOGICAL AMENDMENT

As we have seen the biggest misunderstanding concerning the Last Man Argument is that it pointed at something - the value of nature - independent of subjectivity. Since we cannot simply switch subjectivity off, we arrive at the problem Bishop Berkeley identified as early as the 18th century. He resolved his dilemma with the statement "esse est percipi" and called attention to the omnipresent subjectivity of God. In Christian theology God is the Creator creating the world continuously (creatio continua). Thus, even after the death of Mr. Last Man God would continue to sustain the world and find value in it. Thus, the intrinsic value of creation still holds, but it is dependent on the Creator.

This is what Holmes Rolston points at in his reformulation of the Last Man Example:

> Suppose, a century hence, that in a tragic nuclear war each side has loosed upon the other radioactive fallout that sterilizes the genes of humans and mammals but is harmless to flora, invertebrates, reptiles, and birds. That last race of valuers, if they had conscience still, ought not destroy the remaining biosphere. Nor would this be for interest in whatever slight subjectivity might remain, for it would be better for this much ecosystem to continue, even if the principal valuers taken out. That verdict would recall the Genesis parable of the first judgement, where,

> stage by stage, from lesser to higher forms, goodness is found at every level.[70]

On a theological level it is God the Creator who serves as the ultimate guarantor of the value of every created being, and this is what takes us beyond the anthropocentric view Routley wished to transcend.

## Chapter VII

# THE TROLLEY PROBLEM

We rarely have to face true dilemmas in everyday life. Opting for or against taking an umbrella in case it rains is not a real dilemma, neither is choosing between a hamburger and vegetable soup for lunch, or even deciding to pass or fail a student as a university professor, though such choices might produce headaches at times. True dilemmas emerge in times of crisis: for example, when one is living under a dictatorship. Totalitarianism seems to create situations where there are no options that could be termed good or that a person might choose in good conscience.[1] Thus, it is no coincidence that films set in the Nazi era often center on the kinds of dilemma produced by these very regimes.

In the film *Sophie's Choice*, based on the novel by William Styron, a Catholic woman who has been deported to Auschwitz with her two children must decide which will die and which will be given a small chance of survival. As the deportees arrive at the concentration camp, the SS doctor sorts them into two columns: one for the weak who will be put to death in the gas chambers, and the other for those capable of hard work who might thus survive. When she approaches the SS doctor she says, "I am not a Jew. Neither are my children! They're

not Jews. They are racially pure. I am a Christian. I am a devout Christian." He replies, "You may keep one of your children. (…) You're a Polack not a Yid. That gives you a privilege, a choice. (…) Choose or I'll send them both over there!" The offer is a terrible one; "Don't make me choose! I can't!" replies Sophie, "Take both children away!" Sophie then pushes her daughter away from her and shouts out, "Take my little girl!"[2] The dilemma that Sophie was forced into is clear. She had two equally bad options: either to let both of her children die or to send one of them to the death chambers and thus give the other a small chance of survival. This description, obviously, is a vague one, but shows the horror involved in both possible choices: "Either way, Sophie will end up doing something she ought not to do (consenting to the death of one of her children) thus failing in her obligation to protect the life of that child."[3]

In his film, *Dekalog, Eight*, Krzysztof Kieślowski presents a similar dilemma, but one that has a positive outcome. Set in the mid 1980s in Poland, Elzbieta, an American professor, arrives in Warsaw where she attends a lecture by an older university professor. She proposes an imaginary case for discussion, which involves a six year old Jewish girl seeking asylum from the Nazis in Warsaw in 1943. However, the couple who were supposed to act as the godparents at her baptism and to take her to a host family turned their backs on her. As it becomes clear, this is a not a fictitious case but a real one: Elzbieta was the little Jewish girl and Zofia was the ethics professor who turned her down. The reason she did so was not because the couple did not want to give false testimony about her being a Christian – as Elzbieta thought. Zofia, who was a member of an underground group working to save Jews, was informed that the couple who had offered asylum to the girl were conspiring with the Gestapo. If they helped the girl, they would be putting the whole group, together with all its efforts to save Jews, in

grave danger. A tough dilemma: save the little girl or protect the underground group, including both the work it was doing and the life of its members. As it turns out, the information about the couple conspiring with the Gestapo was false, and the girl managed to survive. But the choice, even after more than forty years, still haunts Elzbieta and Zofia.[4]

Both courses of action - saving Elzbieta or the underground group - can be justified: to save the greater number of lives. But this answer seems unsatisfying. The principle of saving the greater number (doing the greater good) does not calm our intuitive doubts. It is not easy to accept or condone the choices Sophie and Zofia made. These fundamentally unresolvable predicaments make these two films work outside the walls of the cinema and beyond the television screen. Can ethics help to resolve such cases?

## THE BASIC TROLLEY THOUGHT EXPERIMENT

The Trolley Thought Experiment presents a similar dilemma situation; it is framed, however, in a politically neutral context. Here is a common presentation of the case:

> You're standing by the side of a track when you see a runaway train hurtling toward you: the brakes have clearly failed. Ahead are five people tied to the track. If you do nothing, they will be run over and killed. Luckily, you are next to a signal switch: turning this switch will send the out-of-control train down a side-track that lies just ahead of you. Alas, there's a snag: you spot someone tied to this side-track, as well: changing the train's direction will inevitably result in this person being killed. What should you do?[5]

Why is this case similar to the previous two? There are lives at stake, and whichever way one decides, someone will be killed. It is an undesirable situation inducing contradictory intuitions. This is why many people would prefer not to discuss such cases, even if they are imaginary or hypothetical. They represent the type of dilemma that puts ethics on trial.

Although there are numerous variations, the dilemma has a standard structure: "The train is usually racing toward five unfortunates and the reader is presented with various means to rescue them, all of which at the cost of another life."[6] There are other, more or less characteristic features that trolley-cases share: the five people on the main track are generally just as innocent as the person on the side-track, and the only thing connecting them is the situation of the runaway train. The usual answer to the dilemma is changing the direction of the train and letting one unfortunate person be run over, on the assumption that "it is better to save five and let one die, than to let five die and one live".

This intuitive answer is often challenged by means of the following scenario:

> You're on a footbridge overlooking the railway track. You see the trolley hurtling along the track and, ahead of it, five people tied to the rails. Can these five be saved? Again, the moral philosopher has cunningly arranged matters so that they can be. There's a very fat man leaning over the railing watching the trolley. If you were to push him over the footbridge, he would tumble down and smash on the track below. He's so obese that his bulk would bring the trolley to a shuddering halt. Sadly, the process would kill the fat man. But it would save the other five. Should you push the fat man?[7]

Now, for most people the previous argument -"it is better to save five people and to let one die, than to let five die and allow one to live" - does not work for the fat-man case. It is not as easy as far as our conscience is concerned to push the fat man onto the tracks as it is to change the direction of the train, even if both actions are merely hypothetical. Moreover, it seems that many people are unable to explain their differing reactions to the two cases. This fact has fascinated not just philosophers, but also psychologists, economists, and representatives of various other disciplines. The interdisciplinary endeavor to solve the challenge posed by the thought experiment resulted in giving it its own name, trolleyology (coined by Kwame Anthony Appiah), which speaks for the career the Trolley Problem has made since its formation.

Although innumerable versions of the Trolley Dilemma exist, the Spur and the Fat Man[8] are the standard ones that professors of ethics use both in their introductory courses and in more advanced considerations of the topic. However, neither the Spur nor the Fat Man can be regarded as thought experiments in themselves, since some basic components of this form of thought are missing. To become proper thought experiments, both dilemmas require a context that raises moral issues, challenges moral beliefs, propositions and theories, and can activate the moral intuitions of the audience or reader through its collision with reality.

## FOOT'S ORIGINAL FORMULATION[9]

In his book, *Would You Kill the Fat Man?*, David Edmonds points out how important it is to research the biographies behind the formulation of the Trolley Problem. It was not created in a vacuum, but its emergence is closely bound up with the lives of

three philosophers and the age in which they lived.[10] Philippa Foot, who first formulated the Trolley Problem in her "The Problem of Abortion and the Doctrine of the Double Effect", and her two philosopher friends, Elizabeth Anscombe and Iris Murdoch, all experienced the dilemmas to which World War II gave rise, as well as the new life philosophy gained in the post-war period due to its special focus on ethical issues. It is no coincidence that the Trolley Problem first appeared in an article dealing with the theory of double effect in relation to abortion. It was a topic that Foot, a humanistic atheist, often debated with Anscombe, a devout Catholic.

The formulation of the Trolley Argument in Foot's now classical article is not very developed yet:

> …he is the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed.[11]

This original account of the case differs from from what have become the standard versions. Foot does not mention the word trolley – a key expression in the formulation of "trolleyology" –, but uses the term "tram" instead. A more important element is the fact that the audience is put in the position of the driver rather than that of an outsider who just happens to be standing next to a railway switch. If we read Foot's article attentively, it becomes apparent that she initially intended to use the example of an airplane and not a "runaway tram": "Beside this example is placed another in which a pilot whose airplane is about to crash is deciding whether to steer from a more to a less inhabited area."[12] This example, which was a daily reality in England during World War II, has been substituted with the

"runaway tram", a much better fit for times of peace. The pilot is put into the driver's seat of the tram to "make the parallel as close as possible".[13]

But the Runaway Tram is only one of the numerous examples that Foot uses to elaborate her views on the doctrine of double effect. She in fact uses them as tools to compare the reasons that lie behind human choice, and make fine distinctions between them. Another often quoted example is that of a judge facing angry rioters:

> Suppose that a judge or magistrate is faced with rioters demanding that a culprit be found for a certain crime and threatening otherwise to take their own bloody revenge on a particular section of the community. The real culprit being unknown, the judge sees himself as able to prevent the bloodshed only by framing some innocent person and having him executed.[14]

She later adds that "the mob have five hostages" to make the parallel between the two cases more obvious.[15] She supposes that anyone hearing the two cases would undoubtedly choose to steer the tram away from the five workers and towards the one but would not frame the innocent man. But what is the explanation for the different intuitions?

Firstly, Foot seems to find the answer in the doctrine of double effect, since it helps us to make a choice by differentiating between intending and foreseeing: namely, it is "one thing to steer towards someone foreseeing that you will kill him and another to aim at his death as part of your plan"[16] It is not part of the tram-driver's plan to kill the man on the side-track. If that man were able to free himself from the tracks before the tram arrived, the dilemma would be resolved. It would practically become not just the right, but also the obvious choice

to steer the tram to the side-track to save the five and kill no one. But the situation of the judge is far more difficult, since he needs to frame the innocent man and have him executed in order to prevent the riot and the bloodshed. The killing of the innocent victim is not just an accidental but an indispensable part of his plan. If he were somehow to escape, it would result in the unwanted outcome.

Foot does not stop here, however. She draws up another scenario, the amended version of which later came to be known as the "transplant case". She touches upon one of the trickiest and most delicate issues which modern public-health systems face: the allocation of resources. She asks her readers to imagine that they are doctors who want to save the life of a patient with the help of a certain drug. It becomes known, however, that the life of five other patients could be saved with the help of the very same drug. They only need one fifth of the standard dose each, but since so little of the drug exists, only one such dose is available. Foot claims that the obvious choice is to give the drug to the five, since "we feel bound to let one man die rather than many if that is our only choice."[17]

But does this intuitive choice justify all actions aiming at saving more lives rather than only one? Do we intuitively say yes to all actions which are done in order to save the most lives possible? Foot's answer to these questions is a resounding no, and she lists some exemplary cases where we would intuitively deny the justness of choices made according to this rule, such as "killing people in the interests of cancer research or to obtain (…) spare parts for grafting on to those who need them" or to "kill a certain individual and make a serum from his dead body" in order to save some individual.[18] She wrote these lines in the 1970s when hopes and concerns with regard to organ transplantation were at their peak; thus, they cannot be taken as simply hypothetical.

But what is the difference between sharing the drug among five patients at the cost of one patient dying, and killing an individual to use his or her organs to save five others? Or as Foot puts it: "Why cannot we argue from the case of the scarce drug to that of the body needed for medical purposes?"[19] The answer is the same as in the case of the runaway tram and the judge: when allocation is done at the cost of the death of an innocent person and we divide the drug up between the five instead of giving it to the one, we do not "aim at the death of an innocent man". When we harvest someone's organs at the cost of his death, however, we do. It is part of our plan to kill him, since this is the necessary price paid for obtaining that person's organs. Although Foot did not intend to provide bioethicists either with a puzzle or a line of argumentation, she nevertheless was able to formulate questions that have been subjects of bioethical debates for several decades now.

But how does Foot explain the intuition which supports changing the direction of the tram and giving the drug to the five, but rejects both the conviction of an innocent charged with murder and the killing of a man to harvest his organs? Foot explains the distinction not with our intuitive awareness of the doctrine of double effect, but with our intuition about positive and negative duties. She claims that our negative duties (to "refrain from injury") are stronger than our positive ones (to aid other human beings); thus, where a collision between the two sets of values occurs we have an intuition that negative duties should be preferred. Concerning the Tram Example, she concludes: "The steering driver faces a conflict of negative duties, since it is his duty to avoid injuring five men and also his duty to avoid injuring one. In the circumstances he is not able to avoid both, and it seems clear that he should do the least injury he can."[20] The same is the case with the allocation of the life-saving drug, although here it is pri-

marily positive duties that are involved. A collision of negative and positive duties, however, is at the heart of the decisions both of the judge and the transplant surgeon: the intuition to which the Trolley Dilemma precipitates can be explained by our preference for negative over positive duties. Foot sees our preference for negative duties as based on solid grounds and she herself is taken aback by her own observation that "even where the strictest duty to positive aid exists, this still does not weigh as if a negative duty were involved", for example, it is not permitted "to commit a murder" even if it was done in order "to bring one's starving children food".[21]

It is important to note that the "runaway tram" was only one of the numerous examples which Foot's article introduced to the field of practical ethics. This is underlined by the fact that Elizabeth Anscombe, in her commentary on Foot's article, does not touch upon the "runaway tram" example, but rather challenges Foot's hypothesis concerning the intuitive preference to give the scarce drug to the five instead of the one dying patient. Her objection is to utilitarian thinking or, more precisely, the universalization of the validity of utilitarian modes of intuition:

> There seems to me nothing wrong with giving the single patient the massive dose and letting the others die, of with refusing to deprive the single patient of care necessary to keep him alive because the hands needed for that care could help in saving the many victims of an accident. (…) I do not mean that "because they are more" isn't a good reason for helping these and not the one, or these rather than those. It is a perfectly intelligible reason. But it doesn't follow from that that a man acts badly if he doesn't make it his reason.[22]

Anscombe not only identifies weak points in Foot's arguments, but the difficulties that surface when ethical theories are based principally upon examples, and rely too heavily upon intuitions.

## THE ORIGINAL TROLLEY EXAMPLE AS A THOUGHT EXPERIMENT

Anscombe criticizes Foot for exaggerating certain arguments in her article. Firstly, the utilitarian argument is a sound reason, but not an exclusive one when arguing that the scarce drug should be given to the five: "It is a perfectly intelligible reason. But it doesn't follow from that that a man acts badly if he doesn't make it his reason."[23] Secondly, giving the drug to the five might be supported by intuition, but there may arise other considerations, as well, such as taking into consideration the financial situation of these patients. Thirdly, not everyone has the intuition that we should always act with the intention to save the most lives – for instance, it is not shared by Elizabeth Anscombe.

So what's wrong with the opposing arguments in which the original Trolley Example was set? First of all it must be taken into account that the "runaway tram" is not a thought experiment in itself. Foot combines it with other examples to make it one. Thus, the reader is confronted with several fictive scenarios, each with particular morally relevant aspects. In examples of the "runaway tram" and the "judge", one is called upon to decide whether one individual or five people will die. The difference between the intuitions triggered in the two cases is used firstly to examine the validity the doctrine of double effect, and secondly to propose a new theory which might explain the different responses. What this technique fails to

achieve, however, is a conviction on the part of the reader that a particular course of action is ethically better or right. Intuitions are simply triggered without any intention to challenge their ethical rightness. Contradictory intuitions are not challenged, and the difference is simply explained in descriptive terms.

Foot makes use of intuitions in an affirmative manner. She does not question them but uses them to support her own theory. The readers of the Experience Machine Thought Experiment could discover the insufficiency of their previously held hedonistic principle. But what can readers of the "runaway tram" discover? Does a differentiation between positive and negative duties explain, or simply describe certain features of our intuitions? And, finally, why should we take the priority of negative duties over positive ones as normative? Foot does not answer these questions and, as Anscombe's criticisms show, there are good reasons to challenge her arguments.

## VARIATIONS OF THE TROLLEY PROBLEM

The original version of the Trolley Example has undergone countless changes over the last fifty years. Debates have evolved around the issue specifically raised by Foot's article, and the Trolley Example itself has developed into an autonomous domain of academic discourse. The original version was altered for several different reasons: subsequent alternative descriptions aimed at a better description of the intuitive structure activated by the scenarios, and also to verify various alternative theories. Although it is mostly philosophers and ethicists who have dealt with the topic, the Trolley Problem has run a brilliant career within the fields of economics, psychology, neuroscience, and law. I will now take examples from

the abundant literature on the Trolley Problem to illustrate the purpose of this type of thought experiment, which is to highlight the conflict between the intuitively discerned good and the actual state of affairs.

The elements that characteristically constitute a thought experiment are also present in every trolley example. These are imaginary scenarios designed to function according to well-defined rules. Their discrete elements are not arbitrary, and the scenario may not be amended beyond the given rules: for example, it is not enough to take the backpack of the man standing on the footbridge and throw it before the trolley. They all describe an imaginary scenario that implies a certain assumption about the value of every human life, bodily integrity, and moral responsibility. These moral considerations hold independent of the individual person. If someone claims that the value of the life of an overweight person is less than that of someone who is not, then pushing the fat man from the footbridge might well be taken to be an acceptable solution. But as long as the audience assigns the same value to every human life and recognizes the equal dignity of every human person, they will find themselves in a dilemma situation each time they have to assess the worth of individual human lives or to decide who should live and who should die.

However, the Trolley Example differs from other thought experiments with respect to their contextual embedding and their relation to the real world. Most trolley cases could happen at any time; we do not require great imaginative capacities to acknowledge that trolley-like situations exist in real life. In such cases, the only option left is a choice between two evils of equal weight, and it is impossible not to be guilty in some way or to some degree. Due to the general presence of railway vehicles, almost everyone living in our modern age understands the danger presented by a runaway trolley. However,

it is the growth of trolleyology that gives rise to further difficulties: imaginary scenarios have been formulated within different contexts, thereby providing a large number of possibilities for analogy. Thus, the trolley problem cannot be treated as an individual thought experiment, but as a whole range of imaginary scenarios designed to achieve different aims and formulated in different contexts. Still, as the history of the trolley problem shows, a well-designed thought experiment can maintain its intuitive force for decades and may not only serve as evidence in support of certain ideas but also facilitate the development of new ethical theories.

Trolley scenarios, however, have another common element. They assert the value of human life and the obligation to protect it. Although these concepts might seem trivial, they are in no way self-evident. One could imagine, for example, a madman who enjoys running people over with a trolley. One of the presuppositions of any trolley thought experiment, therefore, must be a shared acceptance of the fundamental value of human life and the obligation to protect it. Readers must have these ideas in common, since without them every trolley example is meaningless.

## THOMSON AND THE RISE OF TROLLEYOLOGY

Although the founder of trolleyology was undoubtedly Philippa Foot, it was Judith Jarvis Thomson who contributed most to its rise. She turned a simple example into a complex and exceptionally challenging line of subsequent thought experiments. In her two pioneering articles, Thomson expanded Foot's original idea and created further parallel examples.[24] While the trolley example was just one among several other imaginary scenarios for Foot, Thomson places the trolley kind of case at

the center of ethical inquiry. Indeed, she was the first author to use the term "trolley problem".

Her contribution is also interesting in that she explicitly rejects the doctrine of double effect, which Foot thought worthy of consideration, and puts the question of rights at the center of her argument, thus omitting any focus on the intention of the acting person.[25] She connects the trolley problem with the question of death and the related issue of euthanasia, a central topic in the field of bioethics. In her 1976 article, "Killing, Letting Die, and the Trolley Problem," she lays the ground for her argument as follows:

> Morally speaking it may matter a great deal how death comes about, whether from natural causes, or at the hands of another, for example. Does it matter whether a man was killed or only let die? A great many people think it does: they think that killing is worse than letting die. And they draw conclusions from this for abortion, euthanasia, and the distribution of scarce medical resources. Others think it doesn't, and they think this shown by what we see when we construct a pair of cases which are so far as possible in all other respects alike, except that in the one case the agent kills, in the other he only lets die.[26]

In these sentences, Thomson describes the horizon of the readers and highlights the presuppositions that play a vital role in how thought experiments are understood. She differentiates between two distinct groups: those who view the distinction between killing and letting die as crucial, and those who think it to be ethically irrelevant. She demonstrates her thesis by formulating the first versions of her trolley and transplant-surgeon dilemmas, and tries to show that the distinction between

killing and letting die does not completely accord with our intuitions. She describes the examples she gives of the transplant-surgeon dilemma as follows, and expects a clear and unambiguous intuitive response, at least to the first one:

> Charles is a great transplant surgeon. One of his patients needs a new heart, but is of a relatively rare blood-type. By chance, Charles learns of a healthy specimen with that very blood-type. Charles can take the healthy specimen's heart, killing him, and install it in his patient, saving him. Or he can refrain from taking the healthy specimen's heart, letting his patient die.[27]

She also thinks that most of her readers will reject the first possibility given in her second example, as well:

> David is a great transplant surgeon. Five of his patients need new parts – one needs a heart, the others need, respectively, liver, stomach, spleen, and spinal cord – but all are of the same, relatively rare, blood-type. By chance, David learns of a healthy specimen with that very blood-type. David can take the healthy specimen's parts, killing him, and install them in his patients, saving them. Or he can refrain from taking the healthy specimen's parts, letting his patients die.[28]

The transplant-surgeon dilemma clearly presents the stringent rules governing every thought experiment in ethics. There are only two options: to kill the healthy person, remove his organs, transplant them, and save the five patients, or not to kill the healthy person and remove his or her organs, and to let the patients die. However, such situations – if they were to ever

actually happen –, would be much more complex. The complexity arises not simply because it is impossible to transplant a spinal cord, but also because important details are omitted from the examples. We don't know, for example, if the recipients of the organs want their lives to be prolonged by means of organ transplantation. We only know that in the hypothetical world of the thought experiment the possible outcomes are guaranteed.

The same is true for the trolley dilemma Thomson designs to contrast with the two surgeon examples:

> Edward is the driver of a trolley, whose brakes have just failed. On the track ahead of him are five people; the banks are so steep that they will not be able to get off the track in time. The track has a spur leading off to the right, and Edward can turn the trolley onto it. Unfortunately there is one person on the right-hand track. Edward can turn the trolley, killing the one; or he can refrain from turning the trolley, killing five.[29]

Thomson uses an example of the trolley dilemma to induce the intuitive judgment approving the driver's decision to change the direction of the trolley and the view that a distinction can be made between killing and letting die. If killing is ethically out of bounds, how is it possible to intuitively approve the driver's decision to change the course of the trolley? Thomson refers here to Foot's distinction, which argues that negative duties are more important than positive ones. Here two negative duties collide, which explains why we approve of the driver's decision to turn the trolley. However, Thomson criticizes Foot when she writes: "Now I am inclined to think that Mrs. Foot is mistaken about why Edward may turn his trolley, but David may not dissect his healthy specimen."[30] To support this

criticism, she constructs another example, in which "Frank is a passenger on a trolley whose driver has just shouted that the trolley's brakes have failed, and who then died of the shock."[31] She makes an obvious distinction here based upon the role the person fills. The driver of the trolley kills whomever he chooses, since he is in control of the steering wheel; in Thomson's version, however, Frank is only a passive observer of the trolley's path. According to Thomson "it seems (…) that if Frank does nothing, he kills no one. He at worst lets the trolley kill the five; he does not himself kill them, but only lets them die."[32]

These two examples– the one where reades must place themselves in the situation of the driver of a trolley, and the other where they are assigned the role of a passenger – clearly demonstrate how Thomson makes use of thought experiments: First, she takes Foot's theory, according to which negative obligations (e.g. not to cause harm to others) outweigh positive ones (e.g. help others) in cases where they come into conflict. In a second step she shows how this formula presents itself in the original thought experiment (i.e. Foot's tram example) and connects it to the intuition that the case induces. Third, she constructs another thought experiment, which induces the same intuition, but to which the previous formula (i.e. the primacy of negative duties over positive ones) does not apply. The intuition thus produces the same judgment in both cases and suggests altering the direction of the trolley. However, these intuitive judgments cannot be justified with the same resolution: namely, the one that Foot offered in her original verions. Thought experiments are therefore inadequate when used as a means to resolve dilemmas in ethics, because they only describe intuitive judgments without offering a normative justification of one's actions in a specific situation.

Thomson reformulates Foot's version of the fat-man scenario as follows:

> George is on a footbridge over the trolley tracks. He knows trolleys, and can see that the one approaching the bridge is out of control. On the track back of the bridge there are five people; the banks are so steep that they will not be able to get off the track in time. George knows that the only way to stop an out-of-control trolley is to drop a very heavy weight into its path. But the only available, sufficiently heavy weight is a fat man, also watching the trolley from the footbridge. George can shove the fat man onto the track in the path of the trolley, killing the fat man; or he can refrain from doing this, letting the five die.[33]

Thomson assumes that the scenario provokes an intuitive judgment different from the previous ones. The fat man ought not to be pushed, and the trolley should be allowed to run over the five men on the main track. As was indicated earlier, the fat man scenario constitutes the real challenge with regard to finding an answer to the trolley dilemma. Why does our intuition allow us to redirect the trolley to another track and so kill one person, but not allow us to push the fat man from the bridge? Thomson explains that redirecting the trolley from one track onto another represents only a simple "distribution" of the danger embodied by the speeding trolley: "If the one has no more claim against the bad thing than any of the five has, he cannot complain if we do something to him in order to bring about that the bad thing is better distributed".[34] What Thomson does here is to apply the Kantian prohibition against using another person as a mere means rather than an end, and supplementing it with the principle of the distribution of dangers. It is worth noticing, however, that she does not refer here to Kant, but rather to the intuition which tells us that "what matters in these cases in which a threat is to be distributed is

whether the agent distributes it by doing something to it, or whether he distributes it by doing something to a person".[35] Thomson formulates the principle of "distributive exception", defining it as follows: "It is not morally required of us that we let a burden descend out of the blue onto five when we can make it instead descend onto one only if we can make it instead descend onto the one by means which do not themselves constitute infringements of rights of the one?"[36]

The starting point here, just as in the case of the violinist, is the question of rights. Thomson claims that in the Spur Dilemma the rights of the individual person are not infringed upon, while the basic rights of the fat man are. But which rights are involved in these two cases, and how can we differentiate clearly between the right not to be pushed to your death from the right not have a death-dealing trolley directed at you. Gorr sees the main point of Thomson's theory in the following distinction: "There is an intrinsic moral difference between (i) bringing it about that the smaller group is threatened by doing something to that group and (ii) bringing this about by doing something to the threatening force".[37] In short, there is a difference between redirecting a threatening force towards an innocent man and redirecting this man towards the threatening force in order to remove or disable it. Although clothed in different words, the Kantian categorical imperative can be identified here: it is immoral to use a person merely as a means and not as an end.[38] This applies both to the trolley scenarios and the surgeon case.

There is at least one significant difference, however: Kant uses the categorical imperative as a normative call, while Thomson's distinctions are of a descriptive nature. They serve to provide a theoretical formulation of the nature of intuitive judgments. This theory condones the turning of the switch but does not allow the pushing of the fat man. They fail, nevertheless, to explain why one action is morally right, while the other is morally wrong.

Thomson seems to acknowledge this deficiency, as she stresses the importance of considering individual cases and not using her distinctions too mechanically. Concerning the one between killing someone and letting someone die, she writes:

> the thesis that killing is worse than letting die cannot be used in any simple, mechanical way in order to yield conclusions about abortion, euthanasia, and the distribution of scarce medical resources. The cases have to be looked at individually. If nothing else comes out of the preceding discussion, it may anyway serve as a reminder of this: that there are circumstances in which – even if it is true that killing is worse than letting die – one may choose to kill instead of letting die.[39]

## SEEING THE FAT MAN'S FACE

The previous examples are faceless scenarios. We are not given any information about the personality or circumstances of the characters; all we know is that, in the one case, they are track-workers and, in the other that the man on the footbridge is fat. The second piece of information, even if it might appear personal, is only a technical element in the thought experiment. His weight only matters insofar as it enables us to imagine that his body would certainly stop the train. He could just as well be a person with a backpack or some other heavy object tied to his body. But would anything change if we could see the faces of these people, or if we knew something important about them? Thomson tries to expand her original trolley scenario by adding some more specific details about the characters involved:

> The five on the track ahead are regular track workmen, repairing the track – they have been warned of the dangers of their job, and are paid specially high salaries to compensate. The right-hand track is a dead end, unused in ten years. The Mayor, representing the City, has set out picnic tables on it, and invited the convalescents at the nearby City Hospital to have lunch there, guaranteeing them safety from trolleys. The one on the right-hand track is a convalescent having his lunch there; it would never have occurred to him to have his lunch there but for the Mayor's invitation and guarantee of safety. And Edward (Frank) is the Mayor.[40]

What has changed and why do most people intuitively feel that the switch should not be redirected in this altered version? Interestingly, the danger does not seem to be as threatening if the track workers know that this danger could arise sometime during their work, especially if it is part of their contract, and they receive due financial compensation.

But what is the basis of the intuition according to which we ought not to turn the switch? In this case, it is not only the value of life human or life, but also the principle of justice. While the convalescent has a right to special protection, this is not true for the track workers, since they are aware of the possible danger of their work and have accepted it by signing a contract. Interestingly, it is not the case itself that produces the intuitive tension but the incongruence between intuitive ethical judgments as they pertain to similar cases: Why do we respond intuitively with a "yes" to the dilemma of the spur and with a "no" to the one with the mayor?

Other factors which might be added to the original Spur Scenario that might alter considerably our original moral sense.

A cartoon on the internet illustrates a situation where, behind the veil of ignorance, we arrive at a decision about whether or not to redirect the trolley. Its message is as follows: "You don't know where you'll be in the trolley problem. However, you have to choose the scenario in advance. Regarding personal interest, would you like the lever to be pulled?"[41]

What are the principles which we were willing to accept as guides when making the choice? Presumably this choice is not as easy as it would be in the future just society envisioned by Rawls, since it is not only a matter of social justice, but of life and death: that is, participating in either the killing or the saving of lives. Behind the veil of ignorance it would be transformed into a perfect dilemma – similar to the case of Tomoceuszkakatiti and Gyugyu. There is only one way out, namely, the balancing of different risk factors. However, this cannot be considered as something general, since there are certainly some who are not willing to take certain risks in such a situation – e.g. taking the role of the driver, or the person at the switch. It all begins to look like the game of the devil's bones.

## THE RECONSTRUCTED TRACK SYSTEM NETWORK

The Trolley Scenario may not only be amended by providing information on the characters, but also the track system may undergo certain changes. Frances Kamm also introduced one of the most disquieting questions of trolleyology by redesigning the track system network. The loop scenario goes as follows:

> The trolley is heading toward five men, who, as it happens, are all skinny. If the trolley were to collide into them they would die, but their combined bulk would stop the train. You could instead turn the trol-

> ley onto a loop. One fat man is tied onto the loop. His weight alone will stop the trolley, preventing it from continuing around the loop and killing the five. Should you turn the trolley down the loop?[42]

If the graphic representations of spur and loop are juxtaposed, it is hard to recognize the difference without a closer look. It is only a short pair of rails connecting the two ends of the spur. However, this short section of rail causes enormous difficulties with regard to the ethical evaluation of the scenario. This is because of the incongruence between the intuitions induced by the fat man and the loop. Someone who argued that the fat man cannot be used as a means to stop the trolley cannot redirect the trolley in the loop case either, not even if he argued in favour of turning the switch in the spur scenario.

But what is the difference between the spur and the the loop? It is the role of the individual in saving the five workers. In the case of the spur, the presence of someone on the sidetrack has no causal relationship with the saving of the five men. In the case of the loop, the five can only survive if the fat man is hit by the trolley. His death is not only the price but also the precondition for their survival. If we apply the doctrine of double effect to the dilemma in its simplified form, those two factors are of great importance. While in the case of spur the death of the single worker is not an intended, but merely a foreseen effect of pulling the switch, in the case of loop, the deadly collision of the trolley with the fat man is intended. If it did not collide with the fat man, the trolley would simply carry on down the track towards the five workers.

This solution might appear unsatisfactory for many of us, since the difference between the two scenarios is only a couple of meters of track. Frances Kamm is also one of the critics who has tried to explain the moral difference between the

two situations by developing the docrtine of triple effect. She demonstrates the main idea behind the doctrine through the following example:

> I intend to give a party in order for me and my friends to have fun. However, I foresee that this will leave a big mess, and I do not want to have a party if I will be left to clean it up. I also foresee a further effect of the party: If my friends have fun, they will feel indebted to me and help me clean up. I assume that a feeling of indebtedness is something of a negative for a person to have. I give the party because I believe that my friends will feel indebted and (so) because I will not have a mess to clean up. These expectations are conditions of my action. I would not act unless I had them. The fact that they will feel indebted is a reason for my acting. But I do not give the party even in part in order to make my friends feel indebted nor in order to not have a mess. To be more precise, it is not a goal of my action of giving the party to do either of these things. I may have it as a background goal of my life not to have messes, but not producing a mess is not an aim of my giving the party.[43]

Kamm makes a distinction between "acting because I believe I will have a certain effect" and "acting in order to bring about (intending) the effect".[44] She holds this distinction to be relevant in the loop case as a way of replacing the doctrine of double effect:

> I claim that doing something *because this will cause* the hitting of an innocent bystander does not imply that one *intends to cause* the hitting or that one does

> anything in order to hit. This is because there is a general conceptual distinction between doing something because it will have an effect and doing it in order to produce an effect.[45]

But what happens when the distinction is applied to the two alternative versions? In the case of the spur, the distinction is not relevant, since there is no intent to have the trolley hit the person bound to the sidetrack, and we do not pull the switch because this will cause the death of that person. The moral decision can be satisfactorily described with the conceptual distinctions that the doctrine of double effect already provides. It is, however, different from the fat man case, where there is a clear intention to push the fat man in front of the trolley. Here, we intend to cause the fat man to be struck by the trolley. This distinction is relevant only in the loop case, since the turning of the switch is carried out because this will cause the fat man to be hit by the trolley.

Kamm summarizes the Doctrine of Triple Effect as follows: "A greater good that we cause and whose expected existence is a condition of our action, but which we do not necessarily intend, may justify a lesser evil that we must not intend but may have as a condition of action."[46] It is, however, questionable whether it has helped Kamm resolve the trolley dilemma . On the one hand, the doctrine of triple effect does not seem to be much more than a complex formulation of the lesser of two evils principle. The latter entails that the dilemma cannot be solved without opting for an evil alternative, and that the choice of the lesser evil is a precondition for avoiding the greater evil. From the perspective of simplicity, the doctrine is also uneconomical.

But there is a more critical point. Michael Otsuka points out that the root of the problem is that the loop case is compared

with those of the spur and the fat man. When talking about a couple of extra meters, we tend to focus less on structural differences than on visual similarities. According to Otsuka "it is harder to morally differentiate looping cases from the Bridge Case than it is to morally differentiate them from the Trolley Case".[47] He claims that point is not the difference between "knowing it will cause" and "intending to cause", but that "the apparently morally significant distinction between treating as a means and not so treating appears to distinguish looping cases from the Trolley Case".[48]

## THE UTILITARIAN RESPONSE

The simplest solution to the Trolley Problem seems to be the application of the utilitarian calculus to individual cases. Ethical dilemmas may seemingly be resolved by thinking about them in terms of a mathematical equation. One ought to choose the option promising the least number of deaths in the end. In the Spur, the Fat Man, and the Loop forms of the trolley dilemma, the trolley ought not to be allowed to run over the five workers. The right action is, in one case, to turn the switch to divert the trolley onto the track with the one single worker and, in the other, to push the fat man from the footbridge in order to stop the trolley. According to utilitarians, these options should be evaluated independent of our intuitive judgments concerning the individual cases.

The weaknesses of the utilitarian calculus present themselves here, too. On the one hand, it is hard to calculate which outcome results in the most pleasure, and prevents the most suffering. What if the five people on the track organize illegal animal fights and cause great suffering to dozens of dogs every day, while on the side-track there lies an enthusiastic friend

of animals who voluntarily maintains an animal shelter for hundreds of dogs, providing them with food and care. In this case – according to a non-anthropocentric utilitarian calculus –, it is better to let the trolley hold its course.

Peter Unger defines moral common sense as follows: "What's morally more weighty is how much you (knowingly) lessened, and how much you (knowingly) increased, the serious losses suffered."[49] He develops the trolley problem from a two-option case to a several-option one in his book *Living High and Letting Die*. In the book's imaginary scenarios "an agent has more than two options, and (…) she must have at least two active options".[50] His aim with using multiple-option cases is to question the validity of earlier intuitive judgments.

In the multiple-option case, called the Switches and Skates, he constructs a complex system network of intercepting tracks, empty and overloaded trolleys, main and sidetracks, with switches and fat men riding remote controlled skates.[51] He offers four different options according to which one may respond to the situation. In the first version, "you do nothing about the situation (…) then, in a couple of minutes, it will run over and kill six innocents who, through no fault of their own, are trapped down the line".[52] The second option is to "push a remote control button" in order to change "the position of a switch track" and lead the trolley away from the six to a line where "three other innocents are trapped".[53] If the third option is chosen, the empty trolley endangering six people may be stopped with the help of another trolley carrying two people, who will lose their lives due owing to the collision. Finally the trolley endangering the lives of six innocent people can be brought to a halt by turning on a "remote control dial", thus starting "up the skates" and thus sending the heavy man "in front of the trolley".[54]

By choosing the first option you let six people die; opting for the second allows you to save six but kill three; with the

third option, you save six and kill two; and with the fourth, six people can be saved by killing one. Utilitarians would probably choose number four, while deontologists would stick with the first option. However, according to Unger, most people's intuitive judgment would be that starting up the skates is the appropriate response in this case.

But Unger designs another imaginary case (The Heavy Skater) to show how much "folks want their responses to seem consistent".[55] In this imaginary scenario the case is presented much more simply by making only the first and the fourth options available. By putting the two essentially identical thought experiments side by side, he demonstrates the extent to which previous trolley examples influence our intuitive judgments. He claims that "had readers confronted the Heavy Skater first, there'd be a strong tendency (…) to respond negatively" both to the Heavy Skater and option four of the Switches and Skates. This shows clearly that our responses to imaginary scenarios are highly guided by our previous decisions and our pursuit of coherence.

The incoherence of intuitive judgments is demonstrated by another pair of thought experiments. The protagonist of the first one is Bob who, in order to secure a peaceful and financially secure retirement, buys a Bugatti. He keeps it in a garage, and one day he decides to take a ride and ends up in a trolley-case-like situation. He is at a switch, and needs to decide whether he should let the trolley run over a young child or let his Bugatti be smashed to pieces by the runaway vehicle: "Bob chooses the first option and, even though the child is killed, he has a comfortable retirement."[56]

Unger contrasts this example with the case of an accountant, Raymond R. Raymond, who is asked to give "99% of his material assets, including both what's in his retirement fund and what's not, to Unicef".[57] Raymond is informed that this money

would help a large number of needy children all around the world to survive. "Understandably, Ray does nothing toward meeting his Big Request and, so, thousands more children die than if he'd met it."[58]

Unger admits that the analogy between the two examples is not completely satisfactory, since the amount of the assets and their significance for their owner and the number of lives saved differ significantly. He also realizes that, in a psychological sense, it is easier to help a child who is physically near us than anonymous children whose lives are saved only through the intervention of a faceless charity organization. Still, Unger succeeds in pointing out the immorality of Western people's reluctance to donate money for humanitarian purposes. He criticizes intuitive judgments on the grounds that they are driven by the "conspicuousness of the need", namely, "the extent to which the need attracts and holds your attention".[59] A suffering child next to us can certainly "hold our attention" more effectively than a short written report about thousands of children dying in a faraway land. Besides the factor of conspicuousness, it is "futility thinking" which characterizes the intuitive judgments made in these cases, which "focus[es] on the vastness of the serious losses that will still be suffered even if you do all that you can do".[60] Saving a child on the spot makes immediate sense; however, sending money to underdeveloped countries seems futile. Singer summarizes his argument as follows:

> As with the conspicuousness of need, futility thinking seems to be anything but a sound basis for intuitive moral judgments. Whether our response to an example is affected by futility thinking will depend on whether we identify the person we help as an individual, or as one of a much larger group. (…) It is hard to see why

> this factor should carry much weight, as compared with questions about how much good we can do, and at what cost to ourselves. When we save the lives of ten or a hundred children, the good that we do to those children is not diminished by the fact that other children are still dying.[61]

From this summary it is clear that analogous thought experiments inducing diverse intuitions basically serve to confirm utilitarian thinking.

There are some questions concerning this reasoning, however. No matter how clear the underlying principles which govern our intuitive apparatus – as in the latter case conspicuousness and futility thinking – they do not go beyond the level of description and fail to provide sufficient justification for the soundness of the utilitarian argument.

## THE TIME TRAVELER

Most trolley examples are lineal. The events, even if they are distant from one another in space, are mostly close in time. The runaway trolley makes impact within a short time-interval. But how do our intuitive responses change if the impact of the trolley is not immediate? This question was raised by my students on the basis of some recent sci-fi television series. These programmes are based on the futuristic ability of human beings to travel in time. The dramatic element consists in the interconnectedness between a past and present state of affairs; the protagonists are thus able to alter the present by changing the past. The thought experiment proposed by my students goes as follows:

> There is a small town where a serial killer has murdered five women. His identity is unknown, but somehow we receive information about who his father was. In present time, the serial killer continues to commit other murders and the police do not seem to be able to put an end to his evil deeds. Luckily we have a time machine that enables us to return to the past and prevent the serial killer from being born by killing his father. There is no other option left. We also know that his father died shortly after the birth of the serial killer, thus he is in no way responsible for how his son has turned out. The question is whether it can be ethically justified to go back in time and kill the murderer's father, thereby prevent the murders from occurring.

Although the thought experiment plays out according to different sequences in time, it presents a similar structure to standard trolley cases. Five people are in grave danger, and their lives can only be saved by sacrificing one other person. The question is whether one life is expendable in order to save five. We can raise the same question in a more trolleyesque form:

> Imagine that a runaway trolley has caused a serious accident in a small town. A trolley designed for sightseeing trips has run over five people. We also know that the trolley has been especially designed and manufactured to carry tourists around the town. The accident has happened owing to the inattentiveness of the driver. The trolley had hitherto functioned flawlessly. We now have a time machine capable of taking us back to a time prior to its construction, and thereby granting us a chance to prevent the accident. The

> only way to stop the trolley-builder from designing the runaway trolley is to push him in front of another trolley. The question is whether it is ethically justified to go back in time and push the trolley-builder in front of another trolley.

One would imagine that most people would answer these questions with a "no". The probable reason is that neither the father of the serial killer nor the trolley builder can be blamed for what has happened. They did not intend the death of the five victims and were not responsible for the events in any other way. They are separated from the accident by a great deal of time. Certainly many people think about the possibility of going back in time in order to kill Hitler or Stalin. Many of them would even support such an idea, if time travel were possible. Far fewer, however, have brooded upon whether it would be ethically justified to kill Alois Hitler and Klara Pölzl, or Besarion Jughashvili and Ketevan Geladze, to forestall the evil deeds committed by their offspring.

The time factor plays a significant role in the trolley problem if there is an immediate connection between the trolley as a threat and the person whose death would result from saving the lives of the five people. The intuitive responses to the three basic trolley scenarios would be significantly altered if it turned out that it was the railway worker or the fat man who tied the five others to the track. In this case, most readers would decide to change the direction of the switch, and also the number of those who would consider pushing the fat man from the bridge to be an ethically justified action would increase. Moreover, there are presumably numerous people who would also find it justified to return to the past in order to push in front of a trolley the trolley-builder who intentionally designed a trolley to run over innocent people. Similar

intuitions may also be called forth if the accident was caused by negligence. Thus, it is of great importance concerning intuitive judgments whether or not somebody can be blamed for the emergence of the threat. There is a significant difference between our intuitive judgment concerning those who are responsible for the current state of affairs and those who are not. The utilitarian calculus would condone the murder of the parents of future dictators, but this goes against intuitive judgements we would make in parallel cases. Our intuitions would only change if the parents turned out to be active participants in the temporal sequence of events that made Hitler and Stalin vicious dictators.

## THE LIMITATIONS OF TROLLEY EXAMPLES

Students often say that trolleyology is just a pastime for bored philosophers. Its purpose is merely to provoke interesting discussions with fellow philosophers, or to serve as a tool to make classes more interesting. However, it is not just philosophers who love being challenged by dilemmas like the Trolley Example. Ferdinand von Schirach's *Terror* was one of the best attended and most discussed theatrical pieces in 2016.[62] It was even adapted for cinema and television. Schirach's play modifies traditional theater by making the audience part of the performance. The theater is turned into a law court where a pilot is put on trial.

The story behind the trial is the hijacking of an airliner flying from Berlin to Munich by a terrorist. He redirects the plane against a packed stadium. Combat planes are sent up to follow the airliner, one of the pilots being Lars Koch, the central figure in the trial. He has to decide (1) whether to shoot down the airliner, thereby causing the deaths of the 164 passengers but

saving the 60,000 spectators who are inside the stadium, or (2) to let the plane fly on, with a very strong likelihood that everyone on the plane and in the stadium will die.

At the end of the play, the audience is asked to cast their votes, taking on the role of the grand jury in the case, tasked with deciding whether the pilot is innocent or guilty. The play and the results of the jury vote can be viewed on the internet. Since the play has not only been performed in Germany, but also in other parts of Europe, as well as in more distant countries such as Japan, it is easy to compare how the voting in different nations and cultures went. The most striking difference can be observed between Germany and Japan. While most audiences in Germany find the pilot not guilty, Japanese audiences tend to think that he is culpable. The most obvious explanation is that while the Germans have a high respect for an individual's conscience, the Japanese prioritize the respect for authority. Either way, the success of the play showed not only that we have different, often culturally dependent intuitions with regard to dilemma cases, but also that we are stimulated when our intuitions are challenged by such differences – at least, in a fictive setting.

Why is this so? Does a sor of catharsis occur when we let our intuitive responses be tested over the two hours or so that we spend watching a play? Do we know ourselves better after attending the fictive juridical process, and being asked to cast our vote? Might philosophers be preoccupied with dilemma-cases in order to experience a catharsis and thereby come to know themselves and the people around them better? And (even) if these questions are answered in the affirmative are dilemmas and the intuitions they elicit deprived of a role in the formulation of ethical theories?

## THE NATURE OF INTUITIVE JUDGMENTS

As we have seen, the procedure for working with trolley examples is the following: "The method of trolleyology involves conjuring up various trolleyesque scenarios and taking note of the (preferably) strong moral intuitions that they elicit. Then he or she tries to formulate a plausible principle (or principles) that unites and makes sense of these intuitions."[63] But this method raises two crucial questions. The first concerns the nature of the principle formulated on the basis of these intuitions. The second is whether the principle formulated to account for these intuitions should be considered at all as constitutive of a normative ethical theory.

It must be asked whether intuitions have any part at all to play in ethics. Problems related to their application are revealed in the way trolleyologists use them: by eliciting intuitions they work towards a principle which "should itself have some intuitive plausibility".[64] Thus, we offend up with a vicious cycle in which intuitions serve not just as the starting point of the argument but also as the entities which justify the conclusion.

Still, intuitions - judgments that arise spontaneously and prior to any rational thought - are crucial to ethics. Intuitive judgments resemble judgments of conscience by virtue of their unconditional nature. One cannot intentionally produce a particular intuition, nor can we make our conscience arrive at a certain judgment. However, intuition and conscience also differ in important respects. Intuitive judgments are facts born in connection with a certain event (e.g. someone sees a young man robbing an old lady in the street and intuitively, without any further considerations, senses it is not right.) The judgment of conscience needs to be taken into account and cannot be set apart from the judgment of reason. The person

listening to his intuition does nothing but refer to the spontaneous judgmentthat has come to him prior to all rational reflection, and which might prove to be irrational. But when someone speaks about a judgment of conscience, he means a judgment at which he has arrived after rational and responsible reflection. He has reached the conclusion that his judgment concerning a particular event is right and can explain why he considers that particular conclusion to be right over another. Others have the opportunity to challenge his insight, or consider it irrational. This does not change the fact that the person's judgment was preceeded by rational consideration.

Judgments of conscience are less problematic in ethics than moral intuitions, and this gives rise to difficulties concerning thought experiments, since intuitions are integral to this mode of thinking. Still, thought experiments might help individuals to progress from intuitive judgments to judgments of conscience. An obvious example is the Parable of the Good Samaritan, which we discussed earlier. The intuitive judgment about the Samaritan's actions, namely that they were morally right, and that he showed himself to be a true neighbor to the victim, turns into a judgment of conscience as soon as the Jewish audience realizes how narrowly they previously defined the term neighbor. Intuition here was only a tool which helped them to arrive at the judgment of conscience so fundamental to ethics.

## SINGER ON INTUITION

But what can we do with moral intuitions that do not develop into judgments of conscience? An answer to this question comes from Peter Singer who clearly opposes an uncritical appeal to intuition in ethical theories. In his article "Ethics and Intuitions" he provides an overview of the theoretical status

of intuitions, from a perspective that is especially critical of trolleyology. He sees the central problem as follows:

> These philosophers thus take the moral intuitions elicited by the cases as correct, and seek to justify them. But every time a seemingly plausible justifying principle has been suggested, other philosophers have produced variants on the original pair of cases that show that the suggested principle does not succeed in justifying our intuitive responses.[65]

He shares the concerns of James Rachel regarding the use of intuitions in ethics. Intuitions should not be used as point of orientation, since we should not build ethical theories upon them. He instead suggests that we take another path, namely "to challenge the intuitions that first come to mind when we are asked about a moral issue".[66] If this recommendation were implemented, then trolleyologists would begin by ethically criticizing the intuitions induced by their fictive constructs.

There is a specific reason why Singer wants to weaken the position of intuitions in ethical theory. His aim is to disqualify one of the major arguments against utilitarianism. It is quite easy to find hypothetical examples in bioethical literature designed to show that utilitarian thinking may lead to conclusions which are incompatible with our intuitions. He describes the steps involved in such undertakings as follows:

> Initially, the use of such examples to appeal to our common moral intuitions against consequentialist theories was an ad hoc device lacking metaethical foundations. It was simply a way of saying: 'If Theory U is true, then in situation X you should do Y. But we know that it would be wrong to do Y in X, therefore U

> cannot be true.' This is an effective argument against U, as long as the judgmentthat it would be wrong to do Y in X is not challenged. But the argument does nothing to establish that it is wrong to do Y in X, nor what a sounder theory than U would be like.[67]

In his article, Singer uses several arguments from philosophy, evolutional psychology, and neuroscience to support his position. First of all, he highlights the evolutionary basis of morality. Ethical concepts, like justice, can be observed not just among humans but also in the behavior of some higher animals. These concepts must have had certain evolutionary benefits. He gives the following example:

> A monkey will present its back to another monkey, who will pick out parasites; after a time the roles will be reversed. A monkey that fails to return the favor is likely to be attacked, or scorned in the future. Such reciprocity will pay off, in evolutionary terms, as long as the costs of helping are less than the benefits of being helped and as long as animals will not gain in the long run by 'cheating' – that is to say, by receiving favors without returning them. It would seem that the best way to ensure that those who cheat do not prosper is for animals to be able to recognize cheats and refuse them the benefits of cooperation the next time around. This is only possible among intelligent animals living in small, stable groups over a long period of time. Evidence supports this conclusion: reciprocal behavior has been observed in birds and mammals, the clearest cases occurring among wolves, wild dogs, dolphins, monkeys, and apes.[68]

It must be noted, however, that the evolutionary explanation of certain behavioral traits – which also involve behavioral expectations – does not mean either that they should be considered to have a moral nature, or be judged to be ethically right. But it seems obvious that people who see themselves as ethical are more inclined to cooperate with others who are ethical (or see themselves as ethical). In some societies, cheating is strictly forbidden, while in others it is tolerated, and both can be viewed as evolutionary advantageous behaviors that lead to success. Singer is right when he proposes that our ethical expectations are deeply rooted in our past.

In any case, he does not fall prey to a naturalistic interpretation of ethical behavior, as he makes the distinction between the origin of ethical norms and their normative value: "So while I have claimed that evolutionary theory explains much of common morality, including the central role of duties to our kin, and of duties related to reciprocity, I do not claim that this justifies these elements of common morality."[69] An evolutionary understanding of the origin of moral norms, however, is important for ethical theory, but "in an indirect way".[70]

Secondly, Singer criticizes Rawls' "reflective equilibrium" theory, which defines the method of ethics as the struggle to reach a state of equilibrium between theory and moral judgments.[71] If they are in a state of imbalance, then one or both must be amended or altered until a balance is reached. Singer sees the same problem in Rawl's theory as in the evolutionary approach to ethics: it tries to follow the example of the scientific method and adjust the theory to the facts. It does not take the peculiar nature of ethics into consideration; namely, that it is not a descriptive or explanative discipline but a normative one. According to Singer, it might not even be disastrous if all intuitive judgments were in contradiction to an ethical theory. "It (the theory- GK) might reject all of them, and still be

superior to other normative theories that better matched our moral judgments."[72] At this point, he seizes the opportunity to formulate his own ethical credo: "A normative moral theory is an attempt to answer the question 'What ought we to do?' It is perfectly possible to answer this question by saying: 'Ignore all our ordinary moral judgments, and do what will produce the best consequences'."[73]

Thirdly, Singer aims to show how unreliable intuitive judgments with the help of psychological experiments. He refers to the research of psychologists Jonathan Haidt and Joshua Greene. In his pioneering paper, "The Emotional Dog and its Rational Tail", Haidt claims that "moral reasoning does not cause moral judgment; rather, moral reasoning is usually a post-hoc construction, generated after a judgment has been reached".[74] Moral judgments are much more based on pre-rational intuitions than rational considerations.

Haidt uses an imaginary scenario in his experiments and asks his readers to explain their judgment in the following case:

> Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that, was it OK for them to make love?[75]

Haidt points out that while most people judge their action to be wrong, they fail to give adequate reasons for their judgment. The reasons seem inadequate because all the dangers associated with sexual relationships between relatives are eliminated in the description. No offspring will result from their sexual affair, and they will not have to face any negative psychological consequences - the shared experience even makes their relationship better and closer according to the given narrative. One may of course raise the question whether the subjects of the research were not tricked by the simple elimination of the negative consequences. One might as well ask whether war can be considered good if it does no harm to anyone or anything. It is not hard to see the point of the example, namely that we carry innate judgments about certain forms of behavior, which were there for good reasons throughout the evolutionary process but, in a situation like the one described above, no longer hold. Singer also points to the research of Joshua Greene who uses modern diagnostic techniques such as MRI to map the changes caused by dilemma situations in the nervous system of his subjects. Greene found that people's emotional responses to the Fat Man case were much stronger than to the Spur one.

Singer places the results of both psychologists into an evolutionary context. The explanation for why we are unwilling (or at least hesitant) to push the Fat Man off the footbridge is evolutionary. For thousands of years, humans lived in small groups where close bodily encounters were of much greater importance than they are today, and these encounters were filled with an emotional surplus. Remote actions, like altering the route of a trolley, were not relevant. According to Singer, this is why we respond with stronger emotions to the Fat Man case than to the Spur one: "The thought of pushing the stranger off the footbridge elicits these emotionally based responses.

Throwing a switch that diverts a train that will hit someone bears no resemblance to anything likely to have happened in the circumstances in which we and our ancestors lived."[76]

The point of Singer's argument is that intuitions are a product of evolutionary processes and derive their validity from past circumstances. It therefore makes no sense simply to take them to be morally right and to build an ethical theory upon them. Intuitive judgments are in need of ethical examination to determine their correctness. This conclusion has enormous consequences for trolleyology because "there is no point in trying to find moral principles that justify the differing intuitions to which the various cases give rise. Very probably, there is no morally relevant distinction between the cases."[77]

The analysis provided by Singer is fundamental, since it makes a clear distinction between descriptive-explanatory disciplines and the field of ethics, thereby helping us to identify the role of intuition in ethical theory. However, his criticism must be amended by some positive observations concerning the importance of intuitions in the field of ethics.

First of all, thought experiments are designed to provoke intuitions, and when these results are challenged a space opens for their rational consideration. When facing a dilemma situation the reader needs to juxtapose competing intuitions about what ought to be done. Once intuitions are made visible responsible solutions can be sought and the rightness of the intuition can be challenged.

Second, intuitions can be endorsed and educated.[78] Not just ethical, but all intuitions are governed by tacit systems of thought which can be shaped and developed.[79] The malleability of intuitions reveals the need for their ethical examination and education. This shows that intuitions fall prey to over-criticism from normative ethics, which may erode their importance to virtue ethics. It is beyond a doubt that the intu-

itions of virtuous people are of vital importance to ethical theory, which may connect character education with a training in moral reasoning.

Third, the question might arise whether intuition can be viewed as an unexamined piece of tradition. Morality is often viewed as the objectification of previous judgments of conscience. But isn't it more precise to describe morality as a blend of judgments of conscience and intuitive judgments? And is it wise to disqualify the latter simply on the grounds that they were born under circumstances different from ours?

Fourth, life would be impossible without intuitions, including ethical ones. They make our actions and the actions around us calculable and predictable, and render our social world reliable.

Fifth, Singer is right in claiming that ethics must not remain at the level of tailoring ethical theories to our intuitions. He misses an important point, however, which is critical to ethical research: intuitive judgments can serve as useful tools to uncover what missing from our ethical theory. For example, Kant failed to listen to his intuition – I suppose he had this intuition – that it was right to lie to save a friend's life. Had he done so he would have been able to review his theory of deontology.

Singer makes the same mistake as Foot did – and what Anscombe criticized in her comment – namely that he identifies rationality with mathematics or, in other words, with utilitarian thinking. Singer writes that

> reasoning can overcome an initial intuitive response. That, at least, seems the most plausible way to account for the longer reaction times in those subjects who, in the footbridge example, concluded that you would be justified in pushing the stranger in front of the trolley. (…) That reasoning leads us to throw

> the switch in the standard trolley case, and it should also lead us to push the stranger in the footbridge, for there are no morally relevant differences between the two situations[80]

The only argument he is able to deliver in favor of utilitarian thinking, however, is that it "does not seem to be one that is the outcome of our evolutionary past."[81]

## APPROACHING INTUITIVE JUDGMENTS

Singer thus questions the value of intuition based ethical theories, but so did Peter Unger. In contrast to Singer, Unger does not place the nature of intuitive judgments in the foreground of his criticism, but rather explores the question of how we approach intuitive judgment when formulating ethical theories. In *Living High and Letting Die*, his book mentioned above, Unger distinguishes between two opposing approaches to intuitive judgments in ethically relevant cases: preservationism and liberationism. The former holds that "our untutored intuitions on cases (almost) always are good indicators of conduct's true moral status".[82] Thus intuitive responses are well fitted to show "our deepest moral commitments, or our *Basic Moral Values* underlying our moral intuitive judgments.[83] Preservationist ethics are based on just such intuitive judgments, which uncover the "true nature of the Values" and also shed light on "the nature of morality itself".[84] Accordingly, preservationist ethics is conservative by nature, trying to match norms with intuitive answers. It is also uncritical, since it relies on intuitive judgments instead of being critical towards them.[85] On the contrary, liberationists are characterized by a critical and suspicious spirit towards intuitive judgments. They hold that "folks'

intuitive moral responses to many specific cases derive from the sources far removed from our values and, so, they fail to reflect the Values, often even pointing in the opposite direction".[86] As the term *liberationists* suggests, their major aim is to liberate humanity from the chains of intuitive judgments, as opposed to *preservationists*, who strive to preserve the exclusive significance of intuition in the foundation of any moral theory.

The core of the debate between the two parties lies in the "different preferences of Preservationists and Liberationists", which "arise from different epistemic meta-intuitions about the psychological process generating our case-based moral intuitions".[87] But does intuition simply mirror our Basic Moral Values, or do heteronomous factors also influence intuitive judgments?

Unger's question is justified, yet dividing the approaches to intuition into two categories misses the point of the problem. On the one hand intuitive judgments are not alien to the Basic Moral Values of a given person. Quite the contrary, intuitive judgments are often unexpected and surprise the very person in whom the intuition arose. (For example the intuition induced by the Parable of the Good Samaritan in the Jewish audience.) Thought experiments can successfully question the moral order present in our horizon only by inducing nonconforming intuitions. On the other hand intuitions do not constitute a complete description of our Basic Moral Values, since they first manifest themselves in an unreflective form. This is shown by the contradictory intuitions induced by thought experiments which differ only slightly, as in the different versions of the trolley case. It must be noted that Unger himself surpasses these radical schemes when distinguishing between first order intuitions, which are won by reacting to a single particular case, and second order intuitions, which evolve through the comparison of different cases.[88]

The contrasting of liberationists and preservationists is useful as long as it makes possible different stances concerning the human faculty of intuition. These answers are, however, too approximative and fail to show the advantages of using intuitions to formulate ethical theories, since they either approach intuition uncritically or reject it categorically.

## TROLLEY DILEMMAS AND REAL-LIFE CASES

The Trolley Dilemma emerged from the head of a philosopher and is therefore an imaginary scenario, nevertheless, it can be used to shed light on real cases. Although dilemma cases are rare in everyday life, they constitute a common problem on the systemic level. This is the case with autonomous vehicles (AVs), which "should reduce traffic accidents, but will sometimes have to choose between two evils, such as running over pedestrians or sacrificing themselves and their passenger to save the pedestrians".[89] But on what principle should autonomous cars decide whether to put their own passenger or pedestrians at risk? The utilitarian calculus may seem an obvious choice here, following the principle that saving the greater of lives is better than saving the lesser. But there is a more practical question as well: who would want to buy a car which, instead of protecting its passengers – in our case, the owner of the car, and his or her family –, "decides" to protect the lives of pedestrians in a dilemma situation? There are probably very few people who would accept being assigned the role of the victims by their own cars in a possible future dilemma situation.

There have been efforts to get round this possible dilemma by providing autonomous vehicles with an algorithm based on previous intuitive judgments, which would provide orientation

and enable "decision-making". The Moral Machine Project[90] at MIT was invented to pursue this goal of aligning "moral algorithms with human values" and to "start a collective discussion about the ethics of AVs - that is, the moral algorithms that we are willing to accept as citizens and to be subjected to as car owners."[91] The question is whether it is possible to provide autonomous vehicles with an ethically founded formula by collecting the intuitive judgments of respondents to imaginary cases. The creators of the program also acknowledge the difficulty of finding an adequate formula to regulate self-driving cars. The core of the problem is that "although people tend to agree that everyone would be better off if AVs were utilitarian (in the sense of minimizing the number of casualties on the road), these same people have a personal incentive to travel in AVs that will protect them at all costs."[92] Utilitarian ethics are often counterintuitive, in some cases, even immoral. A family with little children would hardly choose a vehicle that functioned upon utilitarian principles, since they would fail in fulfilling their responsibility to protect their children.

James Keenan, approaching the problem from the perspective of virtue ethics, calls attention to the importance of the virtue of fidelity. He proposes a new list of cardinal virtues – justice, fidelity, self-care, and prudence –, and refers to Carol Gilligan when claiming that "the human must aim both for the impartiality of justice as well as the development of particular faithful, partial bonds".[93] Fidelity is paralleled with justice, which must tally with the claim for universality. In contrast, "fidelity rests on partiality and particularity".[94] It is defined as "the virtue that nurtures and sustains the bonds of those special relationships that humans enjoy whether by blood, marriage, love, citizenship, or sacrament".[95] But how is it possible to find a balance between justice and fidelity, the claim to stay clear of "favoritism" and the duty to nurture and foster "par-

ticular relationships".[96] Self-driving cars illustrate these difficulties: is it possible to be just – since we are making decisions concerning self-driving cars ignorant of whether we will find ourselves in the role of a pedestrian on the sidewalk or a passenger in the car –, and to fulfill the claim for fidelity, knowing that the self-driving car might prefer to protect someone other than the passengers in the car.

## TROLLEYS AND PERSONAL RELATIONSHIPS

Trolley Thought Experiments cause similar problems when it is one of our relatives who happens to be standing on the sidetrack. In such a situation, it is knotty question whether we should turn the switch in a different direction; and even more so if that particular relative is guilty of causing the dangerous situation, for example, if he is the one who has tied the five persons to the other track. Utilitarian arguments and loyalty to kin are in conflict in these trolley examples, since one has to choose between letting relatives or strangers live.

The situation described above is not entirely alien to real life. This dichotomy often appears in bioethical debates, especially concerning early-life decisions. It is not by chance that Foot formulates it in connection with bioethical questions (e.g. abortion, allocation of a scarce drug). Thus, there is reason to think that the issues raised by trolley dilemmas are highly relevant to bioethical questions. János Kis, for example, describes the dilemma of Siamese twins as follows:

> Sarah makes her choice. Sarah gives birth to Siamese twins, who are attached at the head. The doctor tells her that if the children remain in this state, they will both die before the age of three, and their short lives

> will be one of suffering. Disjoining them is only possible with an operation that will surely kill one, but the other may survive and grow into a healthy adult. Sarah asks about the chances of survival. The doctor spreads his hands and says, 'To be honest, I don't know.' Sarah has to make her choice in a state of complete uncertainty. It is now harder for her to make a choice than if the doctor had said, 'The other one will surely remain alive.' But the uncertainty does not deprive the mother from the possibility of deliberation. She has to make her choice, and has to take responsibility for it.[97]

This example does not only exist hypothetically, but also in reality. Such is the case of "the Maltese Siamese Twins". Jodie and Mary – the names are pseudonyms – have a lot in common with the example mentioned above and also with trolley scenarios. The sisters were conjoined twins who, after being faced with the diagnosis, were sent to the United Kingdom so that the appropriate medical expertise and equipment for the complicated birth of the twins may be available. The children were born on 8 August 2000, but soon after their birth it became clear that their lives were imperiled owing to their conjoined state. The doctors suggested surgery, as a result of which Jodie, the healthier, would probably survive and live a physically normal life, but Mary would die. The parents were against the operation and, as devout Catholics, emphasized the sanctity of Jodie's life. They also stressed the burden of familial separation, and the problem of the possible discrimination that Jodie might suffer after her return to Malta. The Court of Appeal finally ruled that the operation to separate the two sisters should take place. The surgery was completed according expectations: Jodie survived, while Mary passed away on 6 November, 2000.

What intuitions does the case trigger? It very much depends on one's role in the story. It is highly probable that the intuitive judgment of the doctors involved in the operation was influenced by their professional obligation to save lives. It was rather the parents who had to struggle with their conflicting intuitions: one compelling them to take equal responsibility for the lives of both their children and the other emphasing the impending danger of the almost certain death of both without the operation, and the consequent imperative to at least save Jodie. They also had an even more powerful intuition with which to struggle: the abhorrence of letting Mary be killed.

Cathleen Kaveny claims that the fundamental questions to be answered in this case are whether "(1) the operation did not constitute the intentional killing of Mary, and that (2) the operation was not otherwise unjust or unfair to Mary."[98] She adds that the "intent or purpose of the operation was to separate the babies; Mary's death figured neither as an end nor as a means in the surgical team's actions; it was a foreseen and unintended side effect. In addition, going forward with the operation was not unfair to Mary".[99] She supports this judgment with two arguments: first, it was "Jodie's right to be free of a physical connection to Mary that not only impinges upon her bodily integrity, but also saps her very lifeblood", and second that "both babies soon would be dead without" the operation.[100]

However, questions may arise concerning this line of argumentation. As Bratton puts it: "Was Mary a parasite putting Jodie's life at greater risk? Was Jodie an individual with an unusual anatomy? Or were Jodie and Mary unique people requiring a unique response from the doctors and the courts?"[101] It is a further question whether parental intuitions may add something to the previous, principle-based judgments. Claudia Wiesemann, in her book *Von der Verantwortung ein Kind zu bekommen [About the Responsibility of Having a Child. An*

*Ethics of Parenthood]*, emphasizes the rift between the court's and the parents' point of view. While the court describes the relationship between Mary and Jodie as a non-mutual, parasitic relationship, the parents saw it as something different: "The parents, on the other hand, saw both of them as their children and had given each their own name. For them it was their two babies, between whom it was impossible for them to make such a fundamental difference."[102]

Wiesemann highlights the imperfection of the perspective provided by trolley scenarios, as she shows the limits of an approach that views the twins as two separate entities. She argues that they both should be viewed in their relationship to each other, in their "common bodily existence".[103] This enables us not only to go beyond the question of rights, but also to ask: "How can the relationship between the two be promoted and strengthened? What is in the mutual best interests of the two siblings?"[104] Thus, the aspect of care does not settle for the information provided by trolley scenarios but asks for more: namely, to go beyond the intuitions and their meaning and relevance, and to give everyone who is somehow involved in the scenario visible voice. Both abstract principles and intuitions fall short when it comes to assessing a situation in terms of care and relationships.

The scenario of the conjoined twins pushes the dilemma situation to the extreme. In terms of the Trolley Example, a parent has to choose between the options of leaving his or her two children on the same track with a runaway trolley careening towards them, or moving them while knowing that there is only a small chance that they will survive. This pushes our intuitions beyond their limits and shows how unhelpful trolley examples are in resolving real dilemma situations. They simply cannot confront the complexity of either our inner reality or the real world. It is possible to hide behind a given formula

by claiming that the most lives ought to be saved, but in such cases these formulas are not only counterintuitive but also fail to evaluate the morality of any action. They gloss over the complexity of situations that involve dilemmas, and obscure idiosyncratic details that do not fit the pattern.

The real-life example of the conjoined twins demonsrates that trolley problems may not be useful for solving ethical questions in real life. They show how our intuitions can be influenced by different, often unnoticed elements in a particular scenario, and warn us to be cautious about how we react to them. However, they fail to untie the Gordian knot produced by contradicting intuitions, as we have seen in the case of the Maltese conjoined twins.

## WHAT DO TROLLEY EXAMPLES TEACH US?

Trolley examples challenge not just our intuitions but also moral norms which are held to be self-evident in daily life. In the case of trolley examples, the imperative "Thou shall not kill!" takes a central position, only to be challenged by the dilemma-situations laid out in the different trolley scenarios. They provide information not only about the intuitions to which they give rise in their readers, but also about how we perceive our world. Trolley thought experiments entail a certain implicit anthropology, which unfolds in the clash between intuitive responses and moral expectations.

Trolley scenarios, and the intuitive responses they provoke, suggest that we live in a world which doesn't fit us perfectly.[105] This not-fitting-into-the-world affects every level of our existence and makes the discipline of ethics so necessary. If ours were a perfect world there would be no sense in acting immorally. The dilemma situations described by the trolley

scenarios highlight this state of not-fitting-in. If we were living in a perfect world, the trolley could stop before reaching the track-workers. But in trolley scenarios, the agent has to choose between two alternatives, which, from the given perspective, are of the same moral weight. It is only possible to choose one of the alternatives, therefore the "agent (…) seems condemned to moral failure; no matter what she does, she will do something wrong (or fail to do something that she ought to do)".[106] This failure is primarily an existential rather than a moral failure. This is affirmed by responses that not only provide information about which alternative the agent would choose, but also go futher by considering the existential relevance of the choice.[107]

Such dilemma situations are rare in real life, though situations might arise from time to time that seem to leave only two equally bad options open. The question remains whether completely unresolvable dilemmas ever emerge in real life, or whether they are only produced by our own narrow horizon, which does not reveal all the possible options on the practical and the moral level. A conflict such as a problematic pregnancy in which the mother's life can be only saved at the cost of the embryo's life, is a typical example: no matter what choice is made, it will cause serious existential harm, even if it can be justified on the basis of a moral principle.

Another important insight gained through the analysis of trolley examples is the insufficiency of the analogous application of impersonal solutions. In case of a conflict in pregnancy it is not alien track-workers who are affected by our choice but people who are in a personal relationship with one another: husbands, fathers, mothers, and children. Critics of abstract ethical thinking claim that textbook examples of principles such as the doctrine of double effect fail to recognize the importance of interpersonal relationships, such as the bond between the child and the mother in the case of pregnancy.[108]

In reality, pregnancy is a human relationship based on responsibility. The anguish resulting from some pregnancies is not due to the biological incompatibility of mother and child but the mother's experience of being unable to fulfill her parental responsibility with regard to the child.[109] Here, abstract rules hardly offer a solution that is both morally and existentially justified.

## Chapter VIII

# THE VIOLINIST ANALOGY

Along with the Vietnam War, abortion was one of the most intensively debated questions in 1970s American public discourse. The Supreme Court's 1973 ruling in *Roe v. Wade*, which concluded that abortion is a fundamental right ensured by the United States Constitution, mobilized both pro-life and pro-choice groups to such an extent that their antagonism has been a defining factor of the American political landscape up to the present day. According to Balkin "Roe was merely the opening event in a political and legal struggle over reproductive rights that continues to this day."[1] No wonder that the journal *Philosophy & Public Affairs* participated actively in the debate over abortion. There are numerous articles from this decade discussing almost every relevant philosophical aspect of the topic.

Every important philosophical question touching upon the issue of abortion had been discussed in the journal, including the moral status of the embryo, the moral relevance of the different stages in embryogenesis, the distinction between abortion and infanticide, the embryo's right to life, and women's right to have control over their body. In his article "Understanding the Abortion Argument", Roger Wertheimer discusses the

different attitudes towards the fetus using student opinions.[2] In "Abortion and Infanticide" Michael Tooley examines "what properties a thing must possess in order to have a serious right to life";[3] John Finnis criticizes the vagueness of the language of rights concerning the morality of abortion, and gives reasons "why the foetus from conception has human rights, i.e. should be given the same consideration as other human beings".[4]

It was Judith Jarvis Thomson's "A Defense of Abortion", with its famous analogy of the violinist, however, that became one of the most influential texts, and which is likely to be "the most widely reprinted essay in all of contemporary philosophy".[5] What makes this essay extraordinary is its opposition to the then-contemporary trend in argumentation among philosophers, theologians and political activists for whom the question of the embryo's moral status – i.e. whether the embryo is a human being or person – was central to the abortion debate. Thomson in her article puts this problem aside. Moreover, she succeeds in avoiding the question of the fetus's right to life in her argumentation about abortion. N. Ann Davis summarizes her achievement as follows: "Whether or not one shares Thomson's views about abortion or is persuaded by the arguments and examples she presents in ADA [*A Defense of Abortion*], it must be acknowledged that ADA has had a lasting influence on the way philosophers think about abortion and other normative issues, and on how they view moral theory."[6] Its significance is not only due to its topic, but also to its exemplary nature concerning the functioning and its power to involve its readers in the course of argumentation:

> The power of *A Defense of Abortion* thus lay not merely in the success of its attack on restrictive views of abortion, nor in the subtlety of an approach to doing philosophy that made the reader a participant rather

> than a mere analyst or observer. In casting both the teacher and the student in the role of participants, it helped turn the teaching of philosophy into a form of (more) democratic collaboration, one that engaged the student and the teacher both with the material and with each other.[7]

However, despite its incontestable merits in popularizing philosophy, the question remains whether it reaches the aim of its author: namely, to function not only as an example, but as an integral part of ethical argumentation.

## THE ORIGINAL THOUGHT EXPERIMENT

In her 1971 article Thomson formulated the Violinist Thought Experiment as follows:

> …now let me ask you to imagine this. You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital now tells you, 'Look, we're sorry the Society of Music Lovers did this to you – we would never have permitted it if we had known. But still, they did it, and the violinist now is plugged into you. To unplug

> you would be to kill him. But never mind, it's only for nine months. By then he will have recovered from his ailment, and can safely be unplugged from you.' Is it morally incumbent on you to accede to this situation?[8]

The description of the imaginary scenario happens to be much longer here than in the case of common thought experiments. Although it contains some fabular elements, these – at least for readers who are not well trained in medicine or biology – are not the medical features, but rather a means to introduce the Society of Music Lovers and the violinist to the story. Although they seem to be accidental elements, they constitute essential – but nonetheless replaceable – parts of the thought experiment: the story would function just as well if it were the Society of Chess Lovers who kidnapped someone in order to save a world famous chess player whom the audience has never heard of before. Still, the thought experiment made a name for itself as the "Violinist", which shows the importance of marketing in propagating philosophical publications. What might cause the medically informed reader some difficulty is Thomson's rather simple and vague description of the interconnection of the two circulatory systems. Such an episode, however, might well be imaginable for the non-professional or general public who take the inevitable and regular progress of medicine for granted. The crux of the thought experiment lies neither in the inclusion of the violinist and the Society of Music Lovers, nor in its medical description, but in the analogy between the described case and pregnancy.

## THE STRUCTURE OF THE THOUGHT EXPERIMENT AND ITS ARGUMENTATIVE CONTEXT

The success of the Violinist Example does not lie in the right choice of protagonist in the person of the violinist, but in its remarkable design, outstanding among other thought experiments. The central concept of the article is abortion, with an analogy between the imaginary scenario described and pregnancy. Just like the embryo in its mother's womb, the violinist is dependent on the kidnapped person's aid for nine months. Beyond that both cases describe situations of asymmetrical relationships. The connection of the two circulatory systems strengthens the parallel view, since the imaginary tubes resemble the umbilical cord. (Certainly there are significant differences too – acknowledged by Thomson –, which will be discussed later in this chapter.) The discontinuation of the physiological connection would lead in both cases – in the violinist-case certainly, in case of abortion mostly – to death for the dependent party. It is possible to terminate the physiological connection, but only at the price of the dependent's death.

But what is the relevance of this analogy concerning the ethical assessment of abortion? Thomson uses it to test certain beliefs and theses, which play a central role in the abortion discourse. She primarily targets those who oppose abortion on the grounds that the fetus is a human being, moreover a person, which fact renders abortion immoral. Thompson gives a general overview of their typical line of argumentation:

> A1 "Every person has a right to life."
> A2 "So the fetus has a right to life."
> A3 "No doubt the mother has a right to decide what shall happen in and to her body…"
> A4 "…a person's right to life is stronger and more

> stringent than the mother's right to decide what happens in and to her body, and so outweighs it."
> A5 "So the fetus may not be killed; an abortion may not be performed."[9]

Thomson's argumentation against these theses constitutes the backbone of the train of thought in the text: she contests the view, which holds that from the statement "the fetus is a person" would follow the imperative concerning the "impermissibility of abortion".[10] Thomson sees this step to be more problematic than it appears at first glance. The starting point of her argumentation is the criticism of the "extreme view" according to which "abortion is impermissible even to save the mother's life" and which does not allow for exceptions, either if the pregnancy was a result of rape, or in "a case in which the mother has to spend the nine months of her pregnancy in bed", or if "the continuation of the pregnancy is likely to shorten the mother's life".[11] The imaginary scenario was thus designed to respond to those who would reject abortion under all possible circumstances.

An important rhetorical component of the argument appeals to how our moral intuitive machinery responds to this imaginary scenario. It seems probable that no one would want to find herself in such a situation: forced bedrest for nine months with an unconscious stranger whose life literally depends on us is hardly a desirable scenario. Thomson also believes that the audience of the thought experiment "would regard this as outrageous", especially if someone were to justify this situation with the argument that "all persons have a right to life, and violinists are persons".[12] Surely, there might be exceptions with regard to the negative intuitive judgment. For example someone might interpret the situation as proof of his or her unique ability to help. But these are exceptions. The

text presupposes that most people have goals in life other than saving someone else's life by taking such extreme measures. Similarly, Thomson assumes a certain intuitive indignation on the part of the audience over the usurpation of the kidnapped person's autonomy. The person in whose position the audience finds itself when hearing the imaginary story did not volunteer for the job. Finally, the audience must accept the death of the violinist and the death of the embryo both as something undesirable and also objectionable. If the death of the violinist or the embryo were desirable – for example, the violinist would make us sick as time passed by, sucking out every drop our vital power by the end of the nine months – the answer to the thought experiment would be less difficult: most people would intuitively opt for the unplugging. Similarly, one would rather choose not to stay plugged in if the violinist had made himself sick on purpose – e.g. by using intravenous drugs, knowing that his acts would lead to severe damage to his health – and had kidnapped and connected us to his circulatory system as part of his plan to recover from the condition he knowingly caused. In the thought experiment presented by Thomson it is not the violinist, but his fans who kidnapped the 'helper', and the violinist cannot be blamed for either for his condition, or for the kidnapping.

The death of the violinist must be something which is not desired by the audience in any way, and is considered a bad outcome. The dilemma in the story –to save an unconscious stranger by staying bound to him for nine months or to sustain our autonomy by unplugging ourselves from his circulatory system and letting him die – results from this negative view of the violinist's death. Most respondents would be happy to see a third option where the violinist could be saved by means of a newly discovered drug, for example. In such a case, the intuitive judgment would overlap with the moral judgment. How-

ever, the obligation to choose between autonomy and physical integrity on the one hand, and the life of a human being on the other, makes this choice more difficult.

Thomson makes two further preliminary assumptions in her argumentation. The first is what may be termed, for the sake of simplicity, the distinction between rights and morality. She does not raise the question of whether it would be morally justified to leave the violinist alone in his grave situation, but asks whether the violinist's right to life implies the duty to waive control over your own body, even if this entails a nine-month hospital stay to provide the sick man with the necessary physiological support to sustain his life. She continues her description of the imaginary scenario with the following question: "Is it morally incumbent on you to accede to this situation? No doubt it would be very nice if you did, a great kindness. But do you have to accede to it?"[13] It is clear from these comments that Thomson moves within the framework of norm-ethics, but turns as the argument progresses towards the question of rights. The right to life and the right to make decisions concerning one's body remain two basic concepts in her line of thought. The only question is whether one of these rights trumps the other, and if so, which one?

Thomson's second presupposition is that it is not necessary to determine the moral status of the fetus in order to ethically assess an abortion under certain conditions. For the argument's sake, she proposes "that we grant that the fetus is a person from the moment of conception".[14] First, she asserts several times that she has "only been pretending (…) that the fetus is a human being from the moment of conception", and that she considers "a very early abortion (…) surely not the killing of a person".[15] She is ready to accept, merely for the sake of argumentation, that the fetus is a human being or a person. Her example of the violinist aims to evoke in the reader certain

mortal intuitions that would ethically justify abortion even if "the fetus is a person from the moment of conception."[16]

To support the sophisticated distinctions made by the use of intuitive judgments, Thomson designs further imaginary scenarios, which are less realistic, even satirical: the case of Henry Fonda's Cool Hands, People-seeds, and the Growing Child Example. Although these are treated less frequently by violinist literature, they constitute important additions to the central imaginary scenario.

## THE GROWING CHILD

The most elaborate additional example in Thomson's article is about the growing child:

> Suppose you find yourself trapped in a tiny house with a growing child. I mean a very tiny house and a rapidly growing child – you are already up against the wall of the house and in a few minutes you'll be crushed to death. The child on the other hand won't be crushed to death; if nothing is done to stop him from growing he'll be hurt, but in the end he'll simply burst open the house and walk out a free man. (…) However innocent the child may be, you do not have to wait passively while it crushes you to death. Perhaps a pregnant woman is vaguely felt to have the status of house, to which we don't allow the right of self-defense. But if the woman houses the child, it should be remembered that she is a person who houses it.[17]

The first question concerning this example is the intuition it induces. Most parents – whether fathers or mothers (since rep-

resentatives of both sexes might be present in the imaginary building) – would hardly consider the physical destruction of their child to be the proper course of action, even if it were done out of self-defense. It is conceivable that most would opt for the life of their child. Of course, Thomson speaks here of a small house and a growing child, not about a parent and her child. But most people would not opt for the destruction of the child, even if it were a total stranger, if they knew that the child would remain alive after outgrowing the house.

In the Growing Child Example, Thomson chooses an imaginary scenario that serves her purposes less than a real example would. The analogy corresponds to the situation when pregnancy endangers the mother's life. The general moral intuition – which, of course, has its exceptions – suggests that it is the life of the mother that must be saved in such a case. It would have suited Thomson's purposes better if she had used a concrete case here, which might have elicited readers' moral approval of abortion, at least in cases when the life of the mother is in danger. Thomson still thinks that the analogy of the growing child justifies the claim that "a woman surely can defend her life against the threat to it posed by the unborn child, even if doing so involves its death".[18]

## THE QUESTION OF OWNERSHIP – THE COAT

While in the case of the growing child the house is a symbol of being locked in one's own body and being at risk, the coat, which is the central object of the following imaginary scenario, represents ownership and protection:

> If Jones has found and fastened on a certain coat, which he needs to keep him from freezing, but which

> Smith also needs to keep him from freezing, then it is not impartiality that says 'I cannot choose between you' when Smith owns the coat. (…) Smith after all, is hardly likely to bless us if we say to him, 'Of course it's your coat, anybody would grant that it is. But no one may choose between you and Jones who is to have it.'[19]

By this example, Thomson creates a dilemma. While the growing child scenario generated an intuitive dilemma between the death of the parent – with whom the audience was expected to identify – and the death of the child, the coat example centers on two 'third persons', Smith and Jones. Furthermore, one of them, (Smith) is the owner of the coat. Thus, the example seeks to justify an intuitive judgment based on ownership, even though the example also entails the question of life and death. There is, however, a counter-intuition that it would be wrong to let either of them freeze. The decision to give the coat to Jones might also be justified in terms of ownership.

Thomson draws a parallel here with pregnancy, claiming that the mother is the owner of her body also while pregnant, and since she owns it and needs it for her own sustenance, the fetus does not have a right to it. The analogy, however, is imperfect. It only fits if the pregnant woman's life is endangered by the fetus.

Thomson refers to the question of justice in pregnancy by using the following example: not only will the stolen coat be returned to its owner, but "justice seems to require that somebody do so".[20] She seems to exclude the possibility of a conscience clause because she is (inexplicitly) referring to the doctors who have expertise in performing abortions when she writes that "anyone in a position of authority, with the job of securing people's rights" should give the coat back to its owner, i.e. perform the abortion on the pregnant mother.

## DUTY AND KINDNESS – HENRY FONDA'S COOL HANDS

Besides the Violinist Example, the imaginary scenario about Henry Fonda's Cool Hands is the other often-quoted part of Thomson's article:

> If I am sick unto death, and the only thing that will save my life is the touch of Henry Fonda's cool hand on my fevered brow, then all the same, I have no right to be given the touch of Henry Fonda's cool hand on my fevered brow. It would be frightfully nice from him to fly in from the West Coast to provide it. (…) But I have no right at all against anybody that he should do this for me.[21]

Thomson draws a parallel between Henry Fonda's deed and the mother's act to lend her body to the fetus. There is an intentional disproportion between the violinist case and pregnancy on the one side and Fonda's deed on the other side: putting a cold hand on a fevered brow – even, if someone has to travel from the West to the East Coast –, is not commensurable with borrowing a kidney or being pregnant. Thomson uses this to make a distinction between "kindness" and something one "can claim from you as a due".[22] Most people might think intuitively that, just because someone is a celebrity, that person is not morally obliged to put his cool hand on the brow of a dying person, even if this person happens to be his most devoted fan. Celebrities are not duty-bound to heal or comfort others.

But what if we alter the thought experiment? How would we intuitively evaluate the situation if we were in the place of Henry Fonda? Would we feel the moral duty to travel to the ill admirer and to heal him with the touch of our hands on his brow? Would we want to raise the duty to help to the level of

universal law? It is reasonable to say that this would be impossible. In such a case, Fonda would be transformed from an actor into a travelling miracle doctor. Still, the question might be raised whether Fonda would be required to help sick people to a reasonable extent if he were endowed with such a miraculous hand. (It does not mean that he would have to fulfill all requests, though, because no doctor – no matter how successful – would heal or perform operations day and night.) The example clearly shows the distinction between duty and kindness. However, it is also clear that placing a cool hand on an ill person's brow is morally different from pregnancy.

## DEFECTIVE CONTRACEPTION – THE BURGLAR

A weakness of the Violinist Example is that it fails to apply in all possible cases of pregnancy. It fits, to a certain extent, cases where pregnancy resulted from rape, but mostly fails to apply to "normal", intended pregnancies, or in cases where a child was conceived due to defective contraception or unprotected sexual intercourse. In order to fill this gap Thomson created the example of the Burglar, which is supposed to highlight the rights and responsibilities that might intuitively arise in such situations:

> If the room is stuffy, and I therefore open a window to air it, and a burglar climbs in, it would be absurd to say, 'Ah, now he can stay, she's given him a right to the use of her house – for she is partially responsible for his presence there, having voluntarily done what enabled him to get in, in full knowledge that there are such things as burglars, and that burglars burgle.' It would be still more absurd to say this if I had

> had bars installed outside my windows, precisely to prevent burglars from getting in, and a burglar got in only because of a defect in the bars.[23]

There are three elements in the example, which make the analogy with pregnancy possible. The first is the character of the burglar. It is not immediately clear whether he should be identified with a rapist or with the fetus. A burglar is certainly an intruder who comes without invitation, but his character is not necessarily connected with violence. His identity is clarified by the other two elements, the open window and the defective bars. The open window suggests that the burglar is meant to symbolize the fetus, since there is no violence involved in the burglar entering the house through the open window. However, the entrance of the burglar might also be interpreted as the violation of someone's private sphere. The defective bars certainly stand for defective contraception, e.g. tubal sterilization. Neither the open window nor the defective bars makes it clear whether the burglar represents a pregnancy due to rape or due to voluntary sex that resulted in an unwanted pregnancy.

On the one hand the example of the burglar certainly evokes strong negative intuitions, since there is hardly anyone who would sensibly want a burglar to break into his house. But may this intuition be applied to pregnancies due to rape or defective contraception, and thereby render the fetus a criminal, as it were? Does such an analogy not place all liability on the shoulders of the mother, rendering her responsible for every decision concerning the fetus? Does it not distort our understanding of pregnancy due to rape when the woman is compared to someone who has irresponsibly left a window open?

On the other hand contraception is commonly not used to prevent pregnancy due to rape, although there might be situations in which rape is unfortunately an everyday horror, e.g.

in the case of war. The question remains whether contraception can be required in such situations, as insurance companies can require window bars in areas with high crime rates. Here it becomes clear that it is one thing to manipulate an object and quite another to manipulate the human body.

The burglar analogy seems to be weaker than that of the violinist, since many elements are opaque and fail to correspond to well-defined aspects of a certain pregnancy situation. Although it is a concrete example to which we can relate – since burglars, open windows, and bars are part of our everyday horizon – it fails to build a connection with the hallmark traits of the analogous situation.

## PEOPLE-SEEDS

The burglar analogy is carried further with a less earthy example using people-seeds floating in the air and waiting to find good soil to grow into full-blown human beings:

> …suppose it were like this: people-seeds drift about in the air like pollen, and if you open your windows, one may drift in and take root in your carpets or upholstery. You don't want to have children, so you fix up your windows with fine mesh screens, the very best you can buy. As can happen, however, and on very, very rare occasions does happen, one of the screens is defective; and a seed drifts in and takes root. Does the person-plant who now develops have a right to the use of your house?[24]

The example is obviously a milder version of the burglar-analogy. A people-seed cannot be viewed as an unjust aggressor or

trespasser, such as a burglar. Thomson also highlights the moment of contraception, as she uses the image of mesh screens in the scenario. She points out – in 1971, when contraceptives were much less reliable as they are today – how drastically absolute prevention (of pregnancy) and protection (of the house from people seeds) would affect the life of women. If someone's goal was absolute prevention of, and protection from, a people-seed entering the house and taking root, that person would have to live his life "with bare floors and furniture, or with sealed windows and doors", just like women would have to undergo a hysterectomy or never leave home "without a (reliable!) army" to avoid pregnancy due rape.[25] This analogy refers not only to pregnancies due to rape, but also to consensual intercourse with defective contraception. Thus the people-seed is described not so much as a burglar, but as an unexpected guest. Interestingly, the people-seed example resembles the idea of preformationism, a popular idea in the 16th and 17th centuries, claiming that the sperm or the ovum contained the complete human organism, mostly in the form of a homunculus.[26] People-seeds seem also to contain everything that constitutes a human being. Thomson's example also seems to indicate – although not in a biological, but in a social or ethical sense – that the new human being, like the homunculus or the people-seed, comes only from the man. She effaces not just the active participation of women in pregnancy, but also the possibility thereof.

This is the primary reason why the intuition evoked by the people-seed example can only be adapted cautiously to cases of pregnancies. Just as the case of the burglar, the people-seeds scenario evokes strong negative intuitions. Most people would find the idea of people-seeds arriving randomly at their homes and taking root in their carpet to be repugnant, just as they would refuse to show hospitality to guests arriving at their

homes at any and all times. (Certainly, someone in need of company or wishing for a child would find such an idea less repulsive.) But does the rejection of the idea of people-seeds arriving randomly in our house imply that the fetus does not have a right to his mother's body?

## BOX OF CHOCOLATE

A more practical example is used by Thomson to demonstrate the difference between virtue ethics and justice- or rights-based ethics:

> Suppose that box of chocolates (…) was given only to the older boy. There he sits, stolidly eating his way through the box, his small brother watching enviously. Here we are likely to say 'You ought not to be so mean. You ought to give your brother some of those chocolates.' If the boy refuses to give his brother any, he is greedy, stingy, callous – but not unjust.[27]

At this point in the essay, it becomes clear how sharp the boundary that Thomson draws between law and ethics, and between norm and virtue ethics really is. Greediness, stinginess, and callousness are ethical terms, as is the word justice. Given this impenetrable border between norms and virtues one can, ironically enough, reasonably conclude that a woman who has undergone an abortion was selfish and irresponsible, while the act itself (i.e. abortion) was justified and moral. This shows that Thomson's concept of morality does not embrace the morality of the individual but reduces it to the horizon of rights.

## THE TWO PROCRUSTEAN BEDS

The examples above clearly show the way Thomson approaches the state of pregnancy and the ethical problems raised by her essay. Abortion is, according to her methodological starting point, a question of rights, which can be described with the categories "right to one's body", "right to life", "right to use another person's body", and justice. Other moral categories, such as the inner morality of the agent, play no role in the moral equation set up in the text. The fetus is described with objective images, such as a burglar, people-seeds, or a box of chocolate. Taken out of their context, these examples hardly resemble what we mean by pregnancy in our everyday language.

The same holds true for the Violinist Example. The question is whether the imaginary scenario resembles any other state of affairs in the real world. John T. Wilcox pinpoints the problem:

> Her argument is basically an argument by analogy: pregnancy is supposed to be analogous to being hooked up to an unconscious violinist. And so what is true (given her general principles) in one case is supposed to be true in the other. The force of the argument clearly depends upon the strength of the analogy. But it is clear that there are great disanalogies between the two cases. Indeed, some wags have said that only someone at MIT could have come up with such an implausible analogy.[28]

The last sentence shows the ideological gap created by the article concerning the abortion debate in 1970s America. As mentioned earlier, it is quite clear which elements of the imaginary scenario constitute the rudiments of the analogy to pregnan-

cy. The dependence of the fetus on the mother is certainly one of these elements, and the duration of nine months may be viewed as a symbolic interface, since pregnancy is commonly held to last nine months, despite numerous counterexamples. Thomson's thought experiment, however, is extremely viable because of its divergence from the phenomenon of pregnancy. This difference is capable of evoking moral intuitions, which are difficult to identify at first glance. Many respondents may intuitively agree that no one has a duty to stay connected to the violinist for nine months. Similarly, many respondents may fail to recognize the analogy with pregnancy. Although there are critics who have raised doubts concerning other arguments in the text, such as John Finnis in his 1973 article, "The Rights and Wrongs of Abortion: A Reply to Judith Thomson",[29] most critics target the analogy between the Violinist Example and pregnancy. Let us consider the elements at the heart of this critique.

Eric Wiland places the elements which render the analogy with pregnancy problematic into four categories. These are "1) the issue of consent; 2) the familial relation between the parties in question; 3) the artificiality of the example, and 4) the distinction between killing and letting die".[30] Thomson does acknowledge in her essay that the analogy may not be suited for all possible cases of pregnancy. There is no need for that, since the only thing she wants to demonstrate is that the personhood or humanity of the fetus does not imply its right to be borne by its mother.

Some, like Peter Singer, may judge the analogy to be sufficient, and yet continue to defend the expectation of remaining connected to the violinist. He claims that a "utilitarian would hold that, however outraged I may be at having been kidnapped, if the consequences of disconnecting myself from the violinist are, on balance, and taking into account the interests of everyone affected, worse than the consequences of remain-

ing connected, I ought to remain connected."[31] Singer accepts the validity of Thomson's analogy, but denies the consequence drawn by Thomson, since he rejects her theory of rights. He claims that "if the life of the fetus is given the same weight as the life of a normal person, the utilitarian would say that it would be wrong to refuse to carry the fetus until it can survive outside the womb".[32]

Singer does not actually accept the personhood of the fetus, but simply assumes it for the sake of applying his utilitarian calculus to the situation as arranged by Thomson. Accordingly, the best one can do in such a situation is to stay connected to the violinist for the next nine months. Similarly, the best thing a pregnant woman can do, if she considers the fetus to be a person, is to carry on with the pregnancy for the next nine months. It is important to note that this formula functions only if the value of the life of a person outweighs all other elements in the utilitarian formula.

## THE ISSUE OF CONSENT

Yet the analogy does not cover the key elements of the situation of pregnancy to the extent that Singer claims it does. Quite the contrary, it is obvious that the Violinist Example was tailored to fit the argumentation provided by Thompson and does not fit our everyday experience of pregnancy. First, the subject of the thought experiment finds herself in an unwanted situation. She was kidnapped and conjoined with the violinist while asleep or in an unconscious state. Typical pregnancies, however, are results of consensual sex. Both parties know that even protected sex may lead to pregnancy. Knowledge of this fact implies a certain tacit consent, which should be obvious to every reasonable adult human being. The Violinist Example

is thus only relevant in the cases of pregnancy due to rape. Certainly, there is no need here to think about sexual assault, since it is also possible to kidnap a potential surrogate mother and impregnate her with the embryo of an alien couple. In this case consent would also be absent. If the embryo is considered a human being and a person, consent, even if it is only tacit, makes the responsibility of biological parents towards their offspring (even in the embryonic stage) clear.[33]

## STRANGERS AND FAMILY

Second, there is a significant difference between the violinist and the fetus. While the former is a stranger, the fetus is bound by genetic ties not only to the mother, but also to the father, the siblings, and grandparents. The significance of these ties is highlighted by the intuitive answers, which emerge when the imaginary scenario is altered. How might our approach to the situation change if it involved a mother who is connected to her child:

> Imagine that you are the father of a teenage daughter. You wake up to find yourself connected with the circulatory system of your daughter. It turns out that she is suffering from an extremely rare and fatal illness, which can only be cured if she stays connected with your circulatory system. Although the ambulance team who arrived at your home while you were asleep did not have the chance to ask for your consent, is it morally incumbent on you to accede to this situation?

Most people would intuitively answer this question with a straightforward "yes". It might be asked whether the daughter has the right to use the circulatory system of her father for the interval of nine months, and also whether this right does not conflict with other duties the father is obliged to fulfil, like taking care of his other children. Still, the intuitive "yes" clearly shows that this situation is different from the case of the violinist. A father withholding support from his daughter fails to fulfil his parental duties, while in the violinist case the stranger cannot be blamed, for he has no such obligation towards the sick musician.

The situation of pregnancy is different from both cases. Someone who is willing to have children, or is simply participating in a sexual act which could result in children, would know – unless he or she is remarkably ignorant – that a new being might be conceived, and that this new being would depend physiologically on the mother for a period of nine months and beyond. There is no human life and development in the nine months prior to birth without this dependency, since the invention of artificial wombs is still pending. There is an even bigger difference concerning the experience of pregnancy. While the victim of the Society of Music Lovers finds himself face-to-face with the violinist – just as the father physically sees his daughter in the modified example – pregnancy is not characterized by a face-to-face experience, but primarily by imagination, a turning inwards to sense the life growing inside, and constant development. Furthermore, pregnancy is usually seen as something positive, while this is by no means true of being kidnapped and forced to participate in a medical procedure.

It seems to be inappropriate to use the term "voluntary" for the relationship between parents and children. Parents do not care for their children on a voluntary basis, but because they understand it to be their moral duty. Moreover, for most

parents caring for their children is something conceived of as natural, i.e. not requiring further consideration or reflection. Francis J. Beckwith resorts to care ethics to criticize the lack of distinction between different kinds of human relationships: "By using the violinist story as a paradigm for all relationships, which implies that moral obligations must be voluntarily accepted in order to have moral force, Thomson mistakenly infers that all true moral obligations to one's offspring are voluntary."[34] Interestingly, Beckwith uses a counterexample to evoke intuitions in the readers which run counter to those called forth by the violinist scenario:

> But consider the following story. Suppose a couple has a sexual encounter which is fully protected by several forms of birth-control (condom, the Pill, IUD, etc.), but nevertheless results in conception. Instead of getting an abortion, the mother of the conceptus decides to bring it to term although the father is unaware of this decision. After the birth of the child the mother pleads with the father for child support. Because he refuses, she seeks legal action and takes him to court. Although he took every precaution to avoid fatherhood, thus showing that he did not wish to accept such a status, according to nearly all child support laws in this United States he would still be obligated to pay support precisely because of his relationship to this child.[35]

The difference again is the implicit consent inherent in every sexual act which could, even with severely mitigated probability, result in pregnancy. Beckwith also points to the fact that people under the influence of alcohol are held to be responsible for their actions. Although these are not actual arguments

that could be used in reasoning about abortion – since they are facts about certain judgments based on intuition – they nevertheless are relevant in challenging Thomson's emphasis on voluntariness, which also relies on intuitions elicited by her imaginary scenarios.

The same is true for Beckwith's claim that „Thomson's volunteerism is fatal to family morality".[36] He explicitly refers to intuition as he claims that "a great number of ordinary men and women, who have found joy, happiness, and love in family life, find Thomson's volunteerism to be counter-intuitive"[37] Family morality is based on a belief "that an individual has special personal obligations to his offspring and family which he does not have to other persons".[38] There seems to be no need to justify this proposition for Beckwith because it happens to resonate with our ordinary intuitions. Although there is indeed a place for ethically justifying the different kinds of duty towards kin and stranger, there is no need for it here; the reference to family morality is used here only to reveal intuitions contrary to those of Thomson.

## THE ARTIFICIALITY OF THE EXAMPLE

Thirdly, Wiland points out "the artificiality of the example".[39] He claims that "the kind of dependency a fetus has on his or her mother is the most normal and natural thing in the world; each of us certainly has been there. So the kind of claim the violinist has on you is quite unlike the kind of claim the fetus has on her or his mother".[40] While the time spent in the mother's womb is a natural part of the biography of all human beings – even if only potentially – kidney failure can hardly be viewed as something natural. Being connected to the mother is the natural state of the embryo, while for the violinist and

his forced alien helper being connected can by no means be understood as a natural state. It must also be noted that no direct moral conclusion might be drawn from these distinctions.

The same problem characterizes Wilcox's claim, who describes the violinist case as "weird" and pregnancy as something "usual": "No one in the history of mankind has ever been kidnapped for the purpose Thomson explains. It is not clear, medically, that anyone could be in the situation described – such that only that person's kidneys could save the violinist. But pregnancy is the opposite of weird. (…) actually few things, except death, are as usual as pregnancy. We all, every man Jack of us, and every woman Jill, too, begin in our mother's body."[41] Though his claim may be true, and it certainly conforms to the intuition of most people, it is hard to establish ethical reasoning by merely pointing out the weirdness of a particular example.

Beckwith makes a better claim at this point. He claims that a case can be made that the unborn does have a prima facie right to her mother's body.[42] If there is "a special parental obligation, which does not have to be voluntarily accepted in order to have moral weight" the unborn baby may have "a natural prima facie claim to her mother's body".[43] From the point of view of parental responsibility it is justified to speak of natural obligations and claims, in the sense that these do not require consent. The natural state of the given human being implies certain obligations and claims, which may differ from stage to stage of development. This is showed by the everyday moral intuition that a "newborn has a natural claim upon her parents to care for her, regardless of whether her parents 'wanted' her (…). This is why we prosecute child-abusers, people who throw their babies in trash cans, and parents who abandon their children."[44] At this point Beckwith makes use of the Thomson's starting assumption, namely that

the embryo should be considered a human being, or even as a person, and shows that the embryo might have certain claims under these conditions: "If the unborn entity is fully human, as Thomson is willing to grant, why should the unborn's natural prima facie claim to her parents' goods differ before birth?"[45] It is a question, though, of how far this parental duty extends. It is clear that "the unborn entity is a human being who is by her very nature dependent on her mother, for this is how human beings are at this stage of their development" and that "the womb is the unborn's natural environment", but do these facts mean that a parent may be morally obliged under certain circumstances to donate his kidney (or his blood or bone marrow) to his sick child if she needs a transfusion or a transplant? Nature certainly does inform our moral duties, but only at the level of factual information.

The distinction between natural and unnatural is certainly appealing and does confirm our everyday intuitions to a great extent. Parents have certainly different, and in some sense stronger, duties towards their offspring than towards strangers. However, it is hard to make the transition from the descriptive to the normative. This may only be achieved if duties and claims connected to certain human conditions (e.g., stages of development and relationships like parenthood) could be established, which may be considered in this sense natural and thereby more binding (in the sense that there would be no need for consent) than those conditions considered unnatural. If this could be achieved, the distinction would become highly relevant for the Violinist Example.

## THE DISTINCTION BETWEEN KILLING AND LETTING DIE

Fourthly, the two examples differ in one crucial aspect: one might be described as killing, and the other as letting die. Accordingly, abortion is nothing but the killing of unborn children, while unplugging the violinist would simply mean letting him die. It is disputable whether there is a clear moral difference between killing and letting die – as James Rachles' analysis of the distinction will show. It is clear, however, that intuitively there is a significant difference between the two cases: walking away and letting the violinist die on the one hand and killing a fetus via abortion on the other. One might certainly point at rare cases of abortion, such as hysterotomy, when the fetus is not actively killed, but removed from the womb. This method also results in withdrawing vital support from the fetus, and is similar to the case of unplugging the violinist. Beckwith again appeals to our intuitions using the following imaginary scenario:

> Suppose a person returns home after work to find a baby at his doorstep (like in the film with Tom Selleck, Ted Danson, and Steve Guttenberg, Three Men and a Baby). Suppose that no one else is able to take care of the child, but this person only has to take care of the child for nine months (after that time a couple will adopt the child). Imagine that this person will have some bouts with morning sickness, water retention, and other minor ailments. If we assume with Thomson that the unborn child is as much a person as you or I, would 'withholding treatment' from this child and its subsequent death be justified on the basis that the homeowner was only 'withholding treatment' of a child he did not ask for in order to benefit himself?

> Is any person, born or unborn, obligated to sacrifice his life because his death would benefit another person? Consequently, there is no doubt that such 'withholding' of treatment (and it seems totally false to call ordinary shelter and sustenance 'treatment') is indeed murder.[46]

Again, the argumentation follows solely on an intuitive basis. The case of the baby is placed at the center, although it is not clear how the example might be used as a general analogy for cases in which withholding treatment occurs. It is obvious though that the case of the baby on the doorstep is fit to serve as an analogy for asymmetrical, parent-child relationships of care. It clearly shows that "withholding treatment" is not a proper term to use in the case of pregnancy or parenthood.

## THE VIOLINIST ANALOGY AND PARENTAL RESPONSIBILITY

The differences presented (the issues of consent, family responsibilities, and artificiality) correspond to Hans Jonas' claims concerning parental responsibility. Jonas argues that "the vertical responsibility of parents for children, which in regard to them is not specified, but global (i.e., extending to everything in them that needs caring for), and not occasional, but continual so long as they are children", is natural indeed.[47] This can be said to be natural, not only due to its (assumed) biological foundation, but rather since it is established by the natural role of the different actors: the role of those who provide care and those who receive care cannot be reversed for any period of time. There is also no need for previous consent, since it is not consent that establishes this relationship, as it

would in cases of contractual relationships. It is irrevocable, since parent-child relationships are not established via consent, or by an external public contract, but subsist, even if one of the parties ceases to conform to its role. Finally it is global because it is not limited to certain aspects, as are contractual relationships, but refers rather to the total existence of the child.

From the deficiencies of the analogy results the improper ethical application of the thought experiment. Thomson approaches the question primarily through concepts, which is understandable, since she wants to contest the legal argumentation of pro-life groups. However, if pregnancy is conceived as a parent-child relationship, legal language comes short at this point. Onora O'Neill points out that the language of rights is incapable of grasping the core of parental care:

> Although children cannot plausibly be said to have a right to the cheerful dailiness of family life, to some fun and attention, to some affection and understanding, most people would think that parents have a responsibility, an obligation, to provide a home and atmosphere which provides some (culturally specific) version of all of these for their children, and that parents who do not do so fail in some of their basic obligations to their children.[48]

According to O'Neill, the language of rights obscures the real meaning of being a parent or a child.[49] However sophisticated the violinist analogy and Thomson's argumentation might be, they narrow the horizon through which pregnancy is viewed. Instead of showing it as a fundamental element of the human condition, or as an existential experience, Thomson presents it as a situation that hardly resembles what is normally expe-

rienced as pregnancy: being forced without consent to provide assistance to a stranger by putting one's body at his disposal. From this perspective the Violinist Example is hardly analogous with what we understand by pregnancy and parental responsibility.

## DEMONIC NATURE

Certainly, rape is one possible case that could be seen as analogous to the Violinist Example. Like the kidnapped donor, a woman who has conceived a child by rape is the victim of a crime. But do unwanted pregnancies due to deficient contraception have the same status as conception by rape? Wilcox claims that in the latter case, "to fill out the analogy, we need something like a demonic nature in the background, taking advantage of our innocence, violating our rights against it while we go about our innocent business, maliciously outwitting our cleverest devices protecting ourselves."[50] Nature thus cannot be assumed as an unjust violator, a rapist, or a kidnapper:

> But what, in ordinary pregnancy, is analogous to the Society of Music Lovers? Is it nature? Is the assumption that nature, in her desire to see that the species is propagated, violates our rights by getting innocent women pregnant? Does Thomson presuppose a sort of bitch goddess, a demonic, malevolent, nature? Or a well-wishing but misguided and unjust nature? Surely just such a presupposition is what we need to make the story complete, to make the analogy complete.[51]

The fetus cannot be viewed as an "unjust aggressor" either and cannot in any sense be viewed as evil. Wilcox also rejects

the analogy of the growing child, since fetuses certainly do not have "evil intentions"; in fact, they have no intentions at all.[52]

The relationship between the parents and the child cannot be seen as something evil, but as a connection calling for a positive response, such as caring and love. Although pregnancy might cause serious difficulties on the physiological level, these are mostly interpreted on the social and the existential level where women's biographies are written. This includes unjust structures, which are more responsible for narrowing options for women than a supposedly demonic nature. As Wilcox writes, "if there is a larger-than-individual force at work here, surely it consists of the social forces and structures which continue to oppress women. But talking in ways which presuppose a mythology about an oppressive nature helps but little, for it diverts attention from the real oppression."[53]

## THE MISSING FATHERS

Another deficiency of the analogy is the missing father: "a striking aspect of her essay is that fathers are virtually absent from her analysis".[54] Paternal responsibility seems to be cast aside in the example of the violinist, if it is taken as an analogy for pregnancy, for no one in that example can be identified with the father. The burden of choice is placed solely on the mother's shoulders, although in reality fathers are determining factors in decisions about pregnancies. Wilcox also highlights this aspect when he writes that "though normal sex requires the willing participation of two parties, women are being left with most of the responsibility for the children thus conceived. If we are looking for real injustices behind the plight of pregnant women, nonmythical analogues to the Society of Music Lovers, here is one place which must be examined".[55]

## THE VOLUNTARY DONOR

One of the biggest weaknesses of the violinist analogy is that the donor was kidnapped by the Society of Music Lovers and was thus forced to be connected to the famous violinist. But how would intuitions change if one were to volunteer to help the violinist?

Michael Davis constructed several examples to examine how our intuitions change if a voluntary donor is substituted for a forced one. The first scenario goes as follows:

> Suppose that you are not kidnapped, that you go to the hospital when you hear that (as in the original story) you alone can help the violinist, and that you allow yourself to be plugged into him, fully aware of the risks. Suppose too that after a month or so you change your mind. You are now quite uncomfortable. Your affairs are going from bad to worse. And so on. Would you have a right to unplug yourself and leave? Certainly it would not be nice of you to unplug. You would disappoint the Music Lovers. You would leave the helpless violinist to die. But would you have a right (in Thomson's sense) to do it? It seems to me you would. What right does anyone have to force you to continue helping just because you have already helped him for a month? Simply doing a favour does not oblige you to do more. If, for example, I allow you to enter my house to escape a storm, do I thereby give up my right to send you off when I choose[56]

Although this example might fit the intuition of some people, it is much less convincing than the original Violinist Example, where the rights of the kidnapped donor have certainly been

abused. It is not obvious, though, that this voluntary donor does not have the duty to stay connected to the violinist. It depends certainly on prior agreement, which would also make intuitive responses less equivocal. An altered version of the original scenario does just that:

> This time you answer an ad along with several others equally qualified to help. The violinist is in such a fragile state that he can only bear adaptation to a "donor" once. You are therefore informed that once you have been plugged into the violinist you cannot be unplugged for nine months without killing him. You are chosen by lot from those willing and able to stay nine months. And you allow yourself to be plugged in, fully aware of the risks. In this case, you would not, I think, have the same right to unplug that you had in [the previous case]. This time you are not simply someone who is helping the violinist. You are also someone who has made it impossible for anyone else to help him.[57]

Davis claims that the first is a "case of dependence", while in the second scenario "you filled a position someone else could have filled instead of you and you thereby made it impossible for anyone else to fill it later".[58] There is a marked difference between the two scenarios, since by taking the position of the volunteer you have abrogated all other chances for the violinist to get help. Although the answer to the first case is intuitively uncertain, the answer to the second one is definite. Davis insists that in the latter case "You have no right to unplug."[59] What these examples show is that the momentum of volunteering does not predetermine the nature of the audience's intuitive judgments.

# CONCLUSION

This book began as an attempt to clarify the role of thought experiments in ethics. In the end, however, it turned out to be much more. It has grown into an anthropology and a call for professors and teachers to make bolder use of thought experiments in lecture halls and classrooms.[1]

When reading and contemplating the thought experiments presented in this book, we learn not only about how decisions are evaluated, but also about human nature. Thought experiments help us approach and resolve ethical dilemmas, to be sure, but their biggest benefit is the illumination of the human good. Over the course of the previous eight chapters we have discerned how important intuitive judgements are to human morality and learned much about human nature. The story of Tomoceuszkakatiti and Gyugyu has revealed to us our desire to see ourselves as beings with a conscience, alongside our urge to be free from abjection and misery. Through the Experience Machine Thought Experiment we have explored our demand for reality and our willingness to relinquish certain pleasures in order to lead an authentic life. The Last Man Thought Experiment has shown that we cannot view the world as a valueless entity, but rather assign an intrinsic value to all of creation. This is true even if the world often creates situations where all possible outcomes of our actions are far from ideal, as we have

seen in the numerous versions of the Trolley Scenario. Finally, the Violinist Thought Experiment highlighted the uniqueness of pregnancy, the distinctiveness of the relationship between the mother and the child, and the exceptional role of family ties in care.

Although these insights do not provide a complete and static description of human nature, they bring us closer to answering the anthropological question. They do so by triggering our intuitions, which makes thought experiments an excellent tool for educational purposes. The use of pointed imaginary scenarios facilitates a deeper involvement of students in the learning process, where the emphasis is on experience and reflection over mere information. This is true even if thought experiments have their weaknesses and often prove inadequate in attempts to resolve ethically difficult questions in the real world. They may even be abused for ideological purposes. Still, in the hands of well-trained teachers with morally sound intentions, thought experiments create a space for self-reflection and critical thinking. The challenge they pose to our intuitions might lead not only to rational conclusions but also result in true insight.

Finally, the thought experiments discussed in this book are still challenging and up to date. Abortion, virtual reality, and the pre-programmed decisions of self-driving vehicles are constantly on the public agenda. The Last Man Example, sadly, has also become a realistic future scenario. The use of these imaginary scenarios in classrooms and lecture halls may equip our students for similar future situations or, even more importantly, motivate them to keep these scenarios from becoming a reality.

# NOTES

## PREFACE

**1** I have partially discussed the ideas presented in this book in the following articles: Kovács, Gusztáv: Narrative Ethics in Religious Education, in: Lichner, Milos (ed.): Hope. Where does our Hope lie?, LIT Verlag, Zürich, 2020. 561-567.; Kovács, Gusztáv: Intuíció és erkölcsi nevelés [Intuition and Moral Education],in: Kovács, Gusztáv; Lukács, Ottilia (eds.): Az elbeszélés ereje, Pécsi Püspöki Hittudományi Főiskola, Pécs, 2019, 75-83.; Kovács, Gusztáv: Ami nem hagy nyugodni. Etikai gondolatkísérletek az igehirdetésben, a képernyőn és az előadóban [What Keeps Me Up. Ethical Thought Experiments in Sermons, on the Screen, and in the Lecture Hall], in: Lázár Kovács, Ákos (ed.): Vallás - média - nyilvánosság: Új társadalomtudományi perspektívák, Gondolat Kiadó, Budapest, 2017, 141-159.; Kovács, Gusztáv: Etikai gondolatkísérletek és intuitív ítéletek [Ethical Thought Experiments and Intuitive Judgements], in: Laurinyecz, Mihály (ed.): Erkölcsteológiai tanulmányok 18., Jel Kiadó, Budapest, 2017, 84-99.; Kovács, Gusztáv: Az utolsó ember. Mit mutat meg Richard Routley (Sylvan) gondolatkísérlete? [The Last Man. What does Routley's Thought Experiment Show Us?], in: Kovács, Gusztáv; Vértesi, Lázár (eds.): A teremtés értéke - Az ember méltósága, Pécsi Püspöki Hittudományi Főiskola, Pécs, 2017, 78-87.

**2** The research presented in this volume is connected to the work of the MTA-PPHF Religious Education Research Group. The publication of this book was funded by the Content Pedagogy Research Program of the Hungarian Academy of Sciences.

## Chapter I
## THE STORY IN YOUR HEAD: TOMOCEUSZKAKATITI AND GYUGYU

**1** The film is based on the novel by Ferenc Sánta. (Sánta, Ferenc: Az ötödik pecsét [The Fifth Seal], Szépirodalmi Könyvkiadó, Budapest, 1963., online: https://konyvtar.dia.hu/html/muvek/SANTA/santa00027a_kv.html, Accessed November 15, 2020.

**2** The Arrow Cross Party, in Hungarian the Nyilas Party, was a fascist party led by Ferenc Szálasi. They were in power from October 1944 to March 1945.

**3** All translations from Hungarian and German sources are mine.

**4** Sánta 1963

**5** Fekete Sándor: Az ötödik pecsét, mint etikai irányregény [The Fifth Seal as Ethical Social Novel], in: A Miskolci Egyetem Bölcsészettudományi Kara tudományos diákköri közleményei, Miskolci Egyetem, Miskolc, 2003, 68-73.

**6** Sánta 1963

**7** Ibid.

**8** Ibid.

**9** Ibid.

**10** According to Elemér Hankiss, the paradox of freedom manifests itself in the fact that rules constitute the basic components of games: “This may sound weird, but it is still true that we can create freedom by restricting our original freedom. This self-restriction is an important element of the game. (...) When we create or accept the rules of a

game, we willingly narrow the range of our possible actions. But we simultaneously expand the field of possible combinations, thereby multiplying our freedom. The rules of chess accept only six types of actors on the board, and each of these is allowed only two or, at most, three types of moves. Yet these few rules create the possibility of an infinite number of combinations. If there were no rules, the chess-pieces would simply remain standing on the board, or we might randomly push them to and fro. Just think of the young child who does not know the rules of chess and loses his patience: how quickly he wipes the chess-pieces off the board. In contrast, initiates can enjoy the enthralling freedom of discovering, day by day and in every game newer and newer, unpredictable, surprising combinations on a board consisting of eight times eight squares. This is almost like experimenting with infinity." Hankiss, Elemér: Az emberi kaland. Egy civilizációelmélet vázlata [The Human Adventure: Towards a Theory of Civilization], Helikon, Budapest, 2014, 303-304.

**11** I use the expression "pragmatics" intentionally, since the purpose of TEs is not simply to create understanding, but rather to influence the audience's system of moral beliefs. (A concise summary of pragmatics can be found in: Akmajian, Adrian; Demers, Richard A.; Farmer Ann K.; Harnish Robert M.: Linguistics. An Introduction to Language and Communication, Cambridge, MIT Press, 2010, 363-418.)

## Chapter II
## HOW THOUGHT EXPERIMENTS MOVE US: THE SAMARITAN AND HIS NEIGHBOURS

**1** In certain cases, thought experiments might simply confirm earlier beliefs held by the audience.

**2** The term "horizon" means cognition shaped and limited by presuppositions, life-experiences, historical experiences, and worldviews.

**3** The audience does not have to accept the intuitive judgement without question. They are welcome to object to it and to attempt to refute their intuitions concerning the story. What is important is the (unintentional) emergence of an intuitive judgement.

**4** In case of thought experiments, the epistemological principle is even more striking: the subject as an observer influences the outcome of the experiment by his simple presence.

**5** Cf. Weed, Jennifer Hart: Religious Language, in: Internet Encyclopedia of Philosophy, online: https://iep.utm.edu/rel-lang/, Accessed November 15, 2020.

**6** Parable, in: Merriam-Webster Dictionary, online: http://www.merriam-webster.com/dictionary/parable, Accessed November 15, 2020.

**7** Despite the fact that the allegorical interpretation of the Bible received heavy criticism in the context of the early Reformation, even from Luther himself, it has remained the most important method of interpretation in both the Catholic and the Protestant traditions. Cf. Plummer, Robert L.: Parables in the Gospels: History of Interpretation and Hermeneutical Guidelines, in: SBJT (2009/3), 4-11. The popularity of allegorical interpretation is aptly displayed in John Bunyan's *Pilgrim's Progress*, which was one of the most beloved reads in seventeenth- and eighteenth-century Protestant circles.

**8** The Catechism of the Catholic Church (118) cites the medieval elegiac couplet describing the four layers of meaning in biblical texts, namely the literal, the allegorical, the moral, and the eschatological: "Littera gesta docet, quid credas allegoria, moralis quid agas, quo tendas anagogia." (The Letter speaks of deeds; Allegory to faith; The Moral how to act; Anagogy our destiny.)

**9** Cf. Roukema, Riemer: The Good Samaritan in Ancient Christianity, in: Vigiliae Christianae (2003/1), 56-97.

**10** Cited by Roukema 2003, 62.

**11** Gnilka, Joachim: A Názáreti Jézus. Üzenet és történelem [Jesus of Nazareth: Message and History], Szent István Társulat, Budapest, 2001, 105-106.

**12** Jeremias, Joachim: The Parables of Jesus, SCM Press, Upper Saddle River New Jersey, 1972, 21-22.

**13** In effect the same question is raised by Simon Beck in his article, where he points out that under today's circumstances parables "cannot work as devices for moral or religious instruction in the way they are usually understood to work". Cf. Beck, Simon: Can Parables Work? in: Philosophy and Theology (2011/1), 149-165, 149.

**14** Gendler, Tamar Szabó: Intuition, Imagination, and Philosophical Methodology, Oxford University Press, Oxford, 2010, 56.

**15** Cited by Liebenberg, Jacobus: The Language of the Kingdom and Jesus. Parable, Aphorism, and Metaphor in the Sayings Material Common to the Synoptic Tradition and the Gospel of Thomas, Walter de Gruyter, Berlin, 2001, 53.

**16** Jeremias 1972, 202-203.

**17** Ibid., 203.

**18** Ricoeur, Paul: Listening to the Parables of Jesus, in: Regan, Charles E.; Stewart, David (eds.): The Philosophy of Paul Ricoeur: An Anthology of His Work, Beacon Press, Boston, 1997, 239-245, 245.

**19** Flavius, Josephus: Antiquities of the Jews, 18,30, online: http://perseus.uchicago.edu/perseus-cgi/citequery3.pl?dbname=GreekTexts&query=Joseph.%20AJ%2018.38&getid=1, Accessed November 15, 2020.

**20** Jeremias 1972, 205.

**21** Lebenswelt is often rendered in English as life-world. One of the standard definitions of Lebenswelt comes from Alfred Schütz and Thomas Luckmann: „By the everyday life-world is to be understood that province of reality which the wide awake and normal adult

simply takes for granted in the attitude of common sense." Schütz, Alfred; Luckmann, Thomas: The Structures of the Life-world. Volume 1, Northwestern University Press, Evanston, 1973, 3.

**22** Holyoak, Keith J.: Analogy and Relational Reasoning, in: Holyoak Keith J.; Morrison Robert G. (eds.): The Oxford handbook of thinking and reasoning, Oxford University Press, New York, 234-259, 234.

**23** Ricoeur 1997, 239.

**24** Ibid.

**25** Cf. Festinger, Leon: A Theory of Cognitive Dissonance, Stanford University Press, Stanford, 1962, 1-31.

**26** Ricoeur 1997, 242.

**27** One may rightly think that here we are dealing with a judgment of conscience. (Although judgements of conscience always contain an element of reflection, this is not true for intuition.) Still, it is no coincidence that Jeremias notes the following about the scene preceding the parable: "That a learned theologian should ask a layman about the way to eternal life, was just as unusual then as it would be today. The probable explanation is that the man had been disturbed in conscience by Jesus' preaching." Jeremias 1972, 202.

**28** Llamzon writes the following about the necessity of a disagreement: "I have another philosopher friend who loves to quote a recognized giant in theology to the effect that all we really can be sure of that Jesus said was, "Amen, Amen." It saddens me to say that the decibel count on the laughter that explodes after that joke is delivered provides an ironic measure of the damage to the spirit that the statement has done. What does it do to all our other beliefs in the beautiful teachings of Jesus? It is a crime to rob the hearers of such words the moral inspiration they and every member of the human race, regardless of the religion to which they belong, draws from the parables of the Good Samaritan, of the Prodigal Son, of the widow's mite. So too, in an ethics classroom, when the teacher leaves the impression that "Amen, Amen" is all the best of us in the field can

tell the hearers of our words, it is clear to me that our students have been robbed, wrongfully, of their natural moral energy. Some lovely landscapes of the soul have been polluted." Llamzon, Benjamin S.: A Humane Case for Moral Intuition, Rodopi, Amsterdam, 1993, 27-28.

**29** Jeremias 1972, 205.

**30** Ibid., 205.

## Chapter III
## WHAT MAKES A THOUGHT EXPERIMENT?

**1** Brown, James Robert: Thought Experiments. A Platonic Account, in: Horowitz, Tamara; Massey, Gerald (eds.): Thought Experiments in Science and Philosophy, Rowman & Littlefield, Lanham, 119-128, 122.

**2** Elgin, Catherine Z.: Fiction as Thought Experiment, in: Perspectives on Science (2014/2), 221-241.

**3** Popa, Eugen Octav: Argumentative moves in a thought experiment, in: Cogency (2015/1), 69-89, 70.

**4** Rescher, Nicholas: Thought Experimentation in Presocratic Philosophy, in: Horowitz, Tamara; Massey, Gerald (eds.): Thought Experiments in Science and Philosophy, Rowman & Littlefield, Lanham, 1991, 31-41.
Because of the diverse and opaque definition of thought experiments the scope of the texts that can be identified as thought experiments is considerably broad. Dominik Perler for instance, when discussing the role of angels in medieval thought, states that "discussions about angels often had the status of thought experiments". Cf. Perler, Dominik: Thought Experiments. The Methodological Function of Angels in Late Medieval Epistemology, in: Isabel Iribarren; Markus Lenz (eds.): Angels in Medieval Philosophical Inquiry, Ashgate, Aldershot, 2008, 143–153.

**5** King, Peter: Mediaeval Thought-Experiments: The Metamethodology of Mediaeval Science, in: Horowitz, Tamara; Massey, Gerald (eds.): Thought Experiments in Science and Philosophy, Rowman & Littlefield, Lanham, 43-64.

**6** Witt-Hansen, Johannes: H. C. Orsted, Kant and The Thought Experiment, in: Danish Yearbook of Philosophy (1976), 48–56.

**7** Cohnitz, Daniel: Gedankenexperimente in der Philosophie [Thought Experiments in Philosophy], mentis, Paderborn, 2006, 32.

**8** Though not under the term "thought experiment", the method was popular in academic circles at the turn of the 18th and the 19th century: Alhough "the term 'thought experiment' was not yet in circulation (...) a family of very similar notions was in circulation during this period, and they are embedded in reflections that prima facie seem highly relevant for a philosophical discussion of thought experiments. We should also make mention of the fact that, between 1750 and 1830, the use of experiments became (a) an essential part of university curriculum, (b) popularized in science books written for the laity, and (c) a primary source of evidence for or against scientific theories". Cf. Fehige, Yiftach; Stuart, Michael T.: On the Origins of the Philosophy of Thought Experiments. The Forerun, in: Perspectives on Science (2014/2), 179-220, 180. It must be mentioned that without using the term itself, a number of important authors, such as Georg Lichtenberg (1742-1799), explored fundamental questions concerning thought experiments.

**9** Cohnitz 2006, 53-54.

**10** Kuhn, Thomas S.: A Function for Thought Experiments, in: Kuhn, Thomas S.: The Essential Tension. Selected Studies in Scientific Tradition and Change, University of Chicago Press, Chicago, 1977, 240-265, 263.

**11** Nersessian, Nancy J.: Why do thought experiments work?, in: Proceedings of the Cognitive Science Society Vol. 13, Lawrence Erlbaum Associates, Hillsdale NJ, 1991, 430-438, 431.

**12** It took almost a hundred years to gain empirical proof for Einstein's theory on gravitational waves. Cf. Castelvecchi, Davide; Witze, Alexandra: Einstein's gravitational waves found at last, online: http://www.nature.com/news/einstein-s-gravitational-waves-found-at-last-1.19361, Accessed November 15, 2020.

**13** "We cannot experimentally test much of the physics that is important in the very early universe because we cannot attain the required energies in accelerators on Earth. We have to extrapolate from known physics to the unknown and then test the implications; to do this, we assume some specific features of known lower energy physics are the true key to how things are at higher energies. We cannot experimentally test if we have got it right." Ellis, George F.R.: Issues in the Philosophy of Cosmology, in: Butterfield, Jeremy; Earman, John: Philosophy of Physics, Elsevier, Amsterdam, 1183-1286, 1233.

**14** Brown, James Robert; Fehige, Yiftach: Thought Experiments, in: Edward N. Zalta (ed.): The Stanford Encyclopedia of Philosophy (Spring 2016 Edition), online: http://plato.stanford.edu/archives/spr2016/entries/thought-experiment, Accessed November 15, 2020.

**15** This is in effect on of the basic questions of kantian philosophy: "how are synthetic a priori judgments possible?" (B19) Kant thought that synthetic a priori judgements were also possible in physics.

**16** Roux, Sophie: The Emergence of the Notion of Thought Experiments, in: Ierodiakonou, Katerina; Roux, Sophie: Thought Experiments in Methodological and Historical Contexts, Brill, Leiden, 1-36, 2.

**17** Galilei, Galileo: Dialogues Concerning Two New Sciences. Translated from the Italian and Latin into English by Henry Crew and Alfonso de Salvio, New York, Macmillan, 1914, online: http://oll-resources.s3.amazonaws.com/titles/753/Galileo_0416_EBk_v6.0.pdf, Accessed November 15, 2020. (Page numbers refer to the online version of the text.)

**18** Ibid., 64.

**19** Ibid., 65.

**20** Ibid.

**21** Ibid.

**22** Ibid.

**23** Ibid.

**24** Ibid.

**25** Ibid., 63.

**26** Ibid., 64.

**27** Ibid.

**28** Ibid.

**29** Williams, David R.: The Apollo 15 Hammer-Feather Drop, online: https://nssdc.gsfc.nasa.gov/planetary/lunar/apollo_15_feather_drop.html, Accessed November 15, 2020.

**30** Serge, Michael: Galileo, Viviani and the tower of Pisa, in: Studies in History and Philosophy of Science (1989/4), 435-451, 435. The picture of Galileo throwing iron balls from the leaning tower of Pisa probably comes from his student, Vincenzo Viviani (1622-1703).

**31** In the following I follow Brown's line of argumentation with minor alterations. He mentions "cannon ball" and "musket ball" in his original example. Cf. Brown, James Robert: The Laboratory of the Mind. Thought Experiments in the Natural Sciences, Routledge, London, 2005, 122-123.

**32** There are numerous authors who stronly criticized Galileo's procedure. Cf. Schrenk, Markus: Galileo versus Aristotle on Free Falling Bodies, in: Logical Analysis and History of Philosophy (2004), 81-89.

**33** Our line of thought does not allow a detailed analysis of the current debate about the role of thought experiments in natural sciences. It seems that representatives of an absolute "sceptical view" of thought

experiments in scientific research - a view, which was hallmarked by Pierre Duhem at the beginnig of the 20th century -, constitute only a small minority in the debate. The importance of thought experiments is recognized by the majority: by the less enthusiastic as “illustrations” and means of “persuading people” (Hull), while the more commited claim that “without thought experiments there would be no real experiments” (Buzzoni), and praise their “heuristic” functioning (Galili). Asikainen, Marvi A.; Pekka E. Hirvonen: Thought Experiments in Science and in Science Education, in: Matthews, Michael R. (ed.): International Handbook of Research in History, Philosophy and Science Teaching, Springer, Dordrecht, 2014, 1235-1258, 1238-1239.

**34** There are thought experiments in physics, which we cannot test due to our technological or physical barriers. Other thought experiments contain conditions, which cannot be fulfilled due to logical or other reasons. To the first type belongs Einstein’s thought experiment of riding a beam of light, to the latter ones Newton’s bucket thought experiment concerning absolute space and Schrödinger’s cat. Cf. Brown 2005, 8-11; 15-16; 23-25. Schrödinger, for example, describes his own thought experiment as a “ludicrous case” since he only wanted to demonstrate that laws of quantum physics simply cannot be applied on the macro-level. The experiment does not rely on some specialized knowledge, but rather on common knowledge that a cat cannot be dead and alive at the same time. Cf. Schrödinger, Erwin: Die gegenwärtige Situation in der Quantenmechanik [The Present Situation in Quantum Mechanics], in: Naturwissenschaften (1935/48), 807-812, 812.

**35** Brown; Fehige 2016

**36** Cf. Cooper, Rachel: Thought Experiments, in: Metaphilosophy (2005/3), 328-347.

**37** Searle, John R.: Chinese room argument, in: Scholarpedia (2009/8), 3100, online: http://www.scholarpedia.org/article/Chinese_room_argument, Accessed November 15, 2020.

**38** Ibid.

**39** Leibniz, G.W.: Monadology, (transl. by Jonathan Bennett), 17. online: http://www.earlymoderntexts.com/assets/pdfs/leibniz1714b.pdf, Accessed November 15, 2020.

**40** Ibid.

**41** Ibid.

**42** Ibid.

**43** Turing, A.M.: Computing Machinery and Intelligence, in: Mind (1950/236), 433-460.

**44** Turing 1950, 433.

**45** Searle, John R.: Minds, brains, and programs, in: Behavioral and Brain Sciences (1980/3): 417-457, 417.

**46** For the purpose of simplicity I altered the description at certain minor points. These changes do not affect the core content of the thought experiment.

**47** Searle 1980, 418.

**48** Ibid.

**49** Ibid.

**50** The thought experiment was criticised by Daniel Dennett, Jerry Fodor, and Karl Popper, among others. The website *MindPapers* lists a total of 1887 with the keyword “chinese room” (cf. http://consc.net/mindpapers/search?searchStr=chinese+room&filterMode=keywords, Accessed November 15, 2020.)

**51** Cole, David: The Chinese Room Argument, Edward N. Zalta (ed.): The Stanford Encyclopedia of Philosophy (Spring 2016 Edition), online: http://plato.stanford.edu/archives/win2015/entries/chinese-room, Accessed November 15, 2020.

**52** Nagel, Thomas: What Is It Like to Be a Bat?, in: The Philosophycal Review (1974/4), 435-450, 446.

**53** Nagel 1974, 439.

**54** Jackson, Frank: What Mary Didn't Know, in: The Journal of Philosophy (1986/5), 291-295, 291.

**55** Ibid.

**56** Ibid.

**57** Cf. Popper, Karl: On the Use and Misuse of Imaginary Experiments, Especially in Quantum Theory, in: id.: The Logic of Scientific Discovery, Routledge, London, 2002, 464-480.

**58** Popper 2002, 465.

**59** "We imagine that we take a piece of gold, or some other substance, and cut it into smaller and smaller parts 'until we arrive at parts so small that they cannot be any longer subdivided': a thought experiment used in order to explain 'indivisible atoms'." Ibid.

**60** Ibid., 466.

**61** Brown; Fehige 2016

**62** Ibid.

**63** Cohnitz 2006, 76. (Author's translation.)

**64** Ibid., 77.

**65** Ibid.

**66** Ibid., 79.

**67** Ibid., 80.

**68** Gendler, Tamar Szabó: Thought Experiment. On the Powers and Limits of Imaginary Cases, Routledge, New York, 2013, 21.

**69** Gendler 2013, 21.

**70** Ibid., 25.

**71** Ibid., 26.

**72** Ibid.

**73** Sánta 1963

**74** Gendler 2013, 21.

**75** Gasparatou, Renia: What would you say then?, in: Sorites (2008/2), 63-70, 65.

**76** Ibid.

**77** This is exactly what causes difficulties when a thought experiment is formulated within a horizon different from ours. Does the Parable of the Good Samaritan still function as a thought experiment if there is a need to explain the conflict between the Jews and the Samaritans for the audience to perceive the moral question in the story?

**78** Tooley, Michael: Abortion and Infanticide, in: Philosophy and Public Affairs (1972/2), 37-65.

**79** Tooley 1972, 44.

**80** Ibid.

**81** Tooley 1972, 60.

**82** Ibid., 61.

**83** Gaspara 2008, 66.

**84** The fundamental meaning of analogical speaking is underlined by this ancient Chinese story:

"Someone said to the King of Liang: 'Hui Tzu is very good at using analogies when putting forth his views. If your Majesty could stop him from using analogies he will be at a loss about what to say.'

The King said: 'Very well, I will do that.'

The following day when he received Hui Tzu the King said to him, 'If you have anything to say, I wish you would say it plainly and not resort to analogies.'

Hui Tzu said, 'Suppose there is a man here who does not know what a *tan* is, and you say to him, »a *tan* is like a *tan*,« would he understand?'

The King said, 'No.'

'Then were you to say to him, »A *tan* is like a bow, but has a strip of bamboo in place of the string,« would he understand?'

The King said, 'Yes. He would.'

Hui Tzu said, 'A man who explains necessarily makes intelligible that which is not known by comparing it to what is known. Now Your Majesty says, »Do not use analogies.« This would make the task impossible.'" Cited by Holyoak, Keith James; Thagard Paul: Mental Leaps: Analogy in Creative Thought, MIT Press, Cambridge Mass., 1996, 183.

**85** Brewer, Scott: Exemplary reasoning: semantics pragmatics and the rational force of legal argument by analogy, in: Harvard Law Review (1996/5), 923–1028.

**86** Jackson, M.W.: The Gedankenexperiment method of ethics, in: The Journal of Value Inquiry (1992/26), 525-535, 530-531.

**87** Souder, Lawrence: What Are We to Think about Thought Experiments?, in: Argumentation (2003/2), 203-217, 206.

**88** According to the principle of double effect actions may have good and bad consequences at the same time. One of the classic examples is that of that of ectopic pregnancy when the embryo is removed together with the fallopian tube and thus the life of the mother is saved. While the latter is an intended consequence of the action, the earlier is just a foreseen, but unintended effect. Traditionally there are four conditions to be met for an action to fall under the category of the principle of double effect: "1. The act to be done must be good in itself or at least indifferent. 2. The good intended must not be obtained by means of the evil effect. 3. The evil effect must not be intended for itself, but only permitted. 4. There must be proportionately grave reason for permitting the evil effect." Aulisio, Mark P.: Principle or Doctrine of Double Effect, in: Post, Stephen G. (ed.): Encyclopedia of Bioethics, Third Edition, Gale, Detroit, 2004, 685-690, 687. For a detailed analysis of the principle cf. Kovács Gusztáv: A puding próbája az evés. A kettős hatás elve a valóság mérlegén [The Proof of the Pudding Is in the Eating], in: id., Varga Krisztina, Vértesi

Lázár (eds.): Kettős hatás: Helyettesíthető-e az etika matematikával, Pécsi Püspöki Hittudományi Főiskola, Pécs, 2015, 84-91.

**89** Foot, Philippa: The Problem of Abortion and the Doctrine of the Double Effect, in: id. (ed.): Virtues and

Vices and Other Essays in Moral Philosophy, Clarendon Press, Oxford, 2002, 19–31, 21.

**90** Foot 2002, 21-22.

**91** Ibid., 22.

**92** Williams, Bernard: Utilitarianism and Integrity, in: Glover, Jonathan (ed.): Utilitarianism and its Critics, Macmillan, New York, 1990, 165-169.

**93** Ibid., 166.

**94** Ibid.

**95** Ibid.

**96** Ibid., 169.

**97** McBain, James: Moral Theorizing and intuition Pumps; Or, Should We Worry about People's Everyday Intuitions Regarding Ethical Issues?, in: Midwest Quarterly (2005/3), 268-283, 268.

**98** Weissmahr, Béla: A lélek halhatatlanságának kérdése az emberi értelem fényében [The Question of the Immortality of the Soul in the Light of Human Reason], in: Szolgálat (1981/1), 11-20, 14.

**99** McBain 2005, 269.

**100** Dennett demonstrates the importance of intuition pumps for scientific thinking with the help of an anecdote from the life of János Neumann (John von Neumann): "There is a famous story about John von Neumann, the mathematician and physicist who turned Alan Turing's idea (what we now call a Turing machine) into an actual electronic computer (what we now call a Von Neumann machine, such as your laptop or smart phone). Von Neumann was a virtuoso thinker, legendary for his lightning capacity for doing prodigious

calculations in his head. According to the story—and like most famous stories, this one has many versions—a colleague approached him one day with a puzzle that had two paths to a solution, a laborious, complicated calculation and an elegant, Aha!-type solution. This colleague had a theory: in such a case, mathematicians work out the laborious solution while the (lazier, but smarter) physicists pause and find the quick-and-easy solution. Which solution would von Neumann find? You know the sort of puzzle: Two trains, 100 miles apart, are approaching each other on the same track, one going 30 miles per hour, the other going 20 miles per hour. A bird flying 120 miles per hour starts at train A (when they are 100 miles apart), flies to train B, turns around and flies back to the approaching train A, and so forth, until the two trains collide. How far has the bird flown when the collision occurs? 'Two hundred and forty miles,' von Neumann answered almost instantly. 'Darn,' replied his colleague, 'I predicted you'd do it the hard way, summing the infinite series.' 'Ay!' von Neumann cried in embarrassment, smiting his forehead. 'There's an easy way!' (Hint: How long until the trains collide?)" Dennett, Daniel C.: Intuition Pumps and Other Tools for Thinking, W.W. Norton, New York, 2013, 1.

**101** Brendel, Elke: Intuition Pumps and the Proper Use of Thought Experiments, in: Dialectica (2004/1), 89-108, 90.

**102** Brendel 2004, 90.

**103** Ibid., 97.

**104** Dennett 2013, 12.

**105** Cf. Cohnitz 2006, 157-164.

**106** Darley, John M.; Batson, Daniel C.: "From Jesrusalem to Jericho". A Study of Situational and Dispositional Variables in Helping Behaviour, in: Journal of Personality and Social Psychology (1973/1), 100-108.

**107** Darley; Batson 1973, 104.

**108** Ibid., 105.

**109** Ibid.

**110** Ibid., 107.

**111** Wisnewski, Jeremy J.; Wellman, Justin; Caplandies, Fawn: Torture, Prejudice, and the Principle of Morally Problematic Correlation, online: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjrwMq874buAhVlmIsKHfqnATAQFjAAegQIAhAC&url=https%3A%2F%2Fwww.academia.edu%2F3662374%2FTorture_Prejudice_and_the_Principle_of_Morally_Problematic_Correlation&usg=AOvVaw3J6RULds7ppRnOM-_pbSVw, Accessed November 15, 2020.

**112** Ibid.

**113** Ibid.

**114** Ibid.

**115** Ibid. The authors are right in saying that it was "bad news for many applied ethicists", since ethicists, like most scholars, aim at formulating general statements.

**116** Wogaman, Philip J.: Moral Dilemmas: An Introduction to Christian Ethics, Westminster John Knox Press, Louisville, 2009, 25.

**117** Sulmasy, Daniel: What is conscience and why is respect for it so important?, in: Theoretical Medicine and Bioethics (2008/3), 135-49, 136.

**118** Sulmasy 2008, 137.

**119** Ibid.

**120** Meynell, Letitia: Imagination and insight: a new acount of the content of thought experiments, in: Synthese (2014/17), 4149-4168, 4161-4162.

**121** McBain, James: Moral Theorizing and intuition Pumps. Or, Should We Worry about People's Everyday Intuitions Regarding

Ethical Issues?, in: Midwest Quarterly (2005/3), 268-283, 269.

**122** McBain 2005, 269.

**123** Cf. Gendler, Tamar Szabó: Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrum, in: Midwest Studies In Philosophy (2007/1), 68-89, 81-82.

**124** Gendler 2007, 82.

**125** Ibid.

**126** 2Sam 12:13

**127** I take the concept of self-transcendence from Lonergan, who speaks of three diefferent but interconnected levels of self-tanscendence: intellectual, moral and religious. Cf. Lonergan, Bernard: Self-tanscendence: Intellectual, Moral, Religious, in: Croken, Robert C.; Doran, Robert M.: Collected Works of Bernard Lonergan. Philosophical and Theological Papers 1965-1980, University of Toronto Press, Toronto, 2004, 313-331.

**128** For a distinction between the cognitive and the existential subject cf. Lonergan, Bernard: The Scope of Renewal, in: Croken, Robert C.; Doran, Robert M.: Collected Works of Bernard Lonergan. Philosophical and Theological Papers 1965-1980, University of Toronto Press, Toronto, 2004, 282-298.

**129** Gooding, David C.: What is Experimental about Thought Experiments?, in: Proceedings of the Biannual Meeting of the Philosophy of Science Association. Volume Two. Symposia and Invited Papers, The University of Chicago Press, Chicago, 1992, 280-290.

## Chapter IV
## THOUGHT EXPERIMENTS IN PRACTICAL PHILOSOPHY AND BIOETHICS

**1** Jamieson, Dale: Singer and the Practical Ethics Movement, in: id. (ed.): Singer and His Critics, Blackwell, Oxford, 1999, 1-17.

**2** Jamieson 1999, 3.

**3** Ibid.

**4** Ibid.

**5** Abbott, Philip: Philosophers and the Abortion Question, in: Political Theory (1978/3), 313-335.

**6** Walzer, Michael: World War II: Why Was This War Different?, in: Philosophy & Public Affairs (1971/1), 3-21, 3.

**7** Thomson, Judith Jarvis: A Defense of Abortion, in: Philosophy & Public Affairs (1971/1), 47-66.

**8** Tooley, Michael: Abortion and Infanticide, in: Philosophy & Public Affairs (1972/1), 37-65.

**9** Besides just war theory, abortion was the most discussed topic in the first volumes of the journal. Cf. Wertheimer, Roger: Understanding the Abortion Argument, in: Philosophy & Public Affairs (1971/1), 67-95.; Finnis, John: The Rights and Wrongs of Abortion: A Reply to Judith Thomson, in: Philosophy & Public Affairs (1973/2), 117-145.; Thomson, Judith Jarvis: Rights and Deaths, in: Philosophy & Public Affairs (1973/2), 146-159.; Brody, Baruch: Thomson on Abortion, in: Philosophy & Public Affairs (1972/3), 335-340.

**10** Daube, David: The Linguistics of Suicide, in: Philosophy & Public Affairs (1972/4), 387-437.

**11** Levinson, Sanford: Responsibility for Crimes of War, in: Philosophy & Public Affairs (1973/3), 244-273.; Nagel, Thomas: War and Massacre, in: Philosophy & Public Affairs (1972/2), 123-144.;

Brandt, R. B.: Utilitarianism and the Rules of War, in: Philosophy & Public Affairs (1972/2), 145-165.; Hare, R. M.: Rules of War and Moral Reasoning, in: Philosophy & Public Affairs (1972/2), 166-181.

**12** Malament, David: Selective Conscientious Objection and Gillette Decision, in: Philosophy & Public Affairs (1972/4), 363-386.

**13** Foot 1967

**14** Rachels, James: Active and Passive Euthanasia, in: The New England Journal of Medicine (1975), 78-80.

**15** Harris, John: The survival lottery, in: Philosophy (1975/191), 81-87.

**16** Although leading theologians seemed to ignore the method of thought experiments in their works, there are some exceptions. Cf. Hauerwas, Stanley: Should War be Eliminated? A Thought Experiment, in: Berkman, John; Cartwright, Michael (eds.): The Hauerwas Reader, Duke University Press, Durham, 2001, 392-425.; Hauerwas, Stanley: Gay Friendship. A Thought Experiment in Catholic Moral Theology, in: id. (ed.): Sanctify Them in the Truth: Holiness Exemplified, T&T Clark, London, 1998, 105-122.

**17** Wilson, James: Embracing complexity: theory, cases and the future of bioethics, in: Monash Bioethics Review (2014/1-2), 3-21, 4.

**18** Wilson 2014, 13.

**19** Ibid.

**20** Nemes László: A bioetika három fajtája [The Three Kinds of Ethics], in: Fundamentum (2006/1), 5-22, 7.

**21** Ibid.

**22** Ibid.

**23** Singer, Peter: Famine, Affluence, and Morality, in: Philosophy & Public Affairs (1972/3), 229-243.

**24** Singer 1972, 229.

**25** Ibid.

**26** Ibid., 230.

**27** Singer 1972, 231.

**28** Ibid.

**29** It also inspired a great number of academic and non-academic works, among them Peter Unger's *Living High and Letting Die: Our Illusion* (New York, Oxford University Press, 1996), in which The Drowning Child Scenario was further elaborated and amended by The Vintage Sedan and The Envelope Thought Experiment.

**30** Goodin, Robert E. Political Theory and Public Policy, The University of Chicago Press, Chicago, 1982, 8.

**31** O'Neill, Onora: Lifeboat Earth, in: Philosophy & Public Affairs (1975/3), 273-292.

**32** Goodin 1982, 8.

**33** Ibid., 9.

**34** "Now suppose that Wilt Chamberlain is greatly in demand by basketball teams, being a great gate attraction. (Also suppose contracts run only for a year, with players being free agents.) He signs the following sort of contract with a team: In each home game, twenty-five cents from the price of each ticket of admission goes to him. (...) The season starts, and people cheerfully attend his team's games; they buy their tickets, each time dropping a separate twenty-five cents of their admission price into a special box with Chamberlain's name on it. They are excited about seeing him play; it is worth the total admission price to them. Let us suppose that in one season one million persons attend his home games, and Wilt Chamberlain winds up with $250,000, a much larger sum than the average income and larger even than anyone else has. Is he entitled to this income? Is this new distribution $D_{2,}$ unjust?" Nozick, Robert: Anarchy, State and Utopia, Blackwell, Oxford, 1999, 161-164.

**35** Goodin 1982, 9.

**36** Ibid.

**37** Ibid.

**38** Ibid.

**39** Ibid.

**40** Ibid.

**41** Ibid.

**42** Goodin 1982, 10.

**43** Ibid.

**44** Thomson, Judith Jarvis: Preferential Hiring, in: Philosophy & Public Affairs (1973/4), 364-384.

**45** Goodin 1982, 10.

**46** Ibid., 11

**47** Ross' central question is: "And if we suppose two men dying together alone, do we think that the duty of one to fulfil before he dies a promise he has made to the other would have any effect on the general confidence?" Ross, William David: The Right and the Good, Clarendon Press, Oxford, 2002, 39.

**48** Goodin 1982, 11

**49** Ibid.

**50** Ibid.

**51** Ibid.

**52** Cf. Kovács, Gusztáv: Igazságos gondoskodás. A gondoskodásetika jelentőségéről a szociáletikában[Just Care. On the Importance of Care Ethics in Social Ethics], in: Acta Sociologica (2012/1), 97-104.

**53** Kovács 2012, 98.

**54** Gilligan, Carol: Images of Relationship, in: North Dakota Law Review (2005/4), 693-728, 694.

**55** Cited by Friedman, Marilyn: Abraham, Socrates, and Heinz: Where Are the Women? (Care and Context in Moral Reasoning), in: Harding, Carol Gibb (ed.): Moral Dilemmas and Ethical Reasoning, Transaction Publishers, New Brunswick, 2010, 25-43, 33

**56** Gilligan, Carol: In a Different Voice. Psychological Theory and Women's Development, Harvard University Press, Cambridge, Massachusetts, 1982, 25.

**57** Schnabl, Christa: Feministische Ethik. Profil und Herausforderungen, in: Salzburger Theologische Zeitschrift (2002/2), 269–282, 272.

**58** Gilligan 1982, 26.

**59** Ibid.

**60** Ibid., 28.

**61** Kovács 2012, 99.

**62** Wendel, Saskia: Feministische Ethik zur Einführung [Introduction to Feminist Ethics], Junius, Dresden, 2003, 69-70.

**63** Friedman 2010, 34

## Chapter V
## THE EXPERIENCE MACHINE

**1** Müller, Klaus: Gottes Dasein denken [Thinking God's Existence], Verlag Friedrich Pustet, Regensburg, 2001, 63-83.

**2** Nozick, Robert: Anarchy, State and Utopia, Blackwell, Oxford, 1999.

**3** Nozick 1999, 42-43.

**4** Nozick 1999, 45.

**5** Feldman, Fred: What we learn from the experience machine, in: Bader, Ralf M.; Meadowcroft, John (eds.): The Cambridge Companion to Nozick's Anarchy, State, and Utopia, 2011, 59-86, 59.

**6** The minimal state is "limited to the narrow functions of protection against force, theft, fraud, enforcement of contracts, and so on". Nozick claims that "any other more extensive state will violate persons' rights not to be forced to do certain things (...); and that the minimal state is inspiring as well as right". Nozick 1999, ix.

**7** Ibid., 44.

**8** Ibid.

**9** Ibid., 42.

**10** Ibid., 43.

**11** Feldman 2011, 59.

**12** Ibid., 65

**13** Ibid.

**14** Ibid.

**15** Ibid., 67

**16** Ibid., 67

**17** Silverstein mentions James Griffin, David Brink, Stephen Darwall, and L.W. Sumner. Cf. Silverstein, Matthew: In Defense of Happiness. A Response to the Experience Machine, in: Social Theory and Practice (2000/2), 279-300, 281-282.

**18** Silverstein 2000, 281-282.

**19** Nozick 1999, 42.

**20** Silverstein 2000, 285.

**21** Nozick 1999, 43.

**22** Ibid.

**23** Ibid.

**24** Ibid., 44.

**25** Ibid., 43.

**26** Ibid., 44.

**27** Ibid., 43.

**28** Ibid., 45

**29** Ibid., 43-44

**30** When the experience machine is discussed in university lectures or seminars, there is always a minority who answers the question with a resounding "yes". Although this does not refute the point of the thought experiment, it certainly shows the diversity of intuitive responses.

**31** Silverstein 2000, 297.

**32** Ibid.

**33** Ibid.

**34** Ibid.

**35** Ibid., 298.

**36** Kolber Adam: Mental Statism and the Experience Machine, in: Bard Journal of Social Sciences (1994/3), 13.

**37** Kolber 1994, 13-14.

**38** De Brigard, Felipe: If you like it, does it matter if it's real?, in. Philosophical Psychology (2010/1), 43-57, 50.

**39** Bostrom, Nick; Ord, Toby: The Reversal Test: Eliminating Status Quo Bias in Applied Ethics, in: Ethics (2006), 656–679, 658.

**40** De Brigard 2010, 50.

**41** Kolber 1994, 14.

**42** Ibid.

**43** Ibid., 17.

**44** Cf. Knobe, Joshua; Nichols, Shaun: Experimental Philosophy, in: Zalta, Edward N.: The Stanford Encyclopedia of Philosophy, online:

https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/, Accessed November 15, 2020.

**45** Weijers, Dan: Intuitive Biases in Judgments about Thought Experiments. The Experience Machine Revisited, in: Philosophical Writings (2013/1), 17-31.

**46** Weijers 2013, 23.

**47** De Brigard, Felipe: If you like it, does it matter if it's real?, in: Philosophical Psychology (2010/1), 43-57.

**48** De Brigard 2010, 47.

**49** Ibid.

**50** Ibid.

**51** Ibid., 47-48.

**52** One explanation is the lack of fear regarding malfunction: „The most notable difference is that the risks involved in each case seem markedly different. In the original Experience Machine thought experiment, our intuitive cognition would have deemed a machine life as risky, despite the stipulation in the thought experiment that the machine works perfectly. This intuition of risk likely arises from all of our previous experience with computerized machines crashing at least once, if not regularly, and often do not provide the quality of performance that they promise. In the Trip to Reality thought experiment, the risk of machine failure and machine underperformance are unlikely to affect our intuitive judgment about a life in a machine because that scenario would be matched to our non-crashing real-life experiences during the intuitive processing of the thought experiment." Weijers 2013, 25-26.

**53** De Brigard 2010, 48.

**54** This is supported by the results for the second neutral vignette, which said: "you can either remain connected to this machine (and we'll remove the memories of this conversation taking place) or you

can disconnect. However, you may want to know that your life outside is not at all like the life you have experienced so far." (Ibid., 49.) In this case 59% wanted to stick with their old lives in the machine, and only 41% risked the fundamentally different life.

**55** Ibid., 49-50.

**56** Kawall, Jason: The Experience Machine and Mental State Theories of Wellbeing, in: The Journal of Value Inquiry (1999), 381–387, 383.

**57** Ibid.

**58** Kagan: Normative ethics, Routledge, New York, 1998, 34-35.

**59** Kagan 1998, 36.

**60** Nozick 1999, 43.

**61** Kolber 1994, 15.

**62** Ibid.

**63** Ibid.

**64** Cf. Albert Szent-Györgyi – Biographical, online: https://www.nobelprize.org/prizes/medicine/1937/szent-gyorgyi/biographical/, Accessed November 15, 2020.

**65** Smith, Basil: Can We Test the Experience Machine?, in: Ethical Perspectives (2011/1), 29-51, 35.

**66** Barilan, Michael Y.: Nozick's Experience Machine and palliative care: revisiting hedonism, in: Medicine, Health Care and Philosophy (2009/12), 399–407, 401.

**67** Barilan 2009, 401.

**68** Cassell, Eric J.: Pain, Suffering and the Goals of Medicine, in: Hanson, Mark J.; Callahan, Daniel (eds.): The Goals of Medicine. The Forgotten Issues in Health Care Reform, Georgetown University Press, Washington, D.C., 1999. 101-117.

**69** Cassell 1999, 105.

**70** Ibid.

**71** The question remains whether the legalization of different forms of euthanasia would result in its widespread use or whether it would simply free the patients from their fears, enabling them to handle the end of their lives and their deaths.

**72** Ibid.

**73** Nozick 1999, 43.

**74** Barilan 2009, 402.

**75** Heller, Andreas; Wenzel, Claudia: Hospizarbeit und Palliative Care. Idee, konzeptionelle Grundlagen und Herausforderung für die Psychotherapie [Hospice Work and Palliative Care. Idea, Conceptual Bases and Challenge to Psychotherapy], in: Psychotherapie-Wissenschaft (2011/3), 150-158. (Author's translation.)

**76** Spaemann, Robert: Moralische Grundbegriffe [Basic Moral Concepts], Verlag C. H. Beck, München, 1991.

**77** Spaemann 1991, 25.

**78** Ibid. 27.

**79** Ibid. 30.

**80** Ibid.

**81** Ibid. 31.

**82** Ibid. 32.

**83** Ibid.

**84** Ibid. 32.

**85** Ibid. 33.

**86** Ibid. 34

**87** This statement is also confirmed by Silverstein: "The experience machine thought experiment appeals to our intuitions as evidence against hedonism. Our intuitions tend to reflect our desires and

preferences. In particular, our experience machine intuitions reflect our desire to remain connected to the real world, to track reality. But according to the account of relation between happiness and our desires (...), the desire to track reality owes its hold upon us to the role it has played in the creation of happiness. We acquire our powerful attachment to reality after finding again and again that deception almost always ends in suffering. We develop a desire to track reality because, in almost all cases, the connection to reality is conducive to happiness. Our intuitive views about what is prudentially good, the views upon which the experience machine argument relies, owe their existence to happiness." (Silverstein 2000, 296.)

**88** Spaemann 1991, 7-8

## Chapter VI
## THE LAST MAN ARGUMENT

**1** Richard Routley later changed his name to Sylvan - meaning "from the forest" - to express his commitment to the environment.

**2** Routley, Richard: Is There a Need for a New, an Environmental, Ethic?, in: Proceedings of the XVth World Congress of Philosophy, Sophia Press, Varna, 1973, 205-210.

**3** Routley 1973, 207.

**4** Mathews, Freya: Environmental Philosophy, in: Nick Trakakis; Graham Oppy (eds.): A Companion to Philosophy in Australia and New Zealand, Monash University Publishing, Melbourne, 2010, 543-591, 543.

**5** Meadows, Donella H.; Meadows, Dennis L.; Randers, Jørgen; Behrens III, William W.: The Limits to Growth. A Report for the Club of Rome's Project on the Predicament of Mankind, Universe Books, New York, 1972.

**6** White, Lynn Jr.: The Historical Roots of Our Ecologic Crisis, in: Science (1967/3767), 1203-1207.

**7** White 1967, 1204.

**8** Ibid., 1206

**9** Ibid.

**10** Ibid., 1206-1207

**11** Ibid., 1206

**12** Leopold, Aldo: A Sand County Almanac, Ballantine Books, New York, 1966, 240.

**13** Freyfogle, Eric T.: Land Ethic, in: J. Baird Callicott; Robert Frodeman (eds.): Encyclopedia of Environmental Ethics and Philosophy, Macmillan Reference, Detroit, 2009, 21-26.

**14** Leopold 1966, 239.

**15** Lo, Norva Y.S.; Brennan, Andrew: The Last Man, in: Huss, John: Planet of the Apes and Philosophy. Great Apes Think Alike, Open Court, Chicago, 265-278, 268.

**16** Koenig, Bernie: Natural Law, Science, and the Social Construction of Reality, University Press of America, Dallas, 2004, 65.

**17** Melden, Abraham Irving: The Primacy of Welfare Rights, in: Carl Wellman (ed.): Welfare Rights and Duties of Charity, Routledge, New York, 2002, 115-134, 131.

**18** Melden 2002, 131.

**19** Robinson's figure also motivated philosophers to reformulate such problems as "the private language". Cf. Ben-Tovim, Ron: Robinson Crusoe, Wittgenstein, and The Return to Society, in: Philosophy and Literature (2008/2), 278-292.

**20** Grey, William: Last Man Arguments, in: J. Baird Callicott; Robert Frodeman (eds.): Encyclopedia of Environmental Ethics and Philosophy, Macmillan Reference, Detroit, 2009, 40-41, 40

21 Routley 1973, 207.

**22** Ibid.

**23** Ibid., 210

**24** Carter, Alan: Projectivism and the Last Person Argument, in: American Philosophical Quarterly (2004/1), 51-62, 60.

**25** Ibid.

**26** Zimmerman, Michael J.: Intrinsic vs. Extrinsic Value, in: Zalta, Edward N.: The Stanford Encyclopedia of Philosophy, online: http://plato.stanford.edu/archives/spr2015/entries/value-intrinsic-extrinsic/, Accessed November 15, 2020.

**27** Peterson, Martin; Sandin, Per: The Last Man Argument Revisited, in: Journal of Value Inquiry (2013/1), 121-133, 126

**28** Kant, Immanuel: Groundwork for the Metaphysics of Morals, Yale University Press, New Haven, AK 4:429

**29** Oddie, Graham: Value Realism, in: Hugh LaFollette (ed.), The International Encyclopedia of Ethics, Wiley-Blackwell, 2013, online: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781444367072.wbiee588, Accessed November 15, 2020.

**30** To the best of my knowledge this thought experiment is my original idea.

**31** Routley 1973, 207-208.

**32** Ibid., 208

**33** Ibid.

**34** ibid.

**35** This can be exemplified by the following scenario. Imagine the following situation: The Department of Greek, Etruscan, and Roman Antiquities of the Louvre is cleaned every day. Janitors use huge stepladders to clean the dust and cobwebs from every corner. One day a member of the cleaning staff leaves a stepladder near

the statue known as the “Venus de Milo”. The next day a group of preschool children visits the museum. As they are passing “Venus de Milo” you see that the abandoned stepladder tilts and starts falling slowly in the direction of the children. The only chance to save the children from serious injury is to push the statue between them and the stepladder. (Suppose you are very strong and can move it very quickly.) However, it is also certain that the statue will be irreparably destroyed by the clash. Is it a moral duty to save the health or even the life of the children at the price of the destruction of the irreplaceable piece of art? (To the best of my knowledge this thought experiment is my original idea.)

**36** Routley 1973, 208.

**37** Ibid.

**38** Ibid.

**39** Ibid.

**40** Ibid.

**41** Ibid.

**42** Ibid.

**43** Ibid.

**44** Carter, Alan: Projectivism and the Last Person Argument, in: American Philosophical Quarterly (2004/1), 51-62, 52-53.

**45** Ibid.

**46** Attfield, Robin: The Good of Trees, in: List, Peter C. (ed.): Environmental Ethics and Forestry: A Reader, Temple University Press, Philadelphia, 2000, 98-113, 106.

**47** Peterson; Sandin 2013, 127-128.

**48** Attfield 2000, 106.

**49** Ibid.

**50** Ibid., 105.

**51** Ibid., 106

**52** Ibid.

**53** Peterson, Martin; Sandin, Per: The Last Man Argument Revisited, in: The Journal of Value Inquiry (2013/1-2), 121–133, 128.

**54** Peterson; Sandin 2013, 128.

**55** Ibid., 130.

**56** Ibid., 131-132.

**57** Ibid., 132.

**58** Ibid., 127.

**59** (Moore 1903, Ch. III. §50) Moore, G. E. Principia Ethica, At the University Press, Cambridge, 1922, 83-84.

**60** Zimmerman, Michael J.: Intrinsic vs. Extrinsic Value, in: Zalta, Edward N. (ed.): The Stanford Encyclopedia of Philosophy, online: http://plato.stanford.edu/archives/spr2015/entries/value-intrinsic-extrinsic/, Attfield, Robin: The Good of Trees, in: List, Peter C. (ed.): Environmental Ethics and Forestry: A Reader, Temple University Press, Philadelphia, 2000, 98-113, 106.

**61** Grey, William: Last Man Arguments, in: Callicott, Baird J. (ed.): The Encyclopedia of Environmental Ethics and Philosophy, Macmillan, New York, 2009, 40-41, 41.

**62** Grey 2009, 41.

**63** Ibid.

**64** Carter, Alan: Projectivism and the Last Person Argument, in: American Philosophical Quarterly (2004/1), 51-62, 58.

**65** Carter 2004, 58.

**66** Ibid.

**67** Ibid.

**68** Routley 1973, 207.

**69** Vucetich, John A.; Bruskotter, Jeremy T.; Nelson, Michael Paul: Evaluating whether nature's intrinsic value is an axiom of or anathema to conservation, in: Conservation Biology (2015), 1-12, 2.

**70** Carter 2004, 53.

## Chapter VII
## THE TROLLEY PROBLEM

**1** Cf. Arendt, Hannah: The Origins of Totalitarianism, Meridian Books, New York, 1958.

**2** Cf. Kis, János: Sophie-tól Blairig. Válasz Laurent Sternnek [From Sophie to Blair. A Reply to Laurent Stern], in: HOLMI (2005/8), 954-963.

**3** Kimble, Kevin: Moral dilemmas that matter, in: International Journal of Philosophy (2013/2): 29-37, 30.

**4** Rácsok, Gabriella: Krzysztof Kieślowski Tízparancsolat című filmsorozatának teológiai-etikai elemzése [The Theological-Ethical Analysis of Krzysztof Kieślowski's Decalogue], in: Igazság és Élet (2011/1), 154-178. 7-8.

**5** Edmonds, David: Would You Kill the Fat Man? The Trolley Problem and What Your Answer Tells Us about Right and Wrong, Princeton University Press, Oxford, 2014, 183.

**6** Edmonds 2014, 10.

**7** Ibid., 184.

**8** The names of the different versions are taken from Edmonds 2014.

**9** Foot, Philippa: The Problem of Abortion and the Doctrine of the Double Effect, in: id. (ed.): Virtues and Vices and Other Essays in Moral Philosophy, University of California Press, Berkeley, 1978. 19–31.

**10** Edmonds 2014, 13-25.

**11** Foot 1978, 23.

**12** Ibid.

**13** Ibid.

**14** Ibid.

**15** Ibid.

**16** Ibid.

**17** Ibid., 24.

**18** Ibid.

**19** Ibid.

**20** Ibid., 27.

**21** Ibid., 28.

**22** Anscombe, G. E. M.: Who is Wronged?, in: Oxford Review (1967), 16–17, 16.

**23** Anscombe 1967, 17.

**24** Thomson, Judith Jarvis: Killing, Letting Die, and the Trolley Problem, in: The Monist (1976), 204-217.; Thomson, Judith Jarvis: The Trolley Problem, in: The Yale Law Journal, (1985/6), 1395-1415.

**25** Cf. Thomson, Judith Jarvis: Self-Defense, in: Philosophy and Public Affairs (1991/4), 283-310.

**26** Thomson 1976, 204.

**27** Ibid., 205.

**28** Ibid., 206.

**29** Ibid.

**30** Ibid., 207.

**31** Ibid.

**32** Ibid.

**33** Ibid., 207-208.

**34** Ibid., 215.

**35** Ibid., 216.

**36** Gorr, Michael: Thomson and the Trolley Problem, in: Philosophical Studies (1990/1), 91–100, 93.

**37** Gorr 1990, 95.

**38** "Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means, but always at the same time as an end" (Gr. 429).

**39** Thomson 1976, 217.

**40** Thomson 1976, 210.

**41** Cf. online: https://pics.onsizzle.com/veil-of-ignorance-trolley-problem-you-you-you-you-you-10553553.png, Accessed November 15, 2020.

**42** Edmonds 2014, 186.

**43** Kamm, Frances M.: Intricate Ethics. Rights, Responsibilities, and Permissable Harm, OUP, Oxford, 2007, 95.

**44** Kamm 2007, 95.

**45** Ibid.

**46** Ibid., 118.

**47** Otsuka, Michael: Double Effect, Triple Effect and the Trolley Problem. Squaring the Circle in Looping Cases, in: Utilitas (2008/1), 92–110., 109.

**48** Otsuka 2008, 109.

**49** Unger, Peter: Living High and Letting Die. Our Illusion of Innocence, Oxford University Press, New York, 1996, 87.

**50** Unger 1996, 88.

**51** Ibid., 90.

**52** Ibid.

**53** Ibid.

**54** Ibid.

**55** Ibid., 92.

**56** Ibid., 136.

**57** Ibid.

**58** Ibid., 136-137.

**59** Singer, Peter: Living High and Letting Die, in: Philosophy and Phenomenological Research (1999/1), 183-187, 184.

**60** Singer 1999, 185.

**61** Ibid.

**62** Schirach, Ferdinand von: Terror. Ein Theaterstück und eine Rede [Terror. A Theater Play and a Speech], Piper Verlag, München, 2015.

**63** Edmonds 2014, 31-32.

**64** Ibid., 32.

**65** Singer, Peter: Ethics and Intuitions, in: The Journal of Ethics (2005/3-4), 331-52, 340.

**66** Singer 2005, 332.

**67** Ibid., 344.

**68** Ibid., 336.

**69** Ibid., 343.

**70** Ibid.

**71** Ibid., 344-345.

**72** Ibid., 345.

**73** Ibid., 345-346.

**74** Haidt, Jonathan: The Emotional Dog and Its Rational Tail. A Social Intuitionist Approach to Moral Judgment, in: Psychological Review (2001/4), pp. 814–834, 814.

**75** Haidt 2001, 815.

**76** Singer 2005, 348.

**77** Ibid.

**78** Graham, Jesse; Haidt, Jonathan; Rimm-Kaufman, Sara E.: Ideology and Intuition in Moral Education, in: 2008 European Journal of Developmental Science, (2008/3), 269-286, 270.

**79** Hogarth, Robin M.: Educating Intuition. A Challenge for the 21st Century, The University of Chicago Press, Chicago, 2001.

**80** Singer 2005, 350.

**81** Ibid.

**82** Unger 1996, 10.

**83** Ibid., 11.

**84** Ibid.

**85** Unger lists the approches of Frances M. Kamm and Judith J. Thomson as preservationists.

**86** Unger 1996, 11.

**87** Nagel, Jonas; Wiegmann, Alex: Moral Intuitionism and Empirical Data, in: Brand, Cordula (ed.): Dual-Process Theories in Moral Psychology. Interdisciplinary approaches to theoretical, empirical and practical considerations, Springer Press, Wiesbaden, 2016, 185-206, 190.

**88** Unger, Peter: Identity, Consciousness and Value, Oxford University Press, Oxford, 1990, 88-89.

**89** Bonnefon, Jean-François; Shariff, Azim; Rahwan, Iyad: The Social Dilemma of Autonomous Vehicles, in: Science (2016), 1573-1576, 1573.

**90** Cf. online: http://moralmachine.mit.edu/, Accessed November 15, 2020.

**91** Bonnefon et. al. 2016, 1573.

**92** Ibid., 1575.

**93** Keenan, James F.: Moral Wisdom: Lessons and Texts from the Catholic Tradition, Rowman&Littlefield Publishers, Lanham, 2004, 151.

**94** Keenan 2004, 151.

**95** Ibid., 151

**96** Ibid.

**97** Kis 2005, 957.

**98** Kaveny, Cathleen M.: Conjoined Twins and Catholic Moral Analysis. Extraordinary Means and Casuistical Consistency, in: Kennedy Institute of Ethics Journal (2002/2), 115-140, 120.

**99** Kaveny 2002, 120.

**100** Ibid.

**101** Bratton, M.Q.; Chetwynd, S.B.: One into two will not go: conceptualising conjoined twins, in: Journal of Medical Ethics (2004/3), 279-85. 279.

**102** Claudia Wiesemann, Claudia: Von der Verantwortung, ein Kind zu bekommen. Eine Ethik der Elternschaft [About the Responsibility of Having a Child. An Ethics of Parenthood], Beck, München, 2006, 93.

**103** Wiesemann 2006, 93.

**104** Ibid.

**105** Cf. Hankiss, Elemér: Az emberi kaland [The Human Adventure], Helikon, Budapest, 2014.

**106** McConnell, Terrance: Moral Dilemmas, in: Edward N. Zalta (ed.): The Stanford Encyclopedia of Philosophy, online: http://plato.stanford.edu/archives/fall2014/entries/moral-dilemmas/, Accessed November 15, 2020.

**107** One of my students gave the following answer to the dilemma: "I would turn the switch, but my life would be ruined anyway."

**108** Wiesemann 2006, 40.

**109** Cf. Gilligan, Carol: In a Different Voice, Harvard University Press, Cambridge, 1982.

## Chapter VIII
## THE VIOLINIST ANALOGY

**1** Balkin, Jack M.: Roe v. Wade. An Engine of Controversy, in: Id. (ed.): What Roe v. Wade Should Have Said. The Nation's Top Legal Experts Rewrite America's Most Controversial Decision, New York University Press, New York, 2005, 3-27, 3.

**2** Wertheimer, Roger: Understanding the Abortion Argument, in: Philosophy & Public Affairs (1971/1), 67-95

**3** Tooley, Michael: Abortion and Infanticide, in: Philosophy & Public Affairs (1972/1), 37-65, 37.

**4** Finnis, John: The Rights and Wrongs of Abortion: A Reply to Judith Thomson, in: Philosophy & Public Affairs (1973/2), 117-145.

**5** Parent, William: Editor's introduction, in: Judith Jarvis Thomson: Rights, Restitution, and Risk, Harvard University Press, Cambridge MA, 1986, vii – x, vii.

**6** Davis, N Ann: Fiddling second: reflections on "A defense of abortion", in: Fact and value : essays on ethics and metaphysics for Judith Jarvis Thomson, MIT Press, Cambridge, 2001, 81-96, 81.

**7** Davis 2001, 87.

**8** Thomson, Judith Jarvis: A Defense of Abortion, in: Philosophy and Public Affairs, (1971/1), 47-66.

**9** Thomson 1971, 48.

**10** Ibid., 48.

**11** Ibid., 49-50.

**12** Ibid., 49.

**13** Ibid.

**14** Ibid., 48.

**15** Ibid., 66.

**16** Ibid., 48.

**17** Ibid., 52-53.

**18** Ibid., 53.

**19** Ibid., 53-54

**20** Ibid., 54.

**21** Ibid., 55.

**22** Ibid.

**23** Ibid., 59.

**24** Ibid.

**25** Ibid., 59.

**26** Huneman, Philippe: Preformation and Epigenesis, in: Werner Dubitzky; Olaf Wolkenhauer; Kwang-Hyun Cho; Hiroki Yokota (eds.): Encyclopedia of Systems Biology, Springer, New York, 2013, 1734-1735.

**27** Thomson 1971, 60.

**28** Wilcox, John T. Nature As Demonic in Thomson's Defense of Abortion, in: The New Scholasticism (1989/4), 463-484, 259.

**29** Finnis 1973

**30** Wiland, Eric: Unconscious violinists and the use of analogies in moral argument, in: Journal of Medical Ethics, (2001/6), 466–468.

**31** Singer, Peter: Practical Ethics, Cambridge University Press, Cambridge, 1993, 148.

**32** Singer 1993, 149.

**33** Cf. Weinberg, Rivka: The moral complexity of sperm donation, in: Bioethics (2008/3), 166–178.

**34** Beckwith, Francis J.: Personal Bodily Rights, Abortion, and Unplugging the Violinist, in: International Philosophical Quarterly (1992/1), 105-118, 111.

**35** Beckwith 1992, 111.

**36** Ibid., 112.

**37** Ibid.

**38** Ibid.

**39** Wiland 2001, 467.

**40** Ibid.

**41** Wilcox 1989, 260.

**42** Beckwith 1992, 113.

**43** Ibid.

**44** Ibid.

**45** Ibid.

**46** Ibid., 115.

**47** Jonas, Hans: The Imperative of Responsibility, University of Chicago Press, Chicago, 1985, 95.

**48** O'Neill, Onora: The "Good Enough" Parent in the Age of the New Reproductive Technologies, in: Hille Haker; Deryck Beyleveld (eds.): The Ethics of Genetics in Human Procreation, Ashgate Publishing, Aldershot, 2000, 33–48, 37.

**49** O'Neill, Onora: Children's Rights and Children's Lives, in: Ethics (1988/3), 445–463, 460.

**50** Wilcox 1989, 266.

**51** Ibid.

**52** Ibid., 267.

**53** Ibid., 268.

**54** Ibid.

**55** Ibid., 269.

**56** Davis, Michael: Foetuses, Famous Violinists, and the Right to Continued Aid, in: The Philosophical Quarterly (1983/3), Vol. 33, 259-278, 263.

**57** Davis 1983, 264.

**58** Ibid., 265.

**59** Ibid.

## CONCLUSION

**1** Both are perfectly exemplified by Tamar Szabó Gendler's 2011 lecture series on *Philosophy and the Science of Human Nature* at Yale University. Online: https://oyc.yale.edu/philosophy/phil-181, Accessed November 15, 2020.