

It's Not the Thought that Counts

*An Essay on the Irrelevance of Intentions to the Moral
Permissibility of Actions*

by

ANTON MARKOČ

Submitted to

Central European University

Department of Philosophy

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Philosophy

Supervisor:

Prof. János Kis

Budapest

2016

Copyright notice

The dissertation contains no materials accepted for any other degrees in any other institutions.

The dissertation contains no materials previously written and/or published by another person, except where appropriate acknowledgment is made in the form of bibliographical reference.

Anton Markoč

Abstract

The dissertation is a defense of the view that intentions with which we act are non-derivatively irrelevant to the moral permissibility of our actions. What we intend might indicate what kind of people we are, or whether and why we should be open to praise or blame, but it is never a reason why we are morally permitted or forbidden to act. The Doctrines of Double and Triple Effect and similar moral principles are rejected. The most prominent arguments in favor of the irrelevance of intentions—those pointing to a category mistake, the impossibility of choosing intentions, or to the oddness of considering them in decision making or of intervening in actions of others due to them—are also shown to be flawed. The conclusion reached is that the strongest evidence for the irrelevance of intentions is negative, based on the failure of case-based, principle-driven, and theoretical justifications to the contrary.

Acknowledgements

I owe the highest debt of gratitude to János Kis, who supervised my doctoral research and whose advice, guidance, and encouragement often went beyond the call of his professional and moral duties, and to Tim Scanlon, who was very generous to invite me to work with him at Harvard, and whose inspiring book *Moral Dimensions* first got me interested in the relevance of intentions to permissibility. I have learned from János and Tim much more than I can acknowledge. They had the profoundest impact on my philosophical development.

I was fortunate enough to have some perceptive people reading and commenting on the manuscript at different stages of its production. Besides János and Tim, I am especially grateful to the dissertation's examiners, Victor Tadros and Simon Rippon, whose insightful criticisms lead to significant improvements in the final version. I have received instructive comments on early drafts of different chapters from Hanoah Ben-Yami, Selim Berker, Ulrike Heuer, Ivan Milić, and Edi Pavlović. I am very much indebted to each of them.

I have also benefited from conversations with these and many other people. I am surely forgetting some who should not be forgotten and I apologize to them. Those I remember talking to include Damir Čičić, Adam Hosein, Ferenc Huoranzski, Yuliya Kanygina, Jeff McMahan, Christopher Megone, Zoltán Miklósi, Andrés Moles, Attila Mráz, Tom Parr, Howard Robinson,

Adam Slavny, David Weberman, and Miklós Zala. Special thanks are owed to Frances Kamm, for lengthy and amiable discussion on trolleyology and ethical intricacies, and to Derek Parfit, for drawing my attention to arguments and distinctions I would have never thought about myself and for sending me his unpublished manuscripts that helped me to clarify and develop my views. The influence of Frances's and Derek's work on my ideas will be visible to any informed reader.

My grammar, orthography, and style would be much worse were it not for invaluable help of Tom Rooney. I also appreciate the help of my sister, Marija Markoč, in double-checking the bibliography and for always being keen to pump her intuitions with my thought experiments.

Work on this dissertation was supported by CEU's Doctoral Scholarship, Write-Up Grant, and two academic achievements awards I had received. My research at Harvard was supported by CEU's Doctoral Research Support Grant. Ministry of Science of Montenegro, Municipality of Bar, and several private companies (DIBIEM, CG Control) provided generous additional support for that research. I am also grateful to the Rončevićs for their hospitality and help in Boston.

CEU, and its Philosophy Department, have been the most congenial environment. I have spent seven wonderful years there. My appreciation goes to all the people who made it possible.

My greatest support have always been my friends and family. I am grateful to my friends, too many to be mentioned here, for tolerating my frequent absences from common events and for being mindful of my research. My parents, Nataša and Savo Markoč, unconditionally supported every decision in my academic career. Even at tough times, they have not failed in love and in moral and financial support. This dissertation is dedicated to them.

Finally, I thank my fiancé, Maryna Batsman, for all her love and forbearance. Having her in my life is a blessing.

Contents

Abstract	iii
Acknowledgements	iv

Introduction	1
1. The Question	1
2. An Overview	10
3. The Method	12

PART ONE

Contenders to the Irrelevance Thesis

1. Double Effect Defects I: Interpretation	21
Introduction	21
1. Why the DDE Matters	22
2. Varieties of the DDE	28
2.1 The Classical Definition	28
2.2 Permissibility and Blameworthiness	30
2.3 Absolutist and Non-Absolutist DDE	32
2.4 The DDE and Moral Absolutism	33
2.5 Ends, Means, and Side Effects	34
2.6 Intention and Mere Foresight	35
2.7 The Counterfactual Test	38
2.8 Intention and Motive	43
2.9 Intending Bad Effects	46
2.10 Intending Good Effects	49
2.11 Subjective and Objective DDE	50
2.12 Due Proportion	54
2.13 Actualist and Possibilist DDE	57
Conclusion	58

2. Double Effect Defects II: Critique	59
--	-----------

Introduction	59
1. The Strength-of-Reasons Objection	60
1.1 The Possibility of Reasons against Intending Harm	60
1.2 The Nature of Reasons against Intending Harm	63
1.3 The Strength of Reasons against Intending Harm	65
1.4 Objections and Replies	76
2. The Closeness Objection	79
2.1 The Relational Views	80
2.2 The Parity Views	86
3. The Explanatory Objection	89
3.1 The Method of Isolation	90
3.2 Explaining the Hard Cases	94
Conclusion	100
Appendix: Triple Effect Defects	101
 3. Intentions and Overdetermined Wrongdoing	 106
Introduction	106
1. Overdetermination and Harm: Conceptual Issues	108
2. The Case-Based Argument	112
2.1 Overdetermination	114
2.2 Overdetermination under Duress	120
3. The Counterfactual Alternative	123
3.1 How the CV Explains the Cases	123
3.2 Problems with the CV	126
3.3 A Rationale for the CV: The Normative Reasons View	129
4. The Argument from Liability	134
Conclusion	137
 4. Intentions and Inchoate Wrongdoing	 138
Introduction	138
1. The Intentions View	139
1.1 The No-Intention Argument	139
1.2 The Comparative Argument	144
2. The Risk-Based View	146
3. Scanlon on Inchoate Wrongdoing	148
3.1 The Facilitation View	148
3.2 Rescuing the Facilitation View	152
3.3 Objections and Replies	155
Conclusion	157

PART TWO

Defenses of the Irrelevance Thesis

5. Intentions and Permissibility: A Confusion of Moral Categories?	159
Introduction	159
1. The Inductive Strategy	161
2. The Deductive Strategy: Thomson	169
3. The Deductive Strategy: Scanlon	172
Conclusion	176
 6. Intentions, Permissibility, and Choice	 178
Introduction	178
1. Permissibility and the Ought Implies Can	180
1.1 ‘Ought’ and ‘Can’	180
1.2 Alternative Possibilities and Thick Act-Types	182
1.3 Objections to P1	185
2. Choosing Reasons for Action	189
2.1 Why We Cannot Believe at Will	190
2.2 An Important Equivocation	191
2.3 Why We Can Choose to Act for a Reason	198
3. How the AFC Proves Too Much	200
Conclusion	203
 7. Bad Intentions and Third Party Intervention	 204
Introduction	204
1. The First Horn	207
1.1 Three Views on the Wrongness of Intending Harm	207
1.2 The Strength of Reasons to Prevent Intending Harm	214
1.3 The Strength of Reasons not to Encourage Intending Harm	218
2. The Second Horn	220
2.1 The Question-Begging Views	220
2.2 Encouraging the Impermissible	222
Conclusion	225
 8. Considering Intentions in Decision Making: What is so Odd about it?	 226
Introduction	226
1. Doing the Unordinary	229

1.1 The Fallacy	229
1.2 The Facts	230
2. Doing the Impossible	232
2.1 Cartesian Psychology	233
2.2 Intending at Will	236
2.3 Act-Types and Possible Intentions	238
2.4 Actualism and Possibilism	241
2.5 The Question-Begging Charges	242
3. Doing the Unnecessary	244
3.1 Moral Self-Indulgence	245
3.2 Tracing the Oddness	247
Conclusion	250

PART THREE

Towards the Irrelevance Thesis

9. Bad Intentions as Wrongs Per Se	253
Introduction	253
1. Agent-Centered Views	255
1.1 Agency	255
1.2 Moral Character and Integrity	256
1.3 Guidance by Evil	259
2. Victim-Centered Views	267
2.1 Respect	267
2.2 Inviolability	270
2.3 Autonomy	272
2.4 Using as a Means	274
3. Impersonal Views	279
Conclusion	282
Conclusion	283
1. The Argument in Brief	283
2. An Error Theory	285
3. Implications	289
Bibliography	293

Introduction

1. The Question

Is it the thought that counts? This is one of the big questions. It is comparable in its sublimity to questions such as “Do we have free will?”, “Is there life after death?”, and “Why anything?”. As with any such question, we must be cautious. Most people are intuitively drawn to some answer. But we cannot give plausible answers unless we understand what we are asked. Since the big questions are also the ones that were debated for centuries with little, if any, progress, it is not unreasonable to suppose that the perplexity is in part due to their ambiguity. To understand the question better, several clarifications are in order.

I shall start from its end. What is meant by the thought counting? It is meant mattering, or being relevant, or important. But mattering for what and in which sense? A common answer is: “For an action’s rightness or wrongness, and fundamentally”. This raises two further questions. First, what is meant by rightness and wrongness here? Second, what does it mean for something to matter fundamentally? I shall take these questions in turn.

We are concerned with moral rightness and wrongness, not legal, or those of etiquette or other normative domains.¹ We are also concerned with a particular sense of moral rightness and wrongness. Although there are many senses of those concepts, some of which are more basic than others (see Parfit 2011a: 164-74), three are especially relevant to our question.

The first and the most commonly assumed sense is of moral obligation. That an action is wrong in this sense means that, morally, it must not be done, i.e., that it is prohibited and that one is obliged or has a duty not to do it. That an action is right means that it is not wrong in this sense, which entails either that it may be done (that it is permissible) or that it must be done (that it is obligatory or one's duty).

The second sense is of moral blameworthiness. That an action is wrong in this sense means that it is the one for which an agent can be legitimately blamed (that it is blameworthy).² That it is right means the negation of it, which implies that it is the action for which an agent can be neither blamed nor praised or the action for which he can be praised (that it is praiseworthy).

The third sense is of moral character. That an action is wrong in this sense means that it exhibits an agent's negative character trait, being, say, callous or deceitful. An action's rightness means the presence of a positive character trait, being, say, compassionate or temperate.

Each of these senses have themselves multiple senses, some of which are definable and some of which are not. That an action is impermissible, for example, might mean that there is, morally, a decisive reason against it being performed (Parfit 2011a: 167). And that an action is

¹ A further assumption is that it makes sense to speak in these quasi-legalistic terms. Some people (e.g. Anscombe 1981: Ch. 4) believe that moral rightness and wrongness are useless notions unless the divine command theory of morality is presupposed while others (e.g. Williams 2006: Ch. 10) reject them because they presumably lead to a too narrow conception of what we have most reason to do, leaving a crucial part of what matters in life aside. I believe that these objections can be answered but I shall not attempt to do it here.

² On some views (Graham 2014), we are blameworthy only for motives and intentions. Those views are, I believe, mistaken. Part of the confusion is due to not taking into account act-descriptions with intention or motive.

blameworthy might mean that some reactive attitude, such as resentment, anger, or indifference, is appropriate towards the agent (Strawson 1962). We can leave these subtleties aside for now.

There are different views on how these senses relate to each other. On one view, they are completely logically independent: neither coextensive nor mutually definable. An action can be impermissible but not blameworthy or show no defect in one's character. And an action is not impermissible because it is blameworthy nor vice versa, and so on. On other views, either some of these senses are coextensive or some are defined by others or both.

We can now see the complexity of the concepts 'right' and 'wrong'. Asking whether it is the thought that counts, we cannot legitimately ask whether the thought matters for rightness or wrongness without explaining what we mean by these. This does not yet answer which sense is relevant to our question. I shall postpone that answer for a moment, turning to the fundamentality of mattering.

Some consideration can matter derivatively or non-derivatively. Mattering fundamentally is a synonym of the latter. There are several ways of derivative mattering. A consideration can signify a future occurrence of a derivative or a non-derivative consideration, such as when one's vicious character indicates the risk of harming other people in future. A consideration can also be caused by, supervene, or be grounded in some other derivative or non-derivative consideration. By grounding, I mean an explanatory relation ordinarily spelled out by locutions 'because', 'in virtue of', 'makes the case'. That relation is distinct from causation and supervenience. Causing is not necessary for grounding as some explanations are not causal, whereas supervening is not sufficient for it as supervenience is not an explanatory relation but a relation of mere covariance (see Fine 2012, Berker 2015).

Since derivative considerations can ground other derivative considerations, what makes non-derivative ones special is that they are, so to say, ungrounded grounders, that is, the ultimate or the fundamental explanations of some phenomenon – in our case, of an action's rightness or wrongness. Hence the fundamentality talk.

Some moral theories are monist in that they posit only one non-derivative consideration. For example, hedonistic utilitarianism is the view that an action is right just when and because it is felicific, that is, just when and because it produces at least as much overall balance of pleasure over pain as any alternative. The fact that an action is felicific, according to this view, is the only fundamental explanation of an action being right. Many other moral theories are pluralist in this respect.

As any other moral question, the question that concerns us here can be answered inside or outside a moral theory. If we are hedonistic utilitarians, we shall deny that the thought counts in a non-derivative way as we shall believe that, non-derivatively, only the balance of pleasure over pain matters. Some other theory might reach an opposite answer. We can then defend an answer to our question by defending a particular moral theory. A more modest but also more promising strategy is to try to answer that question not assuming the truth of any moral theory and then see what, if anything, follows from it about the truth of some theory.

We can thus assume that the thought *might* matter fundamentally. But what do we mean by the thought? We might mean the belief that an action will have certain effects or be done in a certain way. On some views, it is only such facts, which can be called morally relevant, whether or not the agent believed or could have believed that they are true, that make it the case that an action is right or wrong (Moore 2005: Ch. 5, Thomson 1986: Ch. 11, 1990, Graham 2010). On

other views, rightness or wrongness of an action depend on what an agent believed, or justifiably believed, morally relevant facts to be, even if the belief turns out to be false (Prichard 2002: Ch. 6, Scanlon 2008: 47-52, Zimmerman 2008). And there are views which state that both beliefs and morally relevant facts matter but that they matter for different senses of an action's rightness or wrongness (Gibbard 2005, Parfit 2011a: Ch. 7).

When some people ask our question, they have in mind this debate. They are puzzled by the questions such as whether a person who sticks the pins in a voodoo doll of his hated enemy, falsely and unjustifiably believing that the enemy will be thereby killed, acts wrongly or whether a doctor who falsely but justifiably believes that some treatment will benefit others acts wrongly if the treatment unexpectedly kills those people.

The best way to answer such questions, I believe, is to go for the ecumenical strategy and distinguish between different senses of wrongness. We might then say that the former agent acts permissibly but blameworthy while the latter acts blamelessly while permissibly, in the sense relative to the evidence, and impermissibly, in the sense relative to the facts. We might add that, since we cannot act on what is in fact the case if we could not believe that that is the case, given some reasonably demanding standard of belief formation, the latter sense of permissibility has little practical, action-guiding, use except for the purpose of asking, hypothetically, which moral principles are true in order to try to follow them (cf. Scanlon 2008: 51-2, Parfit 2011a: 161-2). We might also add that, while the voodooist has a moral character flaw, the doctor does not.

The other thing some people mean when asking whether the thought counts is whether the belief that an action is right or wrong, whether or not that belief is true or justified, is relevant for that action's rightness or wrongness. On some views, the answer is affirmative, as conscience

is always and decisively binding. On other views, believing that an action is right is insufficient to make it right but believing that it is wrong is sufficient to make it wrong (Aquinas 1981: I-II, 1a 2 ae, q. 6 a. 8, q. 19 a.6, Foot 2002b: Ch. 2). So while a conscientious Nazi, who believes that killing Jews is his moral duty, does not act rightly when he kills them, the person who sticks the pins in a voodoo doll of his hated enemy, believing that he acts wrongly, does act wrongly.

Because these views might be easier to defend concerning an action's blameworthiness or viciousness than its permissibility, it pays to revise them taking into account different senses of wrongness. However, there are cases where the belief that an action is wrong does not make it wrong even in the former senses, such as when an agent justifiably benefits others but falsely and unjustifiably believes that he acts wrongly. When Huckleberry Finn helped Jim to escape slavery, he believed he will go to Hell, but that belief did not make his praiseworthy action blameworthy, nor did acting on it reveal a flaw of his moral character (Parfit 2011a: 158, cf. Bennett 1974).

The same might not be true about the voodooist. His action might still be blameworthy and show a character defect because he believed it to be wrong. But it is less obvious that he acts impermissibly because of the same reason. This is especially contentious when the belief that an action is wrong is not accompanied by the belief that the sufficient reason for acting is to harm the victim or do some other bad thing (as in the voodoo case). It is, for example, hard to believe that Catholics who consensually use artificial birth control act impermissibly just because they believe that using it is impermissible but cannot resist the temptation. Such actions might be hypocritical but it is far from obvious that they are impermissible.

Nonetheless, such actions might still be impermissible because the connection between permissibility and blameworthiness or moral character might be tight or because beliefs that an

action is wrong do fundamentally affect permissibility in a way which, though not very intuitive, has a sound theoretical justification.

My aim in the dissertation, however, is not to resolve this but a similar and more pressing problem. I have just said that the belief that there is a sufficient reason to harm someone might confound the assessment of the relevance of the belief that an action is wrong for its wrongness. I said it because the former and not the latter belief might be what makes an action wrong. That former belief is often a motivating reason for acting. Therefore, when asking whether it is the thought that counts, what we might be asking is whether motives are relevant for the wrongness of actions. Moreover, since motivating reasons, when they are conclusive, give rise to intentions to act (cases of weakness of will aside), we might be asking whether an intention to harm or to do other bad things is relevant for the wrongness of an action it leads to. We might similarly ask whether the belief that benefiting others is a reason for acting, and a corresponding intention to benefit, count for an action's rightness. If the former motives or intentions are bad, and the latter are good, our question reduces to asking whether bad motives or intentions count for wrongness of actions and whether the good ones count for their rightness.

Most people in fact have these in mind when asking whether the thought counts. Rather than wanting to know whether the voodoo 'killer' acts wrongly because he believes he does or because he believes he will thereby kill his enemy, they want to know whether he acts wrongly because he intends to kill his enemy out of hatred. Such questions might also arise confronted with cases of otherwise permissible harming, as when a doctor kills a terminally ill patient at his request intending only to kill him, moved by sadistic urge, revenge, or his professional duty, and

even with cases of permissible benefiting of others, as when one saves a drowning child seeking only praise and fame.

We can now get back to the different senses of wrongness. It is generally acknowledged that bad motives and intentions do count fundamentally for actions's wrongness when wrongness is understood to mean blameworthiness and display of a morally flawed character (Mill 1998: 65, n. 2, Moore 2005: 95-8, Bennett 1981: 97-98, Rachels 1986: 93-5, Thomson 1991: 295-6, 1999: 517, Norcross 1999: 121-2, Enoch 2007: 81, Scanlon 2008: 20-36, 52-6). The voodooist, a sadistic or a revengeful doctor, and the vainglorious rescuer are all bad people and open to blame for the way they do what they do.³

There is, however, much less agreement on whether the same is true of the way motives and intentions matter for the permissibility of actions, apart from various kinds of non-derivative mattering, such as when one's intention to kill signifies the risk of a murder, or when the gravity of a wrong is predictable on the basis of a motive the intent to kill springs from, say, revenge or greed, or when a discriminatory intention grounds the impermissibility of a discriminatory action by being itself grounded in the fact that that action prevents the victim from enjoying important goods or opportunities, etc. Interestingly, most people believe that good motives and intentions, such as the motive of benevolence or the intention to benefit, do not affect the permissibility of actions – impermissible actions do not become permissible or less wrong because they are done with good intentions or motives. However, there is a striking disagreement about whether bad

³ When Kant (1997a: Ak, 4: 390, 398-402, 406, 1997b: Ak, 5:82) says that an action lacks moral worth if done for a wrong reason, or more precisely, from any motive apart from the sense of duty, he does not mean, I believe, that the action is necessarily impermissible but that it has a property in the vicinity of blameworthiness or a flawed character display (cf. Scanlon 2008: 99-105, Parfit 2011b: 170-4, Markovits 2010). Kant (1997a: Ak, 4: 398-9) distinguishes between acting in accordance with duty from acting from duty, and does not think that the former must never be done. It is only the duties of virtue, embodying respect to oneself and others, he (1996: Ak, 6: 383) claims, which require us to act from the right motive (but see Herman 2011).

intentions and motives render otherwise permissible actions impermissible or impermissible ones more wrong. The asymmetry is intuitive but there is no reason why it must hold and why bad and good intentions cannot fall or stand together (Kagan 1989: 177-81, Quong 2013).⁴

The goal of this dissertation is to resolve these and akin disagreements. Although I shall mainly discuss the relevance of bad intentions to permissibility, I shall examine some interesting ways in which good intentions could be relevant. Less precisely but more attractively put, I argue that all intentions are fundamentally irrelevant to the moral permissibility of actions. In a slogan: *It's not the thought that counts*. More precisely, my claim is that we have good reason to believe that what we intend does not count, not even *pro tanto*, in favor or against of what we are morally permitted to do and that the strongest evidence for this fact is negative – at the moment, I see no knock-down argument for the irrelevance of intentions apart from demonstrable failure of the best case-based, principle-driven, and theoretical justifications to the contrary. This does not give us decisive reason to believe that intentions are irrelevant but, as I will argue, it grants that conclusion at this moment in the debate (On the same grounds, but in less depth, I argue that motives are also irrelevant to permissibility.)

I shall refer to these statements as *the Irrelevance Thesis*. I defend the Irrelevance Thesis through an argumentative detour which I shall now outline.

⁴ Some defenses of the relevance of intentions to permissibility are Anscombe 1981, 2006, Cavanaugh 2006, Donagan 1977, Fitzpatrick 2003, 2012, Edmonds 2014, Finnis 2011a, 2011b, Foot 2002b, García 1993, 1995, Hills 2003, Kolodny 2011a, Liao 2012, Lippert-Rasmussen 2010, 2014, McMahan 1994, 2009a, Nagel 1986, Quinn 1989, Shaw 2006a, Sverdlik 2011, Tadros 2011a, 2011b, 2012, 2015a, 2015b, Walzer 1977, Wedgwood 2009b, 2011a, 2011b, 2014.

Some critics are Bennett 1981, 1995, Davis 1984, Di Nucci 2014, Foot 2002a, Glover 1977, Hellman 2008, Kagan 1989, 1998, Kamm 1996, 1999, 2004, 2007, 2008, 2014, Norcross 1999, Nye 2013, Parfit 2015b, Persson 2013, Prichard 2002, Quong 2013, Rachels 1986, Ross 2002, Scanlon 2000, 2008, Sverdlik 2011, Thomson 1991, 1999, Walen 2004, 2006, 2013.

2. An Overview

The dissertation consists of nine chapters divided into three parts. Each part and each chapter is a thematic unit which can be read on its own, provided that the reader is familiar with the debate. To follow the main line of the argument, my advice is not to skip at least the first two chapters and the last one.

The first part, “Contenders to the Irrelevance Thesis”, criticizes four familiar contender views to the Irrelevance Thesis. The first two chapters are devoted to interpreting and criticizing the Doctrine of Double Effect. According to this doctrine, it is permissible to harm other people in pursuit of a good provided that there is no better way to bring it about, that it is sufficiently important to bring it about in such a way, and that the harm is not intended but merely foreseen. The second chapter comes with the appendix, in which the Doctrine of Triple Effect is rejected. According to it, there is a morally relevant difference between acting with intention to harm and acting only because of, or on condition of, harm occurring.

The next two chapters discuss the views that intentions are relevant to permissibility in particular kinds of cases. The third chapter criticizes the view that good intentions, such as those to prevent suffering, render otherwise impermissible actions permissible just when they are one among several actions which overdetermine some wrong (holding the others fixed). The fourth chapter criticizes the view that bad intentions, such as those to harm, render inchoate wrongs of those kinds impermissible.

The conclusion of the first part is that, if intentions matter to permissibility, they do not matter in ways that these views propose. The second part, “Defenses of the Irrelevance Thesis”, assesses and finds wanting four prominent defenses of the Irrelevance Thesis.

The fifth chapter criticizes the objection that the belief that intentions are relevant is a category mistake, confusing the relevance of intentions to blameworthiness or to moral character assessment with their supposed relevance to permissibility.

The sixth chapter criticizes the argument that intentions are irrelevant to permissibility because it can be permissible or impermissible to act for some reason rather than another (and, by extension, with one intention rather than another) only if one can choose to act for a reason but, since that choice is impossible, one can be permitted or forbidden to act, *simpliciter*, and one cannot be permitted or forbidden to act for a reason (and, by extension, with an intention).

The seventh chapter criticizes the argument that, if bad intentions render acting with them impermissible, third parties either have sufficient reason to prevent or to fail to encourage those actions or they can permissibly fail to prevent or even encourage impermissible actions, both of which are supposedly implausible.

The eighth chapter criticizes the objection that, if intentions were relevant to permissibility, one would have to take them into account in decision making, which would be somehow odd.

The conclusion of the second part is that, if the Irrelevance Thesis is true, it is not for the reasons usually appealed to. Together with the conclusion of the first part, this leads to the third part, “Towards the Irrelevance Thesis”, where agent-centered, victim-centered, and impersonal

justifications of bad intentions as wrongs *per se* are examined and rejected. This completes the argument for the Irrelevance Thesis.

The dissertation ends with an error theory of the belief that intentions are relevant and an analysis of some implications of the truth of the Irrelevance Thesis for moral conduct, religious ethics, and criminal law. I shall now briefly address some methodological issues.

3. The Method

The method employed in this dissertation is wide reflective equilibrium. On that method, one's considered moral judgments about some case or an issue, that is, one's intuitions filtered from various distorting factors (biases, ideology, etc.), are first contrasted with the principles used to explicate those intuitions as well as with one's general, moral and non-moral beliefs, even if only indirectly connected to the case or the issue. The judgments and principles are conscientiously reflected upon and, if contradictions between them are noticed, they are revised or abandoned until an equilibrium is reached or, more realistically, approached (Rawls 1999: 18-9, 42-5, 507-8, cf. Daniels 1996, Scanlon 2003).

This is not to say that a research needs to proceed in exactly those steps. They are part of an ideal methodology and only partially represent what actually happens. More often, the starting points are incompletely filtered judgments, revisions are affected by distorting factors, reflection is not careful enough, etc. Nonetheless, shifting back and forth between intuitions and principles is a landmark of normative and practical ethical inquiry, used even by those who reject reflective equilibrium, be it because they deny epistemic authority to intuitions or because they take it as

supreme (see Singer 1972, Kagan 1989, Kamm 2007, cf. McMahan 2000: 107). Purely top-down or bottom-up approaches are in minority nowadays, and, in my view, rightly so.

Thus stated, however, the method becomes almost trivial as it is difficult to see what one should do, being epistemically responsible, if not conscientiously examine the relevant principles and intuitions on their merits (Scanlon 2003: 151). The question, then, is not whether reflective equilibrium is the right, or the most fruitful, method, as it probably is, but what its content should be. This raises numerous questions. Should we consider as relevant ours and the intuitions of our contenders or the average intuitions of some folk or philosophical population? How much weight should our own considered judgments have, as opposed to each other, those of other people, and moral principles? What should we do when we reach conflicting equilibria conscientiously, and starting from equally plausible considered judgments? I shall now reflect on these questions.

Confronted with diversity of intuitions, we ought not to privilege those of philosophers as they are as prone to biases as anyone else's (see Tobia, Buckwalter, and Stich 2011, Schwitzgebel and Cushman 2012). Neither it is much helpful to rely on responses from a sample population of the common folk. Those responses might compel us to provide additional reasons for accepting or rejecting some view or principle when they demonstrate major opposition to or disagreement with it. But, by themselves, the responses do not count, or at least not sufficiently, for or against a position. We ought to reject a view which would try to make a case that, for instance, intentions are relevant or irrelevant to permissibility in some case just because the majority of responders intuitively agree with it. These are not the sort of questions that can be settled by a popular vote.

The best way to rely on intuitions is to consider a case or a principle, arrive at your own intuition about it, try to eliminate the distorting influences, and then contrast it with considered

judgments of those who thought about the same, interpreting them as charitably as possible. We might then encounter at least three types of situations.

Firstly, we might reach intuitive agreement. This is true of some very general issues about intentions, such as that intending harm matters morally, or that good intentions do not suffice to make actions right (“the road to hell is paved...”). However, an intuitive agreement does not give us sufficient reason to be confident that we are right. Even if, intuitively, some proposition looks self-evidently true, we need an independent rationale to back it up. Wide reflective equilibrium is said to level the playing field of justification for intuitions regardless of their generality (Scanlon 2003: 151). The same holds, I believe, for intuitions regardless of their content.

This goes against the views (and interpretations of the method) according to which we need not defend a case-based judgment or a principle if it seems intuitively very plausible, or if it fits well with our other judgments, or if it nicely explains a set of relevant cases. What I propose is not to take even the most self-evident beliefs as infallible, not to think that we have a case just because our views are internally consistent, and not to infer the truth of our beliefs just from their actual or apparent explanatory success. Only by those commitments can we avoid the status quo bias about our deeply held judgments (cf. Scanlon 2003: 150, Kagan 1989: 14).

Secondly, we might disagree but one or all sides might take their views with a reserve, allowing for their fallibility. This is an easy case since the only thing we ought to do is to engage in further examination of those views until we reach some agreement.

Thirdly, we might disagree but one or all sides might be convinced that they are right. This is a difficult situation when the convictions are justified as, unfortunately, it is true of most debates in philosophy, including ours. The disagreement about the relevance of intentions is peer

disagreement; it is pervasive but not unreasonable. A common concern is that such disagreements threaten philosophical progress. We might first ask whether we would continue to disagree under the ideal conditions, knowing all the relevant empirical facts, using the same normative concepts, deliberating carefully, not being under distorting influences, not asking the questions that have imprecise or no answers, and so on (Parfit 2013). It is more likely that, under such conditions, we would share at least some important normative beliefs. We can then aspire to come closer to the ideal conditions.

Part of the disagreement about the relevance of intentions to permissibility is due to using different concepts. In the first chapter I shall argue that we use ‘action’, ‘intention’, ‘motive’, and ‘permissible’, in different, often incompatible, ways, which leads us to talk past each other.

Part of the disagreement lies in different intuitions about whether some action is morally impermissible, even when all the sides use the same concepts. Such disagreements are practically very important, as they leave those inclined to follow moral demands with conflicting guidance. We can come closer to resolving these disagreements in various ways. One is to inspect for flaws in rationales behind the conflicting intuitions about the permissibility of the actions discussed. Since one sense of the term ‘impermissible’ is indefinable, meaning just ‘must-not-be-done’, we can use this strategy only if we appeal to other, definable, senses. For example, we might take the fact that an action is impermissible to mean that, morally, decisive reasons speak against it being done. We can then examine the kind and strength of those reasons. Another way is to consider intuitions about structurally similar cases with the same or somewhat different actions and, after reaching intuitive agreement, to draw analogies or make cautious generalizations. I shall discuss numerous cases where such strategies can help us move from the stalemate.

Finally, part of the disagreement is about why an action is impermissible, when everyone agrees that it is. In such situations, examining structurally similar cases would generate the same kind of disagreement. Therefore, the way to proceed is to re-examine the explanations. For the most part, this is a purely theoretical endeavor. In part two, I critically assess the rationales for the irrelevance of intentions and, in part three, I do it with the rationales for their relevance.

Importantly, even if, after approaching the ideal conditions, some relevant disagreements persist, the possibility of progress, or even its pace, is not threatened. Agreement is not necessary for progress, and for at least two reasons. First, we might reach agreement on one level of debate while continuing to disagree about another but the former might be all that is needed to resolve issues that matter. Second, we might come to a better grasp of an issue through disagreeing about it. We might discover novel ways to understand it, or see merits and demerits of solutions we did not see before. I shall argue that familiar arguments for the Irrelevance Thesis are unsound. But it would be foolish to deny that they significantly advance our understanding of the problem.

I now want to remark on two further methodological issues. The first is about appealing to contrived cases, a methodological device often employed in ethics, this dissertation being no exception. The practice is sometimes criticized because the cases are too detached from the real world. Although this is not true of all the cases I discuss, it is true of some. But there are at least three good reasons to appeal to such cases.

First, it is hard to find real-life cases fit for purpose. An often discussed example in the debate about the relevance of intentions is that of bombing non-combatants in a just war in order to demoralize the enemy and thereby force it to surrender. It is often claimed that it would be impermissible to do it but that it would be permissible to bomb a military facility to achieve the

same goal, merely foreseeing non-combatant deaths as a side effect. The difference is said to lie in the intention to kill behind the former action. We can find out whether those claims are true considering a contrast pair stripped of confounding factors or a real-life analogue, such as the Luftwaffe raids of Britain in 1940-1941. The shortcoming of real-life cases is that our judgments are limited by facts as they happened. We would need to seek for Luftwaffe's strategic and terror raids performed in similar ways, with similar non-combatant casualties, and so on, which might prove an impossible task.

Second, our evaluation of the permissibility of raiding non-combatants as a means and the relevance of the intention to kill for it would be influenced by our existing moral reckoning about Luftwaffe being an instrument of the Nazi regime, the Nazis being the unjust side in the war, and alike (cf. Tadros 2011b: 7-8, Frowe 2014: 5). We need cases where distorting factors are eliminated or significantly curtailed. Hypothetical examples present themselves naturally.

Third, manipulating variables serves to test our intuitions about the moral importance of some phenomenon. We might find out that, under one set of conditions, it is intuitively very clear that some fact plays a role in explaining the impermissibility of an action but that, under similar conditions, with one or a few variables set aside, that fact becomes irrelevant (or vice versa, see Kagan 1988). For instance, we might think that the intention to bomb non-combatants as a means of forcing the enemy to surrender matters to the impermissibility of such bombing since, under ordinary conditions, a bomber with that intention in mind would cause heavier casualties than a bomber intending to destroy a military facility. However, we might also think that, supposing the number of casualties in both cases is fixed, the relevance of the intention to kill is less intuitive (cf. Scanlon 2008: 31-2). The reason why I consider contrived cases, then, is that they exclude

non-derivative ways in which intentions can be relevant. If we focus on real-life cases only, it is much harder to manipulate the variables for that or other purposes.

The second issue concerns methodological demands. On some views, if we argue that a factor is irrelevant in a certain case, contrary to the intuition of many people, we are expected to offer an alternative explanation of that case. Some people believe that this is especially true if we embrace reflective equilibrium. For instance, with regard to our question, McMahan (2009a: 370, cf. Nelkin and Rickless 2014: 131) writes:

I think the theoretical case for the relevance of intention to permissibility is a matter of reflective equilibrium, a matter of the integrity and coherence of our core moral beliefs. (...) [M]any of our firmly held moral beliefs (...) are explained in a simple and seemingly plausible way by the principle that one's intention can affect the permissibility of one's action. (...) If (we believe that intentions are irrelevant to permissibility), [w]e must (...) examine each case in which we thought that intention makes a difference to permissibility and find an alternative explanation and defense of our intuitions in terms of what exceptions there are to certain broad prohibitions.

But as I understand it, the method of reflective equilibrium does not state that a mere coherence of beliefs, or their ability to explain a set of cases, counts sufficiently in favor of accepting them, even if, in addition, they seem obviously true to many people. That intentions are relevant to permissibility does not seem obviously true for otherwise there would not be such a deep disagreement about it. Neither does appealing to intentions explain the permissibility of actions, or so I shall argue.

However, even if everyone thought that intentions are relevant, if one found a compelling argument to the contrary, he would not be obliged to offer alternative explanations of each case

where intentions seem to be relevant. Beside being unreasonably demanding, that requirement is fallacious. If we have good reasons to believe that a factor is irrelevant, those reasons remain if we do not know how to fill in the resulting gap. The burden of explanation, I believe, befalls all sides in the debate.

I shall sometimes take that burden, although offering only tentative proposals. The lack of sufficiently defended alternatives might be seen as one of the drawbacks of this dissertation, even if one agrees with what I have just said. If our methodological standards are high, that worry is justified. Looking at the contemporary ethical theory, one gets the impression that the standards are indeed high. It is customary to expect that one examines the implication of his position for the truth of (non)consequentialism, to offer a solution to the Trolley Problem, etc. My belief is that our goals should be modest. By setting them lower, we are more likely to avoid big mistakes and to make progress. Our steps towards the truth should be more like steps of a tightrope walker between some two skyscrapers or hills – minuscule but more significant than the grand ones that end in the abyss. In this dissertation, my hope is to make a couple of such minuscule steps.

PART ONE

Contenders to the Irrelevance Thesis

1. Double Effect Defects I: Interpretation

Introduction

The Doctrine of Double Effect (DDE) is the best-known principle that states the conditions under which intentions non-derivatively render actions morally impermissible.⁵ According to the DDE, roughly, it is morally permissible to harm others in pursuit of a good if and only if, or just when, the good is proportional, in some relevant respect, to the harm and the harm is not intended either as an end or as a means but is a merely foreseen side effect.

My aim in this two-part piece is to offer interpretative analysis of the DDE and to defend three objections to it. The current state of the debate is partisan to the extent that people either think that the DDE is plainly false or that, in some specified sense, it must be true.⁶ This is in part because they talk past each other. I shall try to reduce unnecessary disagreement clarifying different versions of the DDE in this chapter. In the next one, I shall criticize the best of them.

⁵ Intentions might derivatively render actions impermissible, by pointing towards more fundamental wrong-making features of actions (see Scanlon 2008: Chs. 1-2). The DDE is silent about the derivative influences of intentions.

⁶ Here is McMahan's (2009a: 345) impression: "Between 2000 and 2007, when I was one of the editors of *Ethics*, I reviewed a number of submissions by junior philosophers in which Double Effect was relegated to a footnote and dismissed as an exploded view that no reasonable person could take seriously. I saw that as evidence of a decisive shift in philosophical orthodoxy." Some senior philosophers are equally dismissive (e.g. Singer 2009). Others, like McMahan (2009a: 370), think it is too soon to give up on the DDE. Wedgwood (2009b) remarks straightforwardly: "In my view, the DDE is entirely true."

The chapter starts with an outline of historical, theoretical, and practical importance of the DDE and the waves of attacks on it. This is to show why the doctrine matters and to connect the second, critical, part by describing the origins of and motivations for the objections that are to be defended in it. I then go on to interpret the assumptions laying behind the DDE. I distinguish between the version concerned with permissibility and the one concerned with blameworthiness, the version concerned with intentions and the one concerned with motives, the absolutist and the non-absolutist DDE, the subjective and the objective DDE, the actualist and the possibilist DDE, etc. Moreover, I give some rationale to the DDE's conceptual apparatus, such as the distinction between intention and mere foresight, intention and motive, and the proportionality condition, thereby paving a way for justifications to come. At the end, I give a definition that encompasses the best elements of all of the DDE's variants.

1. Why the DDE Matters

Origins of the DDE go back at least to Aquinas (1981: II-II q. 64 a. 7, cf. II-II q.169). Aquinas famously argued that one is morally allowed to kill a culpable aggressor in self-defense provided that he does not use more force than necessary and that he does not intend to kill the aggressor but only to save his life, even though he might foresee the aggressor's death as inevitable (*Self-Defense*).

Philosophers and theologians after Aquinas modified the DDE in countless ways. Since most of them were associated with the Roman Catholic Church, the DDE became and still is sort

of a semi-official view of Catholic ethical theory.⁷ The DDE's influence is visible, for instance, in the *Catechism of the Catholic Church* (para. 2279, 549) which deems euthanasia—understood as physician-assisted intended killing of a terminally ill patient at his request—impermissible but also states that “the use of painkillers to alleviate the sufferings of the dying, even at the risk of shortening their days, can be morally in conformity with human dignity if death is not willed as an end or a means, but only foreseen and tolerated as inevitable”. Similarly, the Catholic Church distinguishes direct from indirect abortion, the former being an act of intended killing of a fetus at a women's request and the latter any procedure where the death of a fetus is a foreseen but unintended side effect of a beneficial and needed treatment to the woman, e.g. hysterectomy due to uterus cancer. Direct abortion is deemed impermissible but indirect abortion is not (*Humanae Vitae*, sec. 14, *Veritatis Splendor*, sec. 80, *Evangelicum Vitae*, sec. 62).

The DDE is also an important part of secular morality. It is often appealed to in order to explain people's intuitions about some controversial cases. For example, it is used in the debates about the permissibility of unavoidable killing of non-combatants in a just war, in particular as a basis for the distinction between strategic (or tactical) and terror bombing.⁸ Most people believe that it would be sometimes permissible to drop the bombs that would destroy the enemy facilities in order to end the war or achieve an important military advantage even if some non-combatants

⁷ For more about the history of the DDE in the Catholic tradition see Mangan 1949, Kaczor 1998, and Cavanaugh 2006: Ch. 1. Bennett (1995: 200, cf. Kaczor 1998: 300-2) claims that the origins of the DDE as we know it today are not in Aquinas but in Jean-Pierre Gury's seminal work *Compendium Theologiae Moralis*. Whatever the truth about this interpretative issue, it should be noted that the DDE is not exclusively Catholic. Its patterns can be found in other religious traditions: in Talmud (Fisher, Ravizza, and Copp 1993: 707) and among Jewish jurists thereafter, e.g. in Maimonides's *Mishneh Torah* and Joseph Karo's *Schulchan Aruch* (Nelkin and Rickless 2015: 405, n. 3).

⁸ To my knowledge, the earliest mention of the DDE with regard to this issue is in Gury's *Compendium* (the relevant parts of it are translated and discussed in Boyle 1980: 529). Some appeals to the DDE in *jus in bello* are Anscombe 1981: 59, Walzer 1977: 152-9, and McMahan 2009a: 359-70. Some critics are Nagel 1979: 60-1, Bennett 1981: 95-116, 1995: Ch. 11, Thomson 1991: 293, Predelli 2004: Ch. 3, Steinhof 2007, Kamm 2012.

would predictably die as a collateral (*Strategic Bombing*). Most people also believe that it would be impermissible to kill the same number of non-combatants in order to demoralize the enemy and persuade it to surrender (*Terror Bombing*). The difference in permissibility, it is said, lies in the intentions behind the bombings; the former pilot or the commander of the air force giving the order foresee but do not intend the deaths of non-combatants while the latter intend their deaths as a means to a good end.

Similarly, the DDE is often invoked in medical ethics.⁹ Take for example the distribution of scarce medical resources. It is believed that it is permissible to withhold some life-saving drug from a patient in need for the whole supply of it in order to give one fifth of the drug to five other patients who will be thereby saved (*Drug Shortage*). However, it is also believed that in an organ shortage crisis it is impermissible to kill or to let a patient die in order to use his organs to save five other patients (*Transplant*). Again, intentions are said to be the basis of the disparity; in *Drug Shortage*, the doctor foresees but does not intend the patient's death whereas in *Transplant* the doctor intends the patient's death as a means of saving the five.

Finally, the DDE appears to explain some well-known cases in normative ethics, such as the notorious Trolley Problem (McIntyre 2014: sec. 2, Edmonds 2014). Consider the following two cases. In *Bystander*, a runaway trolley is heading toward five persons trapped on a railway track who will be killed unless something is done. I am a bystander who can press the switch that will redirect the trolley on a side track. Unfortunately, one person is trapped on the side track, so if I press the switch, the trolley will hit and kill the one. In *Bridge*, again a runaway trolley is heading towards five trapped people but there is no side track. Rather, I am on a bridge above the

⁹ Literature on the DDE in medical ethics and bioethics is vast. See, e.g. Marquis 1991, Boyle 2004, McIntyre 2004, Sulmasy 2007. The echoes of the DDE are present in the Supreme Court's decision in *Vacco v. Quill* (1997).

track and nearby me is a very large man. I know that if I push the man off the bridge and onto the track the train will hit and kill him but the collision will stop the train and the five will be saved. Although in both cases I would kill one person and save five others, most people believe that it would be permissible (but not obligatory) to redirect the trolley in *Bystander* but impermissible to topple the large man in *Bridge*.¹⁰ The reason for it, the DDE tells us, is that in *Bystander* I would not intend to kill the one while in *Bridge* I would intend to kill the man as a means of saving the five.

The DDE was and still is the bedrock of much Western moral thinking and it is not an exaggeration to say that it became a part of common sense morality (Nagel 1986: 179, Kagan 1989: 84-5).¹¹ It was not until fifty years ago that the DDE started to get attacked by influential philosophers.

To understand the motivation of the first wave of attacks, it should be noted that along with the Doctrine of Doing and Allowing (DDA)—which states that it is morally worse to cause harm than to let it happen¹²—the DDE makes the core of deontological constraints or prohibitions on the consequentialist calculus of harms and benefits (cf. Quinn 1989: 334, Kagan 1989: 83-7, 1998: 102, Kamm 2007: 21). These principles impose constraints on what we may do in making things go best. It is then not surprising that the first attacks on the DDE came from

¹⁰ Studies confirm the uniformity of these intuitions across societies, cultures, gender, etc. See, e.g., Hauser 2006, Banerjee et al. 2010.

¹¹ Mikhail (2011) argues that the DDE did not become but always was a part of common sense morality. The DDE is a crucial part of what he calls the universal moral grammar, a complex system of moral concepts and principles that is innate, largely unconscious, and universally shared. To justify this thesis, Mikhail brings studies of intuitions in a series of trolley cases stipulating that the data is best explained by the participants's tacit and innate knowledge of the DDE. However, the latter assumption remains unjustified and, given that the application of the DDE in trolley cases is highly controversial, the argument cannot run. This is so even if we cannot offer better explanation of these cases as long as we have good reasons to reject the DDE's explanation.

¹² The best anthology about the DDA is Steinbock and Norcross 1994. A recent defense of it is Wollard 2015.

committed consequentialists, and often in tandem with attacks on the DDA (Williams 1957, Hart 2008, Glover 1977, Bennett 1966, 1981, 1995, Rachels 1986, and Kagan 1989).

The first wave brought two powerful objections (besides several minor ones). The first is that, due to the constraint on intending harm being too strong, the DDE intuitively forbids too much (Kagan 1989: 132-65). The second states that, because intentions can be fine-grained, the constraint on intending harm cannot be accounted for in a way that it applies to intending that which is not harm but which is so close to it that it must be similarly forbidden without applying to foreseeing but not intending harm (Hart 2008: 123-4, Bennett 1981: 106-16, 1995: 204-13). Call these *the Intuitive Objection* and *the Closeness Objection*, respectively.

The second wave appeared with non-consequentialists joining the attack (Thomson 1991, 1999, Scanlon 2000, 2008, Kamm 2004, 2007, 2008, 2012). They made a departure from earlier non-consequentialists who held that giving up on the DDE means sliding into consequentialism (see Anscombe 2005: 190-2, Foot 2002b: 88).

Apart from the arguments supporting the general thesis about the irrelevance of intentions to permissibility—some of which were the same as those of the first wave¹³—the second wave brought a novel inductive objection to the DDE. The objection states that if we reflect closely on the examples taken to corroborate the DDE we shall see that they are explained better by other principles. This might lead us to claim that the weak explanatory potential of the DDE is a good indicator of its falsity (Scanlon 2008: 28-36, 69-87). Call this *the Explanatory Objection*.

¹³ An important common objection states that the defenders of the DDE make a category mistake: they confuse permissibility with other moral categories, such as blameworthiness or the agent's character assessment (see Bennett 1981: 97-8, Rachels 1986: 93-5, Thomson 1991: 295-6, 1999: 517, Scanlon 2008: 20-36, 52-6). I do not think that this objection succeeds but I shall not discuss it here.

The outline I have just given is crude. The first wave was not exclusively consequentialist (e.g. Foot 2002a)¹⁴ nor was the second wave non-consequentialist (e.g. Norcross 1999, Persson 2013). I put it in this way to show that, despite different profiles and motivations of the authors, their opinions converged in the prospect that urged moving beyond the DDE.¹⁵

The place of my argument is inside the debunking tradition. Its aim is limited to rebutting the DDE. I shall not comment on how the DDE's falsity relates to the truth of consequentialism or the plausibility of other deontological constraints. My only aim is to take the three objections and defend them against some recent criticisms. Why only these? First, I think they are the most powerful ones. I show how the DDE can avoid some other objections with minor reformulations.

Second, the focus of these objections is the DDE. It is not the wider thesis that intentions are relevant to permissibility.¹⁶ Some good, but to my mind unsound, arguments have been given against that thesis and, by extension, against the DDE. I shall not consider them in this chapter. It therefore fits the purpose to consider objections directed solely against the DDE separately.

Third, these objections have a common dialectic. They take the moral relevance of the distinction between intending and merely foreseeing as given and show what, together with the other assumptions behind the DDE, it entails with respect to moral obligations in or explanations of the relevant cases. An important objection takes the opposite route. It states that, because no

¹⁴ Taken by the DDE's appeal in medical cases, Foot has later abandoned her criticisms. See Foot 2002b: 89-92, esp. 91, n. 6.

¹⁵ The DDE is implicit in laws concerning war, euthanasia, and many other practical matters. Some of its critics seem to think that it is not only moral beliefs that should be revised but laws relying on them as well. For example, regarding physician-assisted suicide, Thomson (1999: 518) writes: "It is [...] unfortunate that [the DDE] have had so powerful an impact on the state of American law in this area." But the revision of the law follows only if law ought to closely follow morality. This is a controversial assumption that cannot be discussed here.

¹⁶ Both Thomson and Scanlon present arguments against that thesis saying that they argue against the DDE. But I think it is plain that the arguments extend much further and that, thus, should be considered separately.

satisfactory explanation of morally relevant difference between intending and merely foreseeing harm exists, we should be skeptical about the distinction's moral relevance, and, consequently, about the DDE's truth (Kagan 1989: 177-82).

Both strategies, I believe, are on the right track and indeed mutually enforcing. However, although the question whether the distinction between intending and foreseeing harm has moral relevance is explanatorily prior to the question whether the DDE is true, we can resolve the latter even if we have not resolved the former. Contrary to a popular opinion, one can acknowledge the DDE's falsity and yet hold that the distinction matters morally but that it plays a different role from the one assigned to it by the DDE. For our purposes, it is therefore sufficient to pursue only the first strategy.¹⁷ But first, we need to understand what the DDE actually states. This is the task for the rest of the chapter.

2. Varieties of the DDE

2.1 The Classical Definition

Anscombe (2005: 251) wrote epigrammatically: "The corruption of non-Catholic moral thought has consisted in the denial of [the DDE], and the corruption of Catholic thought in the abuse of it." I agree about the latter. Countless versions of the DDE have been proposed. We should begin by discarding the implausible and, for our purposes, irrelevant ones.

¹⁷ This would not be sufficient, however, to establish the thesis that intentions are irrelevant to permissibility. That thesis is true only if it is true that reasons to avoid intending harm cannot count against an action even on *pro tanto* level and one of the first steps in that direction seems to be to show that reasons against intending harm are on a par with those of foreseeing it, that is, that no morally relevant difference between the two exists.

We can start with one classical definition (see Mangan 1949: 43, Connell 2002: 880). According to it, the DDE sets out four conditions which, when jointly satisfied, make it the case that one is morally permitted to act. These are: (1) The act itself is good or at least indifferent (it is not in itself bad); (2) The good effect is intended and the bad effect is foreseen but unintended (i.e. merely foreseen); (3) The bad effect is not a means to producing the good effect; and (4) the good effect is sufficiently important for bringing about the bad effect.

There are two ways to understand this definition: as a biconditional and as a conditional. On the former reading, the DDE does not only license certain behavior once the four conditions are satisfied but also prohibits any behavior violating any of those conditions. The DDE is then not just a principle which determines when actions are morally permissible (as the conditional reading would suggest) but a principle which determines when they are impermissible. Although the confusion about the two readings is endemic in the literature, I believe that, traditionally, the DDE was thought not only or even primarily as a licensing principle but as a principle of moral prohibition. For example, in the famous passage, Aquinas does not claim only that, if I satisfy the four conditions, I may kill a culpable aggressor in self-defense. He claims that it is unlawful to do it if, moved by animosity, I intend to kill him. My impression is that most other defenders of the DDE, both traditional and contemporary, understand it in this way (see, e.g., Anscombe 2006: 231-2, McMahan 2009: 345-6, Wedgwood 2012a: 384).¹⁸

¹⁸ I also take it that the DDE's explanatory purpose is to explain both why certain acts are permitted and why others are prohibited. It is a difficult question what the scope of these acts is. One might make the point that, traditionally, the DDE was meant as a very wide-range principle of morality but that is surely not how it is nowadays understood. In the section on the fourth condition, I shall briefly discuss ways to build in other moral principles into the DDE as well as ways in which such principles might replace the DDE in explaining the relevant acts. A stricter delineation of the DDE's scope is an important task for its defenders but critics as well (indeed, it is what they must agree on to reduce talking past each other). I cannot possibly undertake that task here.

It might be said in reply that if we understand the classical definition as a biconditional, the DDE will prohibit way too much. On the natural reading of the third condition, the DDE will prohibit permissible harming as a means of benefiting oneself or others, such as killing in self-defense or performing some surgeries. This worry can be met. I shall suggest below that on the most charitable reading of the DDE, it prohibits harming as a means only when one intends such harm, thus suggesting that in cases of self-defense and similar kinds of harming as a means one need not intend the harm one takes as necessary means to one's ends.

At the end of the chapter, I shall also give a definition of the DDE which is faithful to the classical definition in its insistence on the prohibition on intentionally bringing bad effects as a means while avoiding many of its other problems I discuss below. Perhaps the DDE would have greater plausibility if stated in the conditional form but its theoretical importance would then be curtailed as well. We want the DDE to tell us what we ought not to do, not only what we may.

I shall now turn to the analysis of the concepts featuring in the classical definition as well as its four conditions.

2.2 Permissibility and Blameworthiness

On this definition, the DDE is about moral permissibility or impermissibility of actions (or, in the old-fashioned discourse, their licitness or illicitness). Some people (Duff 1982, Nagel 1986: 180, Chan 2000: 424-33, Chappell 2002: 223-5) suggest that the relevant version of the DDE is about an action's blameworthiness (or moral culpability). If so, the DDE might be uncontroversial to the extent that it seems intuitive that acting with intention to harm as an end is blameworthy even

if, all else being equal, acting with mere foresight to harm is not. But it would be controversial to the extent that it is less intuitive that there is a difference in blameworthiness between acting with intention to harm as a means and an identical causing harm as a merely foreseen side effect (see Bennett 1995: 221-4, Roughley 2007: 103-8, DiNucci 2014: Ch. 12). I think that those worries can be answered but, since the majority of defenders and the critics of the DDE take it in the former way, it is that version which will be the target of my critique.

There is a further complication. Permissibility has two senses: definable and indefinable (Parfit 2011a: 165-74).¹⁹ On the indefinable sense, impermissibility and permissibility cannot be helpfully explained in terms of other concepts. Impermissible means just ‘must-not-be-done’ and permissible ‘may-be-done’. In other words, if an action is impermissible in this sense, it has an indefinable property of that which must-not-be-done, and it is debatable whether any action has that property (Parfit 2011a: 173).

There are many definable senses. For the purposes of this inquiry, it suffices to introduce the following two. On the moral-decisive-reason sense, impermissible means that which one has, morally, a decisive reason not to do, although not in the sense that there is a moral, as opposed to non-moral, reason counting decisively against an action but in the sense that the action is morally outweighed by decisive reasons in the overall weighing. This sense might have to take the word ‘moral’ as indefinable (Parfit 2011a: 170). On the divine command sense, impermissible means that which is forbidden by God. This leaves open that an action is impermissible not just because God prohibits it but because it exhibits a property (e.g. sinfulness) which counters God’s will.

¹⁹ The concept ‘wrong’ has even more meanings, embodying ‘permissible’, ‘blameworthy’, etc. For a partial list see Parfit 2011a: 165-6.

Anscombe (1981: Ch. 4) claims that moral obligations and prohibitions make sense only in the divine-command sense. And Aquinas (1981: II-II q. 64 a. 7) claims that a soldier who kills a foe in self-defense or a judge who orders killing a robber “sins if moved by private animosity”. His conception of permissibility seems tied to his conception of sinning. However, when most secular philosophers discuss the DDE they use permissibility either in the indefinable or in the moral-decisive-reason sense. The disagreement is then in part because the opposing sides appeal to the irreconcilable senses of permissibility.

I shall explore how the latter senses feature in arguments for and against the DDE. My objections do not target the versions of the DDE relying on the divine command sense. For the reasons I cannot go into here, I think that that sense is, nonetheless, implausible.

2.3 Absolutist and Non-Absolutist DDE

A related worry is whether permissibility is too rigid a concept for the best version of the DDE. An action is either permissible or impermissible; it cannot be more or less permissible. However, some might claim that acting with intention to harm as a means or as an end is worse, or harder to justify, than acting with a mere foresight to harm as a side effect but that the former need not be impermissible if, other things being equal, the latter is not (Quinn 1989: 344, McMahan 1994: 201, Hills 2003: 134, Wedgwood 2011a: 385-6, Nelkin and Rickless 2014: 127, 2015: 376). This version of the DDE is often called non-absolutist, as opposed to the absolutist DDE, an instance of which is the variant stated above.

The non-absolutist DDE, however, is unsatisfactorily underspecified. Its cautious claim that intending harm is worse than foreseeing it but not invariably and conclusively worse comes at the price of not identifying the extent to which the former is worse. Without this specification, the non-absolutist DDE cannot explain why, in some but not other cases, intending harm is as bad as to render an action impermissible. This creates a problem if we want the DDE to entail a unifying answer about the paradigm cases such as those mentioned in the previous section. If the non-absolutist DDE implies, as it intuitively should, that these are clear cut cases of permissible/impermissible actions, then the question is why other similar cases are instances of just more/less wrong actions. I am not saying that no explanation could be given and, in the next chapter, I shall discuss non-absolutist revisions of the DDE. But to the extent to which we need an operative version of the DDE, it is preferable to start with an absolutist one and then see whether, given some objection, a non-absolutist variant can do better.

2.4 The DDE and Moral Absolutism

We can now turn to the four conditions of the classical definition. The first one is ambiguous (Bennett 1995: 197). That an action should not be bad in itself cannot mean that there must be no reason against it being performed, since that would prohibit bringing about any harm as a merely foreseen side effect. Nor can it mean that an action should not be impermissible in the given circumstances as that would make the other conditions obsolete. It can only mean that an action should not be one of those which are always impermissible, whatever the consequences.

This makes sense if one is a moral absolutist. Anscombe (1981: 58, 64, 66) takes the

DDE to arise from and complement a tradition that deems some actions absolutely forbidden. Boyle (1991a: 477, 486-7) adds that, unless one is an absolutist, one is not entitled to appeal to the DDE. However, apart from its history, there is nothing in the DDE that makes it linked to moral absolutism. And, whereas a plausible absolutist moral theory must be complemented with a set of exceptions of which the DDE is the best known candidate, it is far from obvious, *pace* Boyle, that the only or the best way to justify the DDE is by assigning to it this argumentative purpose (Kamm 1991: 572, cf. Boyle 1991b: 566). Since most people nowadays, including most defenders of the DDE, are not absolutists in this sense, we might as well do away with the first condition.

2.5 Ends, Means, and Side Effects

The second condition brings several issues that are essential for the proper understanding of the DDE. It states that good effects should be intended but bad ones merely foreseen. We can first distinguish between the effects which are ends we are trying to achieve, those which (we believe) are means to those ends, and those which are foreseen or foreseeable but which are neither our ends nor means to those ends and which are called side effects (Parfit 2015b: 4). In cases the DDE deals with, the ends and the means are intended but the side effects are not. In *Tactical Bombing*, for instance, the intended end is winning the war, the intended means is destroying the munitions factory, and the foreseen but unintended side effect is killing non-combatants.

Although there are three distinct effects here, only two are presumably morally relevant. Hence the double effect talk. If an action is permissible, the end sought is good effects while the

foreseen side effects are bad (the former are sometimes called greater goods and the latter lesser evils). The intended means are not bad for otherwise the action would be impermissible, as in *Terror Bombing*. They are either good or neutral. If an action is impermissible, it is its end, the means to its end, its foreseen side effects, or any combination of these that are bad and intended. The foreseen side effects are then intended as a further end or as a means and they are not *merely* foreseen.²⁰ In cases that are mostly discussed, either the intended end or the intended means are bad.

2.6 Intention and Mere Foresight

The underlying assumption behind the claims just made is that it is in principle possible to distinguish intended harms from merely foreseen or foreseeable harms, even if, in practice, it might not be always easy to do it (Hills 2007: 259). This is a conceptual assumption; it bears no normative baggage. Even if we can distinguish the two, it does not follow that the distinction has any moral relevance.

The conceptual claim was denied for reasons we need not go into (see Sidgwick 1981: 202, Chisholm 1970: 636-7). These doubts are dismissed by counterexamples where, intuitively, a foreseen effect does not entail an intention to bring it about. If *Tactical Bombing*, *Bystander*, or *Drug Shortage* are not sufficiently intuitive for this purpose, it is because they are not the cases we encounter on a daily basis. There are, however, everyday cases that intuitively confirm the

²⁰ An outdated synonym for mere foresight is ‘oblique intention’, as opposed to ‘direct intention’, which is what we nowadays call just ‘intention’ (e.g. Bentham 1907: IX. 24, Hart 2008: 120-1).

distinction. Suppose I intend to go for a run knowing that, in all probability, I will get tired. But I normally would not intend to get tired; I would only foresee that happening (*Runner*).²¹

Not only do we need not intend the foreseen causal effects of our actions; we need not intend the effects logically entailed by them. If I intend to kill five people, I need not intend to kill a prime number of people, or the sum of two and three people, etc. We also need not intend what is causally entailed, under normal conditions, by what we intend. If I shoot you in the head, I might intend to blow your brains out but not to kill you, even if I know that, in all likelihood, one entails the other. Such cases rarely occur but they need not be conceptually impossible. And, as we shall see, they are theoretically very important as they give rise to the Closeness Objection. It is interesting that, while ordinary usage of ‘intention’ is in line with the judgment in *Runner*, it is less clear whether it conforms to the judgments in the latter examples. The notion of intention that features in these debates might be a term of art which is not fully coextensional with the ordinary language notion (cf. Boyle 1991a: 479). I shall say more about this technical notion in a moment and come back to the hard cases in the next chapter.

In paradigm cases the DDE deals with, the foreseen side effects are causally independent of the intended ends in the sense that they neither contribute to causing those ends nor they are caused by them (Parfit 2015b: 4). The deaths of non-combatants is not needed for the war to end and neither is the death of the one for the sake of saving the five. And it is not the winning of the war which kills the non-combatants or the saving of the five which kills the one.

²¹ Some people argue that the converse is true: intention implies the belief that the intended event is likely to occur. This, it is sometimes said, is what distinguishes intentions from hopes (see e.g. Audi 1973: 387-9, 1993: Ch. 2, Harman 1976: 434-7). It is debatable how low one’s expectations have to be for one to count as intending something since we do not want to say that a person who buys a lottery ticket and wins at one in a million odds intended to win. Because nothing about the DDE hinges on this debate, we can set it aside.

However, there are cases where merely foreseen side effects are causally related to the intended ends in one of the two ways. The means to the intended end might be unintended even though the agent finds them necessary for achieving that end (Kagan 1989: 167-8, Broome 2002: 91-2). Suppose chemotherapy is known to acutely damage a patient's organs but the damage is the necessary means of killing cancer cells. A doctor who would prescribe that treatment might not intend to damage the organs despite knowing that doing so is the means of curing the patient (*Chemotherapy*). I do not expect everyone to share my intuition about this and there is certainly an ordinary language usage of 'intention' on which the doctor must intend to damage the organs. But it is plausible, or so it seems to me, that on an equally ordinary usage, he need not.

Some authors disagree. Anscombe (1981: 59) writes: "It is nonsense to pretend that you do not intend to do what is the means you take to your chosen end". However, as *Chemotherapy* shows, this is far from nonsense. The same line of thinking seems to be behind the Catechism's prohibition of euthanasia and direct abortion. But if I am right, the doctors performing these need not intend patients or fetuses to die (how often they intend it is a different, empirical, question). What is more, the followers of the DDE who take Anscombe's view and the DDE as prohibiting and not only licensing certain behavior must conclude that actions such as the chemotherapy are impermissible or explain why they are exceptions. It is best, I believe, to allow for the possibility of merely foreseeing the means we deem necessary for achieving an end.

Some foreseen effects are caused by the intended end, as it is the case in *Runner*. More interestingly, some such effects are themselves necessary means to the achievement of an end but are unintended by the agent. If I intend to give a party for me and my friends to have fun, I might foresee that this will leave a mess. I also know that my friends will feel indebted and help me

clean it up. Assuming I do not like cleaning and would not give the party if my friends would not help me, their feeling indebted and helping me are the necessary means for me giving the party. This is a stretched sense of ‘means’, meaning only that I give the party under the condition that my friends help me clean up. But I need not and normally would not intend them to feel indebted or to help me clean up (*Party*, Kamm 2007: 95).

2.7 The Counterfactual Test

Although the distinction between intention and mere foresight is intuitively appealing, it is much harder to give a strict delineation between the two and to justify it.

The most well-known tool for these purposes is the so-called Counterfactual Test. On one plausible statement of it, a necessary condition for intending an effect is a reaction of the agent if, unexpectedly, the effect had not occurred. Three clarifications are in order. First, the test gives a necessary condition for intending. If we understand passing it as a sufficient condition, we run into the problem of counting as intended the effects that the agent acts only because of (Kamm 2007: 96, 124, n. 17). However, as *Party* shows, there is a conceptual difference between acting because an effect will occur and acting with intention to bring about an effect.

Second, the test assumes that the counterfactual where the relevant effect does not occur is the one where the agent knows all the relevant facts (cf. Shaw 2006b: 202). This rules out the epistemic objection that, if the non-occurrence of an effect was miraculous or beyond reasonable expectation and the agent had truly and justifiably believed that the effect was necessary for the achievement of an end or that it was the end’s most likely side effect, then he would falsely but

justifiably believe that the end did not occur either, which would affect his subsequent reactions (Bratman 1987: 142, n. 6, Bennett 1995: 216-7, Kamm 2012: 28-9).

Third, the test leaves open what the agent's reaction should be. There are at least three versions of it. We can ask whether the agent's plan would have been thwarted had the effect not occurred (Searle 1983: 103, Donagan 1991: 496, Uniacke 1994: 106-7), whether he would have pursued other actions to bring the effect about (Fried 1978: 24, Bennett 1995: 218, 221-2, Kamm 2007: 96), and whether he would have felt disappointed, or would not be glad, had the effect not occurred (Bennett 1981: 102-4, 1995: 221, Masek 2010: 584). If the answer is affirmative to any of these questions, the agent intended the effect, or so the versions of the test imply.

These questions are not accidental; each indicates what are thought to be necessary (but not sufficient) conditions for intending: to be settled on a plan of action, to pursue the available means to bring an end about, and to have a positive attitude towards that end occurring. I shall return to these conditions in a moment but I now want to show why an objection often brought to the Counterfactual Test fails.

Many people claim that, while the Counterfactual Test distinguishes intending harm as an end from merely foreseeing it, it fails to distinguish intending harm as a means from merely foreseeing it (Davis 1984: 119, Marquis 1991: 530, Bennett 1995: 222, Uniacke 1994: 107, Shaw 2006b: 203-4, DiNucci 2014: 33-5). Recall *Terror Bombing*. These people claim that, if the enemy is demoralized and the war ended but, miraculously, no non-combatant died, the terror bomber's goal would be achieved, he would not retry to kill non-combatants, and he would be glad that they did not die. All this would not be true of a sadist or a revengeful bomber who intend to kill non-combatants as an end. But the terror bomber is not like them; he would be

happy to make an omelette without breaking any eggs. The Counterfactual Test implies that the terror bomber did not intend to kill the non-combatants as a means of winning the war, which cannot be true.

This objection fails because it confuses the intention to *kill* non-combatants as a means of winning the war with the intention to harmlessly affect or use them for that purpose or with the intention to bomb them, not caring whether they end up being harmed or not.²² The confusion is understandable given that the ordinary notion of ‘terror bomber’ is underspecified. Assuming the war is miraculously won, the first intention is satisfied only if the victims are killed, the second, only if they are not harmed, and the third, either way. If the terror bomber’s intention is the first, then he is very much like the sadistic or the revengeful bomber, who would not be glad had non-combatants survived and who might even pursue other means to kill them (cf. Delaney 2007: 104-5). However, if his intention is the second or the third, his plans would not have been thwarted had the war been won but non-combatants survived. He would not be dissatisfied nor would he retake action to kill them. The Counterfactual Test correctly tracks these judgments.

It might be objected that the intention to harmlessly affect non-combatants by bombing them or the intention to bomb them while being indifferent whether they are harmed are as bad as the intention to kill them. This is true but it misses the point; the Counterfactual Test is not the test of an intention’s badness but a test for distinguishing intention from mere foresight (to find a principle that counts these intentions as morally on a par is an important task and I shall discuss it in detail in the next chapter).

²² This shows, more generally, that we might intend conjunction, disjunction, or any logical connection of events, but also condition intentions on (a likelihood of) certain events occurring (cf. Shaw 2006b: 211-2, Pruss 2013: 63-5).

A greater worry is that the given analysis presupposes that intending necessarily involves a non-instrumental attraction to an effect, such as killing, but it does not. A bomber can intend to *kill* non-combatants as a means to an end while not being attracted to their deaths in such a way. While I agree that this is a controversial assumption that needs defense, at the moment I can only appeal to cases which seem to confirm it, such as *Chemotherapy*. Those not convinced by them will not be convinced by my reply to the critics of the Counterfactual Test.

But that does not matter much. I did not mean to defend the Counterfactual Test as the only test for the distinction between intention and mere foresight but to show that it might serve the purpose better than usually thought. A different issue does matter here. The defenders of the DDE better not appeal to the Counterfactual Test together with the assumption that intending an effect entails a non-instrumental attraction to it. Although the conjunction of the two succeeds to avoid the often brought objection, it also implies that on the most natural reading of, say, *Bridge*, I do not intend the large man's death, or the man's falling, or even his being pushed or affected. This puts a high bar on intending that might come handy in hands of the DDE's critics but surely not in those who wish the DDE to discriminate against such cases.²³

Let me now return to the three versions of the Counterfactual Test. As I have already said, they correspond to necessary conditions for intending (the list is far from complete). Examining them will help us see how one might justify the distinction between intention and mere foresight.

²³ There are further problems with the Counterfactual Test. Its action-guiding applications are limited, the epistemic condition being what holds it back most obviously. More often than not, we do not have all the relevant information. A strategic bomber might not be able to know whether he destroyed a munitions factory (e.g. the area where the factory is located might be covered with thick clouds, Kamm 1996: 255, n. 10), so he might take non-combatants's deaths as a sign of success. Seeing them alive would be a bad news for him. But it would not follow that he intends to kill them even if he would not abort the raid until he saw them dead.

The first idea is that intending involves being settled on a plan of action, although not irrevocably as one might give up on it, e.g. no longer seeing it as worthy of pursuing (Bratman 1987: Ch. 3, Mele 1992: 142-6). In this respect, intentions differ from beliefs that an action will have certain effects, i.e. from mere foresights, and from desires to do it, both of which lack the intentions's executive component. Mere foresight is motivationally inert; just believing that an action will have such and such effects never leads to acting. And even if desire is unlike that,²⁴ it clearly has no commitment side. For example, there is nothing unusual in desiring to go for a run and not moving from bed. But it seems unusual, indeed, it is an epitome of irrationality, to intend to go for a run and do nothing about it.

It is generally agreed that intending involves means-end reasoning. But a further feature is sometimes mentioned: that one must exploit all the available means to pursue an end that fails to be achieved at first (Bratman 1987: Ch. 10). This is not always the case. Some intentions are conditional upon certain means, or certain circumstances. I might intend to go for a run only if the weather is not bad or only if my favorite pair of sneakers is clean.

However, it seems that intending an end always involves taking a positive stance towards it (Kagan 1989: 170-1, Scanlon 1998: 33-4, Hills 2003: 137-8). One has to believe that the end is worth of pursuit, i.e. that there is some reason to pursue it. In intending a means to an end, one does not believe that the means are good in itself, as one takes them as instrumentally good (but as good, nonetheless). If I intend to shoot the aggressor in self-defense, I shall believe that there is a sufficient reason to shoot him but I need not believe that shooting him is good as such, even

²⁴ On some views, desires are necessary for motivation (see, e.g. Smith 1994: Ch. 4). On others, beliefs about there being sufficient reasons for acting can alone motivate (see, e.g. Nagel 1970: Ch. 5, Scanlon 1998: 33-41). Nothing about our question hinges on which of them is true.

if he is liable to be shot and even if shooting is the only way to defend myself. Clearly, no such things are true of foreseeing effects of our actions.

Believing that the intended end is good is joined by a positive emotional response to that event occurring, such as gladness or contentment. I do not think that one can stay unconcerned to the occurrence of an event even if one intends it only as a means (cf. Kagan 1989: 170-1). This is what motivates the version of the Counterfactual Test asking whether one would be disappointed or at least not feel glad had the effect not occurred.²⁵ Again, all this depends on whether we take intending an effect to imply a non-instrumental attraction to it.

We saw the Counterfactual Test's merits and demerits. It is likely that we would be better off with some other test (I do not know what that test would look like). That being said, I shall sometimes (implicitly) appeal to the Counterfactual Test primarily because of its prominence in the discussion rather than because I think it is the best tool for distinguishing intention from mere foresight.

2.8 Intention and Motive

The belief that there is a sufficient or a conclusive reason to act in some way usually, and at least partially, motivates one to act in that way. That belief, sometimes called an operative reason, is often a motivating reason; that is, it is what moves one to act in some way. Motivating reasons

²⁵ According to Knobe (2003), we often judge an effect as intended when it is bad and unintended when it is good. There is then a problem with drawing the distinction between intention and mere foresight in a non-moralized way. Much has been written about this problem, often called the Knobe effect. My opinion is that Knobe's experiments are deeply flawed because they fail to make clear a textbook distinction between two senses in which an action can be intentional: a broader sense, meaning that an action is knowingly performed, and, a narrower, meaning that it is intended. Since we usually ascribe blame to causing bad effects in the broader sense but praise only to causing good effects in the narrower sense, the subjects are likely to uncritically shift between the two (see Scanlon 2010).

give rise to intentions (akratic behavior aside). We intend what we take to be reasons for acting (Bennett 1981: 99, 1995: 194-5, Scanlon 1998: 33-4, 66, 2000: 306-7). However, motivating reasons are not identical with intentions. For one thing, we might be motivated by events which we cannot affect. My main motivation to go for a run might be the fact that it is a nice weather outside (or the belief that a nice weather is conducive to running) but I cannot intend the weather to be such and such, as that is true independently of my will. What moves us to act and what we intend need not, therefore, be the same.

Part of the disagreement between the defenders and the critics of the DDE appears because they take different (presumably) necessary conditions for intending as sufficient. Some people (e.g. Wedgwood 2009b) claim that being settled on a plan is all there is to intending while others talk about intentions and motivating reasons as if they were one and the same (Anscombe 2000: 9-10, Fried 1978: 22, Duff 1990: 37-73, Scanlon 2008: 11). Those accounts are too narrow. A better one might include the stated conditions and, perhaps, others. It goes beyond the scope of my discussion to develop such an account.²⁶

A disagreement in the neighborhood seems more worrisome. Recall that, according to the second condition of the classic definition, the DDE forbids intending bad effects. Now one and the same intention can stem from morally opposing motivating reasons which are not beliefs of the stated sort (or, possibly, not beliefs at all). Compare a doctor who administers a lethal drug to a terminally ill patient in anguish, and at his request, solely out of concern for the patient dying with dignity with the doctor who does the same solely out hatred towards the patient (*Doctor* and *Bad Doctor*). Both doctors believe that there is a sufficient reason to administer the drug and

²⁶ That task might prove futile as intending might be a family resemblance concept. What is more important, I think, is to examine the moral relevance of different meanings of intending.

intend to administer it. But it is only the latter doctor who believes that the patient's death would be good as such (cf. Thomson 1999: 514).

As *Chemotherapy* shows, the former doctor need not intend to kill the patient as a means of improving his condition. He might intend only that the patient is affected in a way such that his illness and suffering vanishes. However, we can assume that, in some such cases, he would in fact intend the patient to die. Would the DDE prohibit his action or only the one of the malevolent doctor? More generally, does the DDE prohibit acting with intention to harm or just acting with such intention when it stems from bad motives?

Textual evidence suggests that most defenders of the DDE assume that it prohibits both intentions, as their formulations focus on the intention to harm, leaving aside the issue of motive from which the intention stem. However, there are some notable exceptions. As already said, in the famous passage Aquinas writes that it is unlawful to intend to kill in self-defense but he also gives a caveat, namely that holders of public authority or soldiers in just war may lawfully intend to kill criminals and combatants but that they sin if being moved by private animosity. There are two ways to interpret the caveat: either Aquinas's notion of sinning grounds the notion of acting unlawfully, in which case he thinks that it is impermissible to kill whenever one acts from a bad motive but not whenever one intends harm, or the two notions are independent, explanatorily and coextensively, in which case he thinks that under special occasions it is permissible but sinful to harm others intending the harm (cf. Hurka 2014a: 414). I leave the interpretation to Aquinas's scholars, but if the former is what Aquinas has in mind, he defends a rather different principle from contemporary friends of the DDE.

If this version of the DDE is the representative one, it needs to be amended so as to avoid prohibiting too little. Otherwise, it would not be able to account for the paradigm pairs of cases. Take, for example, *Bystander* and *Bridge*. While in *Bystander* I do not intend to kill the one in order to save the five, in *Bridge* I might intend it but typically not in revenge, or out of hatred, or some other bad motive. The version of the DDE would then imply that in *Bridge* the bystander acts permissibly. I see no way to amend that version to account for the paradigm cases without collapsing into the other one.

A further problem with it is that it might prohibit too much. Here, it all depends on which motives are taken as bad. Consider vainglory. A bystander might redirect the trolley intending to save the five, knowing that there is a significant risk of the trolley hitting him before it kills the one, but his desire for boasting about the sacrifice might be as big as to make him take the risk. Whether that action is permissible is the question that takes on the very plausibility of the DDE, which is why I shall postpone discussing it until the next chapter. I bring it up here to show that the defenders of the motive-version of the DDE cannot avoid specifying the scope and kind of motives that bear on the permissibility of actions.

2.9 Intending Bad Effects

This leads to an important question: what are the bad effects whose bringing about intentionally is prohibited under the classic definition? Since the DDE is a principle that lists the conditions under which an action becomes morally permissible, it cannot take the term ‘bad’ in ‘bad effects’ in the morally deontic sense but only to refer to non-deontic badness of certain events.

To begin with, we can rule out some consequentialists views that count an intention or a motive as bad just when and because it results in an action whose (predictable) consequences are suboptimal.²⁷ Some bad intentions and motives, such as the intention to kill a tyrant terrorizing a population purely out of revenge or hatred, lead to optimific results. Similarly, in *Bad Bystander*, everything is the same as in *Bystander*, but I redirect the trolley intending only to kill the one as he is my bitter foe. My intention leads to the best consequences and yet it is bad. Moreover, if the DDE were to take such a principle for delineating bad from non-bad intentions, it would end up allowing what it normally would not (e.g. *Bad Bystander*, *Bridge*) and prohibiting what it should not (e.g. *Runner*, as I could do greater good benefiting those in need instead of going for a run).

The DDE then needs a very different principle delineating bad from non-bad intentions. It could also take some events as bad for granted or go along with intuitions without justifying their badness. Both the defenders and the critics of the DDE usually do the latter. The first and the obvious candidate for a bad event is harm. Intention to harm other people should then count as a bad intention. The prohibition on intending harm, it is sometimes said, is what distinguishes the DDE from similar principles, such as the Kantian dictum that we should not treat other people as mere means, even if we treat them harmlessly (Lippert-Rasmussen 2010: 549-51, Persson 2013: 141).

Notice that ‘harming other people’ is ambiguous between causing harm *to them* and using or treating *them* in ways that are harmful to them. When, in *Self-Defense*, I kill the aggressor, I cause him harm with the aim of saving my life. But I do not use him, no more than I use falling rain when wearing a raincoat to protect myself from getting soaked (Parfit 2011a: 221-2). If I

²⁷ These should not be confused with the indirect consequentialist views which derive the rightness of actions from optimific motives and intentions (see Adams 1976, Sverdluk 2011).

were to use the aggressor, I would want his presence to help me achieve my goal. But my goal would be best achieved had he not been there (Parfit 2011a: 222, Persson 2013: 141, Kagan 1989: 140-2). However, in *Bridge* my goal of saving the five is best achieved if the large man is on the bridge. I therefore use him in the way that is harmful to him, besides causing him harm. Since the DDE would prohibit acting in both *Self-Defense* and *Bridge*, granted that my intentions are to kill, it generally must prohibit both acting with intention to use others in harmful ways and with intention to cause harm to them.

It is sometimes said that the DDE applies to causing right-infringing harms, such as the right of innocent, non-threatening people to live (Kamm 2007: 131, Øverland 2014: 481). This is true of many cases the DDE aims to explain, e.g. of *Tactical/Terror Bombing* and *Trolley/Bridge*. But in other cases it is not. In *Doctor/Bad Doctor*, the patient requests the doctor to kill him. Given his state, it is reasonable to think that the patient's consent suffices for forfeiture of his right to life, thereby giving the doctor the permission to kill him. Similarly, on some views (e.g. Thomson 1991) I get the permission to kill in self-defense if and because the aggressor forfeits his right to life by culpably endangering me. If those views are correct, and the DDE concerns only right-infringing harms, it has no implication in *Self-Defense*.

Another question is whether the DDE prohibits intending harm to oneself. The DDE is sometimes criticized on the basis that it would prohibit supererogatory suicide, such as that of a soldier who throws himself on a grenade to save his comrades. The argument is that, since the soldier must intend to kill himself as a means of saving others, the DDE prohibits the heroic act (Kagan 1989: 144-5, cf. Duff 1982: 9-10). Appealing to cases like *Chemotherapy*, we might say that the soldier need not intend to kill himself. But it might be replied that, if the defenders of the

DDE accept that we need not intend what we believe to be casually necessary to our ends, they must also accept that it is permissible to kill a fellow soldier as a means of saving the others and oneself. The inference is debatable as the DDE might make further claims about killing others as a mere means. (More about that in a moment.) Another route for the defenders of the DDE is to limit its application to cases of harming others. Since the current debate revolves only around such cases, I shall leave aside the issue of the DDE's application in cases of harming oneself.

A much more important class of effects that the defenders of the DDE should count as bad are those that are not harmful under the description the agent intends but that are harmful under any ordinary understanding of harm. Shooting a person point-blank is normally considered harmful since it nearly always results in death or great injury. And yet, a miracle could occur and the person shot could get away unharmed. We could then imagine someone who shoots a person in the head but does not intend to harm him but only to get the bullet through his head. Were the victim to end up unharmed, the shooter would rejoice. Since the DDE must prohibit such actions, it must take the shooter to intend bad effects. As we shall see, it is hard to find an account that succeeds in doing so while preserving the moral relevance of the intention/foresight distinction.

2.10 Intending Good Effects

On the second condition of the classical definition, it is not only intending bad effects that is forbidden but intending good effects that is required. This requirement is too strong. We might intend part of the good effects and merely foresee the other part. I might intend to redirect the trolley in order to save the four, foreseeing that I shall thereby save the five.

Moreover, the good effects can be merely foreseen side effects of those we intend, either in the sense of being caused by the latter or themselves causing them. I might redirect the trolley towards the one, foreseeing that that will save one person on the side track, which will then have the effect of saving five others behind him, but intend to save only the first of them (cf. Kamm 2007: 22, 117). I might also redirect the trolley, not intending to save the five but only to please their families, taking the saving of the five as a necessary means to it. Assuming that I do not intend the death of the one on the main track in none of those cases, the DDE should not censure those actions despite that I do not intend all the good effects.

2.11 Subjective and Objective DDE

The third condition states that bad effects are not means to producing the good effects. When it comes to this addendum, waters are muddied. In contemporary debates the DDE is understood in two different ways, which can be called subjective and objective (Bennett 1995: 199-200, Nye 2013: 260). The subjective DDE makes essential reference to the agent's intention and prohibits causing (or allowing) harm in pursuit of a greater good with the intention to harm as an end or as a means. The objective DDE prohibits only causing (or allowing) harm as an end (without good reason for it) or as a means to a greater good, whether or not the agent intends the harm.

This objective version is sometimes called the Means Principle. But that complicates the things even further, for two reasons. First, the objective DDE prohibits intending harm as an end, not only as a means. Second, the Means Principle might be understood subjectively, prohibiting intending harm as a means (cf. Tadros 2011b: 116, Walen 2014: 428-9). It is sometimes thought

that, properly understood, the DDE is the subjective Means Principle plus some other conditions, i.e. that it prohibits intending harm as a means but not as an end (McIntyre 2001: 226-9). But that cannot be true. Suppose that I throw the man off the bridge in *Bridge* intending only to kill him. I foresee that I shall save the five but I do not care about it. The DDE should clearly prohibit my action (Persson 2013: 143).

On other views, the objective DDE cannot be part of the doctrine since the DDE allows bringing about bad effects as a means to good effects when the former are not causally necessary for the latter. In *Track Trolley* (Kamm 2007: 93-4), all is the same as in *Bystander* except that, when I press the switch to redirect the trolley, the trolley goes onto the side track only after that track is moved down and the one is pushed towards the trolley so that his body collides with the trolley and brings it to a halt before it reaches the crosspoint. His death is a means to saving the five but it is not a causally necessary means; if he was not present on the side track, the five would be saved anyway.

The notion of means featuring in the DDE indeed seems to be that of causally necessary means, where causation might be understood counterfactually: something is a cause only if the effect would not have occurred were it not for the cause. The truth of the conditional might be analyzed using the standard possible world semantics. For example, although the five could be saved in *Bridge* even if the large man survived the collision with the trolley, the possible world in which that happens, in terms of comparative similarity, is far from the actual world. However, the worlds where the five are saved and the one is not harmed in *Bystander* or *Track Trolley* are close to the actual world. That is why the death of the one is a means, in the relevant sense of it, of saving the five in *Bridge* but not in *Bystander* or *Track Trolley*.

However, if the defenders of the DDE find redirecting in *Track Trolley* impermissible, they will want to broaden the notion of means. Moreover, that restriction makes the DDE excuse too much (Kagan 1989: 143). Here is a more obvious case. Suppose that there were two men on the bridge in *Bridge* and I pick just one and throw him off. His death is contributing to saving the five but only in the sense that it is causally sufficient but not necessary for it. The possible worlds where the five are saved and he is not harmed because I throw off the other man for the same purpose are close to the actual world. If the DDE prohibits only harming as a causally necessary means, it does not prohibit my action in this version of *Bridge*, which is implausible.

To find an appropriate balance between these notions of means is the Procrustean task for the defenders of the DDE. The safest option for them is to take the DDE to prohibit harming as a means in any causally contributory manner. *Track Trolley* or self-defense against innocent shields (Kagan 1989: 138-9, Uniacke 1994: 120, n. 36) would then be impermissible according to it.

Getting back to the distinction between the subjective and objective DDE, we can see that their implications are strikingly different. The subjective DDE would not prohibit pushing the large man in *Bridge* if one could do it without intending to kill the man. And the objective DDE would not prohibit redirecting the trolley in *Bad Bystander*. The DDE is expected to prohibit such actions, so it has to combine both variants. This is exactly what the classical definition, on the biconditional reading it, does. But that comes at a price.

On the classical definition, the DDE sets four conditions jointly sufficient for an action to be permissible. This gives an easy target to the critics of the DDE as they can refute it showing that an action is permissible if any of the conditions is not fulfilled; e.g., if one intends harm or if one causes it as a means. An illustration of the latter is the well-know *Loop* case, in which

everything is the same as in *Bystander*, except for a side track where one person is trapped which loops back towards the main track from the other direction, where the five are trapped. If I redirect the trolley, it will kill the one but the trolley will come to a halt, so the five will be saved. I would therefore kill the one as a (causally necessary) means of saving the five (Thomson 1985: 1402-3). If the action is permissible, as it seems to some people, the objective DDE is false.

However, I need not intend to kill the one as a means of saving the five even if I in fact kill him as a means. As already explained, we need not intend the means we take as necessary for achieving an end. However, if the DDE is a conjunction of the objective and the subjective DDE, *Loop* suffices to refute it. The defenders of the DDE can of course argue that it is impermissible to redirect in *Loop*, an option I discuss in the second chapter. But they have two other options: to defend the subjective DDE as *the* version of the principle, not allowing for the possibility of not intending to kill in *Bridge*, or to restate the DDE disjunctively, as taking the subjective and the objective condition as each necessary for permissibility. The DDE would then state that an action is impermissible if either bad effects are intended or if they are caused as an end or as a means to good effects. If they succeed to defend either option, *Loop* becomes an irrelevant or insufficient counterexample.

Unfortunately, going for the disjunctive reading has an unwanted effect of prohibiting any harming as a means by itself and not only together with intention to harm, including those kinds of harming as a means of benefiting, such as some surgeries or self-defense (assuming again the biconditional reading of the classical definition). Since the DDE must not prohibit such actions, the only plausible way to understand the third condition is that harming as a means and intending to harm as a means are not individually but jointly necessary for permissibility. In other words, it

is best if we take the DDE to be the conjunction of its subjective and objective variant. Its friends should then take *Loop* seriously. In the next chapter, I shall criticize the DDE thus understood, arguing, among other things, that *Loop* is indeed a powerful counterexample.

2.12 Due Proportion

The fourth condition, or the requirement of due proportion or the proportionality condition, as sometimes called, distinguishes the DDE from a similar although implausible principle licensing unrestrained harm in pursuit of a good as long as one does not intend the harm. However, when it comes to interpreting it, the waters are no less muddled than with the third condition. There are at least three ways to make sense of the claim that the good effects must be sufficiently important to justify bringing about the bad ones.

First, the condition might require weighing the good against the bad effects (of acting and omitting to act); if the former outweigh the latter, the action is justified (Donagan 1977: 161, cf. Woodward 1997). On this reading, the proportionality condition subsumes under its scope the so-called necessity condition, according to which one is required to pursue the least harmful option from the available ones; this is clearly true if all other options are ruled out as suboptimal. Thus stated, however, the DDE is but a form of maximizing act-consequentialism, constrained by the prohibitions on intending bad effects and causing them as a means. This makes it prone to some standard objections to act consequentialism (McIntyre 2001: 222-3, Kamm 2007: 22).

On one of those objections, the DDE cannot make room for the DDA, the principle that doing harm is worse than allowing it. By itself, the DDA is not sufficient for permissibility; if it

were, it would prohibit redirecting the trolley in *Bystander*. It needs to be supplemented at least by the DDE. Some people (Kagan 1989: 154-8, Woollard 2015) claim that for similar reasons the defenders of the DDE might take the DDA as part of the proportionality condition. The doctrines would then form a deontological unity, standing or falling together. This is the second reading of the proportionality condition.

Since the defenders of the DDE cannot fit the DDA into the proportionality condition by endorsing an absolutist constraint on causing merely foreseen harm, as that would undermine the very basis of the DDE, they need some threshold-type constraint, prohibiting causing such harm if the good does not outweigh it to a certain degree – for instance, if it is not at least several times the size of the harm (Kagan 1989: 157). What is more, since the DDE has to permit causing harm in *Bystander*, we can assume that the bad effect of one person being killed, under the proposed ratio, is outweighed by the good effect of five people being saved.

However, there are many instances of merely foreseen harming of one person as a side effect of saving the five which are impermissible. In *Driver*, rushing to the hospital to save five people, I foreseeably but unintentionally run over and kill one person on the road, there being no other road to take or the way to avoid running over that person. Similarly, in *Gas*, I use some gas as the only way of saving the lives of five patients, foreseeing but not intending the death of one irremovable patient by toxic fumes spreading into his room (Foot 2002a: 29, 2002b: 80-1, Kamm 2007: 22, 131-2, 433-4). Such cases are used to show that even the revised DDE will not suffice.

There is another way to cash out the proportionality condition that exhibits all the virtues of the previous two but avoids their vices. It states that the other conditions get off the ground only when there is a sufficient reason to do the action (Boyle 1980: 528, n. 7, 1991a: 476-8, cf.

Anscombe 2005: 231). The deliberate vagueness of this interpretation is to its benefit because it avoids consequentialism of the first interpretation and upholds deontological constraints of the second. If, as in *Driver* or *Gas*, an action is intuitively impermissible, this interpretation implies that the DDE is out of business as there is a conclusive reason against the action.

One might worry that stretching the proportionality condition as much as to include any morally sound principle makes the DDE obsolete. But as long as its other conditions withstand criticism, the DDE remains relevant. The soundness of those conditions is therefore of primary importance.²⁸

That being said, we ought to be cautious not to appeal to the proportionality condition in an ad hoc manner, especially when the other routes to save the DDE become closed. At the very minimum, the DDE should explain its paradigm cases. The debate would make no sense if, say, the relevance of the constraint on intending harm and on causing it as a means was questioned in *Bridge* but the defenders of the DDE replied that this shows that, after all, the DDE does not even get off the hook in *Bridge* as there was never a sufficient reason to act. If due proportion is to be understood as there being a sufficient reason to act, in terms of mere structural similarity *Bystander* must be *Bridge* minus the two constraints, *Strategic Bombing* must be *Terror Bombing* minus the constraints, and so on for the other pairs. If the pairs do not seem similar in this way, they either were not good candidates to be explained by the DDE (or, perhaps, by no principle of that kind) from the start or the proportionality condition needs to be understood differently. My argument in the second chapter will suggest the former.

²⁸ As Quinn (1989: 334, n. 3) puts it: “[I] am concerned here not so much with how choices with a ‘second effect’ can be justified as with whether, *ceteris paribus*, the structure of intention makes a justificatory difference. That seems to me the fundamental question.”

2.13 Actualist and Possibilist DDE

Some people (FitzPatrick 2003, 2012) argue that the DDE is not concerned with what an agent actually intends but with what he could intend. On this view, an act is permissible if it is possible for the agent to act with the intention the DDE permits. Whether, say, he actually intends harm is not what matters. This implies that in *Bad Bystander*, for example, redirecting is permissible even if I redirect it intending to kill the one because I could do it intending to save the five. Call this the possibilist DDE, as opposed to the traditional, actualist, one.

One could defend the possibilist DDE while also thinking that other conditions matter. On one version of this view (FitzPatrick 2012: 112-8), actual intentions affect the permissibility of such actions only if the agent who intends harm does not let benevolent others do the same without intending harm, e.g. I refuse to let you redirect the trolley intending to save the five and do it intending to kill the one.

That specificity aside, why would the permissibility of acting depend on what one could or could not intend? A possible reply is that an action is impermissible if there is no adequate justification of it and a justification is adequate only if one could act on the reasons that justify an action without thereby intending harm, since intending harm is a wrong *per se* (FitzPatrick 2003: 320, 2012: 102-6). The problem with this reply is that, even if we grant that intending harm is a wrong *per se* and that a justification is inadequate because of these reasons, it does not follow that an action is impermissible because one could not but intend harm. It is a contingent fact that impermissible harmful actions are performed with intentions to harm. But the reason why they are impermissible is not the fact that they could not be done without such intentions but that they

are harmful in morally problematic ways. One of those ways might be the fact that one intends harm but this is exactly what is denied on this view. The appeal to possible intentions is at best trivial and at worst superfluous.

Conclusion

This ends my interpretative inquiry into the nature of the DDE. We saw that there is no one but a variety of often opposing principles understood as the DDE and assumptions comprising it.

Where does all this leave us? Can we distill the best version into one definition? I believe that something like the following principle represents the hopes of the defenders of the DDE and is well worth criticizing: It is morally permissible to cause or allow harm to others, or something close to harm given the best theory of it (harm*), in pursuit of a good if and only if, or just when (1) there is a sufficient reason to pursue the good, and (2) the harm or the harm* is not a causally contributory means to the good, and (3) the harm or the harm* is not intended by the agent, either as a means or as an end, whatever the motive behind it, but is a merely foreseen side effect. In the next chapter, I shall defend three objections to the DDE understood in this broad way.

2. Double Effect Defects II: Critique

Introduction

I shall defend three objections to the DDE introduced in the previous chapter. As explained, they have a common history and dialectic.

The first objection is that the constraints on intending harm and on harming as a means are best understood in terms of reasons not to do these things and, when considered in a variety of cases, these reasons do not intuitively seem to outweigh the competing ones counting in favor of actions. The second objection is that there is no plausible account of the difference between intending harm, intending that which is close to harm, and foreseeing harm such that it tracks the DDE's implications about permissibility in the relevant cases. The third objection is that those cases are neither explained by the fact that one intends (or does not intend) harm nor by the fact that one harms (or does not harm) others as a means.

I shall call these objections *the Strength-of-Reasons Objection*, *the Closeness Objection*, and *the Explanatory Objection*, respectively.

1. The Strength-of-Reasons Objection

The Strength-of-Reasons Objection originates from the Intuitive Objection, which states that the DDE intuitively prohibits too much. Similarly, the Strength-Of-Reasons Objection states that the constraint on intending harm (or broadly, bad effects) is indefensibly strong because it prohibits causing or allowing harm in pursuit of a good, regardless of its kind or gravity, if one intends it. Importantly, the objection states that we can back up that intuition if we think of the constraint on intending harm in terms of reasons not to intend harm or not to act with intention to harm,²⁹ and, assuming that those reasons count against an action's permissibility, weigh them against reasons counting in favor of it (cf. Kagan 1989: 134). We should then intuitively realize that, in relevant cases, the former do not come out as stronger than the latter. The same is true of the nature and the strength of the constraint against harming as a means.

1.1 The Possibility of Reasons against Intending Harm

Spelling out the constraint against intending harm in terms of reasons against intending harm can be questioned arguing that there are no such reasons.

On some views (Kamm 2008: 168, 2014: 380, Rachels 1986: 92), there are no reasons against intending harm since, when one intends harm, an action one would perform, if otherwise permissible or required, does not lose any of its right-making properties. The reasons counting sufficiently or conclusively in favor of it do not disappear. One can then decide whether or not to act on the basis of those reasons only, leaving the intention he would act with aside.

²⁹ I use the two interchangeably. One can of course intend something and not act, as in case of incomplete attempts, but I leave those cases aside here.

That argument begs the question. The issue is not whether reasons that count in favor or against an action disappear once one intends harm, as they obviously do not, but whether they are the only reasons that bear on the permissibility of that action. The DDE suggests that reasons against intending harm (or acting with intention to harm) are also at play. We then need to ask how they feature in the overall balance. If the DDE is true, they outweigh the other reasons, so one must take them into account in deciding what to do.

On another view (Scanlon 2008: 58-62), reasons against acting with intention to harm can exist only if one can act but not with that intention or only if one can act for a reason other than to harm (as intentions are based on one's reasons for acting). But since one cannot choose what to believe is a reason for acting, the argument goes, one cannot choose one's reasons for acting. One cannot act for a reason other than the reason to harm, if one already believes that that is the reason for acting, nor can one act but not with intention to harm, if one already intends harm. Consequently, one cannot be required to act but not with intention to harm but only not to act, *simpliciter*. Reasons against acting with intention to harm do not exist above and beyond reasons against acting, *simpliciter*.

The problem with this argument is that, although one cannot choose what to believe is a reason for acting in the sense of being able to resist believing the evidence one finds convincing, one can choose what to believe in the sense of looking for, assessing, and being exposed to new evidence. This makes it true that one can believe otherwise. If one is convinced that harm is the reason for acting, he has the ability to reassess it and act for a different reason. Moreover, since he had that ability prior to taking harm as the reason for acting, he could have known better. He therefore could have had a reason not to act with intention to harm when he acted with it.

It might be also argued that, even if reasons against acting with intention to harm exist, they must be derived from other reasons, such as those not to harm. For example, a victim aware that he will be harmed with intention to harm might be harmed more, just by being aware of it, than a victim aware that he will be harmed as much but with no such intention. However, the cases the DDE deals with are contrived in the respect that all derivative relevances of intentions are blocked in them. Besides assuming that the consequences of contrastive actions are the same, we are to assume that victims who are targeted with bad intentions are not aware of it (cf. Kamm 2014: 482-3). On this view, then, there are no grounds for reasons against acting with intention to harm in such cases.

It is true that the one on the side track is as dead in *Bystander* as in *Bad Bystander* (where I redirect the trolley only with intention to kill him). But it is also true that it is not different to be killed as a side effect than to be killed as a means, other things being equal.³⁰ In terms of how victims fare, *Bystander* is indistinguishable from *Bridge*. However, for non-consequentialists, the gravity of harm is not all that matters for permissibility. The ways in which victims are harmed are crucial. Therefore, to deny that a victim is wronged more by being harmed as a means than as a side effect since there is no wrongdoing over and above harming, period, is to beg the question against the DDE. The same goes for such a critique of the constraint on intending harm.

One might reply that the DDE is incomplete without an explanation why intending harm is wrongdoing. I agree. As said in the previous chapter, one way to attack the DDE is to deny that there is a moral difference between intending and merely foreseeing harm by criticizing the best

³⁰ This includes the absence of knowledge caveat. Were it to be lifted, the victim might prefer to be killed as a means than as a side effect, saying that in that way his death would at least do some good (cf. Parfit 2015b: 13-4). It also assumes that the term ‘harm’ is in the limits of its ordinary meaning. On some views (e.g. Duff 1996: 338-9, 2007: 128), intentional harming is greater suffering for the victim even if he is unaware of being harmed in that way.

justifications of the distinction.³¹ However, what I am interested in here is whether, given that there is a genuine difference between the two and that to intend harm is to wrong *per se*, the constraint against intending harm stands in the way the DDE implies.

Some defenders of the DDE might object to defining the constraint on intending harm in terms of reasons not to intend harm, saying that impermissibility is unanalyzable in terms of reasons or any other notion. But that impermissibility is indefinable need not imply that there is no explanation for an action's impermissibility but only that such an action has the property of 'must-not-be-doneness' (Parfit 2011a: 170). The property might exist due to another property, e.g. of an action being done with intention to harm. We could then ask why the latter property grounds the former and whether that is true across cases. The debate would boil down to issues I turn to below. I doubt that 'must-not-be-doneness' is a property of any action. I shall then assume the moral-decisive-reason sense of impermissibility. But nothing of importance hinges on that assumption.

1.2 The Nature of Reasons against Intending Harm

Let us then suppose that there *could* be reasons against intending harm. Some people (e.g. Heuer 2014: 29-30) claim that those reasons might be exclusionary, excluding reasons counting in favor or against an action rather than outweighing them. Exclusionary reasons are reasons to refrain from acting for reasons, or on the balance of reasons, that one would otherwise had and act on.

³¹ There are three groups of justifications: that intending harm affects one's agency for the worse (Nagel 1986: 180-4, Wedgwood 2011a: 392-3, 397), that it wrongs the victims (Quinn 1989: 348-51, Cavanaugh 1999, 2006: 147-58, Tadros 2011b: 130), that it increases impersonal badness in the world (cf. McMahan 2009a: 357-8). I think they are all untenable.

The excluded reasons are defeated if the exclusionary ones are undefeated but their balance is not thereby altered. The function of exclusionary reasons is to prevent acting on the balance of reasons but it is not to change that balance, tipping the scale on one side or another (Raz 1999: 35-48, 182-94).

For instance, the balance of reasons in *Bystander* is on the side of redirecting. I have enough reason to redirect. But suppose that I promised the one on the side track that, if we ever find ourselves in that situation, he as the victim and I as the bystander, I shall not redirect the trolley. The promise, some would say, makes it true that I should disregard the balance of reasons favoring redirecting. They do not cease to be in force but I should not take them into account in decision making. Authoritative directives, advices, and incapacities are some other examples of exclusionary reasons. Exclusionary reasons are not undefeatable (Raz 1999: 27-8, 40). If there were a million people on the main track, the promise would not be a defeater.

Reasons against intending harm are not, I believe, exclusionary. It might be thought that they could not be exclusionary since responding to them entails refraining from acting with the problematic intention. That would be a mistake. Responding to exclusionary reasons might result in inaction, as illustrated by the example above. Nonetheless, I do not see anything in the nature of intending harm that would make it the case that the one who intends harm would be obliged to disregard other considerations rather than to weigh them against the fact that he intends harm. Reasons not to intend harm are obviously unlike reasons to keep promises or reasons to follow authority. They also lack the pragmatic rationale of incapacities, such as tiredness. If the DDE is true, it seems more natural to say that reasons against acting with intention to harm outweigh the reasons in favor of acting rather than they preempt considering them. That being said, nothing of

what I argue below would change if reasons against intending harm were exclusionary. We could still ask whether they are defeated by other reasons in certain cases.³²

1.3 The Strength of Reasons against Intending Harm

Against this background, we can now consider some standard cases. In *Self-Defense*, a culpable aggressor tries to kill me. I kill him first, justifiably believing that there is nothing I can do to avoid getting killed or any other way to defend myself. Since I kill him as a means of saving my life, the objective version of the DDE as well as the version that takes the constraint on harming as a means and the constraint on intending harm as jointly necessary for permissibility imply that I act impermissibly, even if I do not intend to kill the aggressor. We may not kill in self-defense unless we kill as a side effect.

Aquinas did not believe this. Nor do many of us believe it, apart from some pacifists. We oppose such views because we think that reasons not to kill are not of absolute strength. Indeed, self-defense is the paradigm exception to the principle that forbids causing death or grave injury to others. There are different explanations for it. On some views (Thomson 1991, Uniacke 1994), by trying to kill me, the aggressor forfeits his claim not to be killed. On others (McMahan 2005), he is liable to defensive killing in virtue of being morally responsible for trying to kill me.

For the purposes of this discussion it does not matter which of these views, if any, is true. If the reason not to kill as a means is outweighed by the reason to save one's life from a culpable

³² What might lead some people to believe that reasons against intending harm are exclusionary is that they are both second-order reasons. Exclusionary reasons are reasons not to act for a reason and reasons not to intend harm might also be reasons not to act for a reason. However, the reasons that exclusionary reasons require abstaining from are normative reasons but the reasons that reasons against intending harm require abstaining from are operative reasons, i.e. what one takes as reasons for acting, since those reasons are the grounds of intentions.

aggressor, we can ask whether the reason not to intend killing as a means is outweighed by it too. Some people (Kagan 1989: 133) think that I must intend to kill the aggressor in *Self-Defense*. I disagree, and so would Aquinas. Suppose, however, that I do intend to kill him as a means of defending myself. On the subjective and the disjunctivist DDE, I act impermissibly.

Anscombe (1981: 68) writes: “To aim at killing, even when one is defending oneself, is murderous. (I fear even this idea is going. A man was acquitted recently who had successfully set a lethal booby trap to kill a thief in his absence.)” There are two reasons why Anscombe’s fear is justified. First, to kill only for the sake of defending private property, in most, if not all, cases, is to inflict disproportionate harm and, thus, to act impermissibly. Second, we have good reasons to forbid preemptive defensive killing of even culpable aggressors since, in many cases, we thereby not only risk harming them disproportionately but rule out the possibility of averting their threats in non-harmful ways.

However, it is a mistake to think that defensive killing with intent to kill is murder, if it is, intention-independently, identical to a permissible defensive killing. Those who think that killing as a means is wronging *per se*, do not think that it overrides the right of an innocent victim to kill a culpable aggressor. Similarly, even if they think that intending to kill the aggressor is wronging him *per se*, they should not think that it suffices to override the right to defend against his attack. The reasons not to intend to kill the aggressor should not, in other words, outweigh the reasons to kill him. I see no way to defend the asymmetry in strength of the two reasons. One might reject the asymmetry, denying that reasons against harming as a means can be outweighed and end up

in a near-absolute pacifism, or accept the subjectivist DDE, thus significantly limiting the DDE's explanatory potential. Either way, these are some non-trivial bullets to bite.³³

Similar reasoning applies to cases where reasons not to intend harm conflict with reasons to act upon one's consent. In *Bad Doctor*, recall, a terminally ill patient in anguish, and with a life no longer worth living, requests to be killed. The only available doctor kills him with intent to kill (say, motivated by hatred). Assume that he wrongs the patient. Does he act impermissibly? Imagine first that we ask the patient whether he consents to be killed with that intention and he replies that he does. It is hard to believe that his consent to be killed with intention to harm along with his anguished state does not outweigh the fact that the doctor intends to kill him. Even those who find harming as a means impermissible would not think that, if the patient consented to be killed as a means, the doctor would not be permitted to kill him as a means of relieving him from suffering. Actual consent has no force when it rests on a false belief or when the agent giving it suffers from cognitive or other incapacities but this is no such case.

Suppose now that the patient refuses to be killed with that intention. Sometimes we may override one's actual refusal to consent that does not spring from a false belief or an incapacity if we thereby make the person much better off. To see whether the patient who consents to be killed with no intention to kill could also consent to be killed with that intention, we therefore need to weigh the reasons to respect his refusal to consent, the reasons to prevent his suffering, and the reasons not to intend to kill him. Starting with the first two reasons, we could reach the view that prevention of suffering does not override the refusal to consent if it involves the victim's death. If

³³ Some people (e.g. McMahan 2009b) think of the morality of conduct in war as an extension of the morality of self-defense. They might want to take this line further and argue that killing combatants on the unjust side is permissible even if the combatants on the just side intend to kill their foes and/or are motivated by hatred. Interestingly, hatred towards Germans was one of the main motives for Soviets to fight in the WWII (see Reese 2011: Ch. 8). I think that, other things being equal, they fought permissibly.

that is correct, the reason why the doctor may not kill the patient with intention to kill is not the intention itself but the patient's consent vetoing the procedure.

We could also think that prevention of suffering outweighs the other considerations. This is the least controversial in cases where we cannot ask the patient for consent—perhaps he is and will remain unconscious—and all other relevant facts are as before (including his general request to be killed). Since most current defenders of the DDE are not moral absolutists, they should agree that the reason not to intend harm does not outweigh the countervailing ones, whatever their strength. If the patient's life is no longer worth living, and there is a conclusive reason to terminate it, then that reason outweighs the reason not to do it intending to kill him. The patient should not be left in such a state even if the doctor's intention is to kill (cf. Thomson 1999: 514-6, McIntyre 2001: 225-6, FitzPatrick 2012: 116-117).

When Quinn (1989: 343, n. 17) discusses such cases, he shares the view that the doctor may proceed despite his intention but explains it differently, saying that the DDE is misapplied in such cases as they do not involve conflicting moral claims of different people. The only relevant claims are those of the patient. However, if the DDE is true, the patient has a moral claim not to be killed with intention to kill. The DDE then fails because it assigns excessive moral importance to the constraint against intending harm and not because it does not apply in non-conflict cases.

Still, Quinn might have thought the DDE to apply to conflicting moral claims apart from the claims not be harmed with this or that intention. Similarly, it is sometimes suggested that the DDE might be restricted to harming non-culpable, non-threatening, and innocent persons, who would not be otherwise comparably harmed, who have not consented to it, and for whom further life is a good, in course of bringing about a greater good to other such people (see Kamm 2007:

130-1, Kagan 1989: 146). Three kinds of cases are relevant here: those where the victims are harmed as a means but the agent does not intend the harm, those where they are harmed as a means but the agent does intend the harm, and those where they are harmed as a side effect and the agent intends the harm. *Loop*, as I interpret it, is an instance of the first, and, as some people interpret it, is an instance of the second, while *Bad Bystander* is an instance of the third.

I have already explained why I need not intend to kill or harm the one in *Loop*, so I shall first assume that what is morally problematic about it is killing as a means. It might be said that what explains this is the fact that harming innocent people as a means of benefiting others is always impermissible. However, if you, your child, and me were trapped in an earthquake and the only means to save your life and the life of your child were to break my arm, you would be permitted to do it, even if I would not consent to it (cf. Parfit 2011a: 231). The claims of you two to save your lives, taken together, would outweigh my claim to get away unscathed.

Since the same would not be true were my life to be at stake, it might be suggested that it is only killing as a means, rather than imposing great injury, that is always impermissible. But, as already said, most friends of the DDE are not moral absolutists. If there were million or even a thousand or a hundred people on the main track in *Loop*, they would concur that it is permissible to redirect.

On that note, consider *Six Behind One*, a variant of *Bystander* where there are six people behind the one on the side track, so that, if the trolley is redirected, it will hit and kill him, which will save the six as a means and the five as a side effect (Otsuka 2008: 100-1). While I need not redirect the trolley intending to kill the one, I must kill him as a means of saving the six since his death is a causally necessary means to it. Yet I would, I believe, act permissibly if I did it. Since

there is no wronging in killing the one as a side effect of saving the five on the main track, he has no claim not to be killed as a side effect. He only has a claim not to be killed, period, but that claim is outweighed by the joint claims of the five to be saved, just as in *Bystander*. We are then left with claims of the five to be saved against the supposed claim of the one not to be killed as a means. It seems clear that that claim does not outweigh the others.³⁴

The same would be true, I believe, in *Five Behind One*, an identical case with one person less behind the one. If so, and if there is no morally relevant difference between *Five Behind One* and *Loop*, it is permissible to redirect in *Loop* for the same reason. We now work under the assumption that I do not intend the death of the one in *Loop*, so the difference between the cases cannot be in what I intend. It might be said that, in *Five Behind One*, I kill the one both as a side effect and as a means but not so in *Loop*. That is true but, since killing one person as a side effect does not *per se* wrong him, it should not affect the way we judge the two cases morally. It can be also said that while redirecting in *Loop* is permissible but not required, redirecting in two other cases is required. However, if it is merely permissible to let the five get killed in *Bystander*, it seems merely permissible to do so in *Six Behind One* and *Five Behind One*.³⁵

We can reach the same conclusion considering another case. In *Loop II*, all is the same as in *Loop* except that the one is not trapped on the side track but before the five on the main track and after the crosspoint. If I do nothing, the trolley will hit and kill the one, which will then save the five, and if I redirect, the trolley will get to the side track that loops back behind the five, hit

³⁴ It might be objected that it is not correct to say that the six are saved as they were not in danger before redirecting. But we would say that the one on the side track in *Bystander* is saved if, after redirecting, I had the option to press the switch again and return the trolley to the main track. Nothing important hinges on this: the argument runs if the six continue to live rather than get saved as a means of the one's death.

³⁵ It might be also said that in *Five Behind One* I cumulatively save ten people, and in *Loop* five. But would the defenders of the DDE want to claim that it is permissible to redirect in *Loop* if ten people were on the main track?

and kill them, stopping just before it hits the one. Since the DDE prohibits not only killing but letting die as a means, it should require redirecting in *Loop II*. In doing so I would kill the five as a side effect of saving the one from being killed as a means. The DDE must then assume that the reason not to let killing as a means is five times stronger than the reason to kill as a side effect.

That balance is, I believe, clearly mistaken. To see why, consider *Four on the Side Track*. In that case, one person is trapped before five others on the main track but after the crosspoint and four persons are trapped on the side track. There is no looping. Doing nothing means letting the one getting killed as a means of saving the five while redirecting means killing the four on the side track as a side effect of saving the one and the five, i.e. of saving six people. If letting die as a means is five times worse than killing as a side effect, assuming no other considerations are relevant, I am required to redirect. But I think it is obvious that I may let the one get killed as a means, thereby cumulatively saving nine people.

If it is permissible not to redirect in *Loop II*, it is permissible to redirect in *Loop*, under the assumption that killing as a means is not that much worse from letting die as a means so that, other things being equal, the latter action is impermissible. Some might say that that assumption is unwarranted. If so, it is killing, period, that must be that much worse than letting die, period. I see no reason for the asymmetry between killing and letting die as a means, on the one hand, and killing and letting die as a side effect, on the other.

We might then ask how stronger is our duty not to kill in comparison to our duty not to let die. Since it is permissible to kill one person to save five others in *Bystander*, it is less than five times stronger. But then our duty not to kill as a means must be also less than five times stronger than our duty to let die as a means. It is then permissible to kill the one to save the five in *Loop*.

We can now suppose that I do intend to kill the one in *Loop*. It is sometimes suggested that I must intend to kill him because I have no reason to redirect except to stop the trolley from killing the five by making it hit the one (Otsuka 2008: 100). One reply is that I need not intend to kill him since I need not adopt other ways to make it true that the trolley hits the one if the initial hitting fails (Kamm 2007: 97). However, this presupposes that intending to do something implies adopting as many ways as needed to do it, which, as argued above, is false as some intentions are conditional on initial settings. I might intend to kill the one but only under the condition that the trolley kills him in the initial hitting. A better reply, illustrated by *Chemotherapy*, is that, even if my only reason to redirect is to save the five by the means of killing the one, I need not intend the means I take to be necessary for my ends, so I need not intend to kill the one.

Be that as it may, suppose that I do intend to kill the one because I do not care about him. On one argument (McMahan 2009a: 359), redirecting in (this version of) *Loop* is impermissible because I would not free the one from being killed by the trolley if I could. Imagine that I could press the button which would free him after I redirect the trolley but before it hits him (*Loop-Option*). Not only I am not required to free the one in *Loop-Option*, I believe, I am not even permitted to do it. If I am required to let the one get killed as a means of saving the five in *Loop II*, then I am required to do it in *Loop-Option*. I could fail to save the five in the former case by redirecting the trolley away from the one and in the latter by freeing him but those differences are morally irrelevant.

It can be replied that *Loop-Option* is relevantly similar to *Bridge* (McMahan 2009a: 359). Both are cases of killing innocent persons as a means of saving five others. If the argument just given is sound, *Bridge* must be different in a morally important way. We can first rule out the fact

that, in *Loop-Option*, I move the threat towards the person while, in *Bridge*, I do the reverse (cf. Thomson 1986: 83-4, 103). I fail to see why, other things being equal, initiating a new threat is morally worse than redirecting an existing one. I am not more culpable in doing the former. In *Loop-Option*, I make it the case that fewer people are killed but so do I in *Bridge*, and I do not cause the trolley to head towards the five in neither case. Suppose also that, in *Loop-Option II*, I can hit the trolley with the bomb which would make it explode after redirected on the side track but the one would be killed not by the trolley but by the shrapnels (cf. Parfit 2015b: 37-8). It is a mistake to think that it would be more wrong to kill the one in *Loop-Option II* because he would be killed by a threat I introduced rather than by the one that had already existed.

If *Bridge* is different, it is for another reason. It seems to me that the reason is to be found in a general prohibition on causing people to fall from bridges, whether they are thereby killed intentionally or merely foreseeably, or as a means or as a side effect. I expound on that thought in the last section.

The remaining question is whether intending to kill the one in *Loop* would tip the balance against redirecting, assuming that killing him as a means would not. If the claim of the one not to be killed as a means of saving the five does not outweigh the claims of the five to be saved, then I doubt that his claim not to be killed with the intention to kill as a means does. To support this doubt it pays again to compare *Loop* with *Loop II*, this time assuming that I intend the one's death in both cases. As before, I find it hard to believe that I ought to redirect in *Loop II*, thus avoiding intentionally letting the one die as a means at the expense of killing the five. Similarly, I see no good reason to avoid intentionally killing the one as a means at the expense of letting the five die in *Loop*.

We can finally consider the strength of reasons against intending to kill as an end on the example of *Bad Bystander*. There is an intuitive disagreement about this case (see e.g. Kamm 2007: 132-3, Liao 2012: 716). Some of those who think that it is impermissible to redirect might have been misled by cases where killing as an end is independently impermissible. For instance, Anscombe (1981: 66) claims: “Naturally, killing the innocent as an end in itself is murder (...); but that is no more than a possible future development for us: in our part of the globe it is a practice that has so far been confined to the Nazis”. On this view, any killing of innocents as an end is impermissible. But we can distinguish killing an innocent as an end with a side effect of achieving greater good from killing an innocent as an end without that side effect. Apart from moral absolutists, everyone believes that killings of the former sort are sometimes permissible. Therefore, the issue about *Bad Bystander* is whether the good done in it suffices to justify killing the one as an end.

It might be said that it does not, even if we accept that the reason not to intend killing as a means does not outweigh the reason to save five people, because intending to kill as an end is much worse than intending to kill as a means. I doubt that this is true. For one thing, the attitude towards the victim is very similar in both kinds of intending. If I really intend to *kill* the one in *Loop*, I must think that it is an instrumentally good thing that he is hit and killed by the trolley. I need not hate him to believe so but I must have a positive stance towards his death. That attitude is not as bad as the hatred that motivates me to kill the one in *Bad Bystander*, but the difference cannot be as big as to make a difference in permissibility.

On one view (Liao 2012: 715-6), although the intention to kill as an end does not render an otherwise obligatory action impermissible, it does render an otherwise merely permissible

action impermissible. Reasons against intending to kill as an end do not outweigh conclusive reasons in favor of acting but they outweigh merely sufficient reasons in favor of it. Therefore, since I am permitted but not required to redirect in *Bystander* (as generally believed),³⁶ I am not permitted to redirect in *Bad Bystander*.

If what I said about the permissibility of redirecting in *Loop* and similar cases is correct, reasons against intending to kill as a means do not outweigh even merely sufficient reasons in favor of acting, as *Loop* and similar cases are instances of permissible but not required actions. If reasons against intending to kill as an end do outweigh such reasons, intending to kill as an end is worse than intending to kill as a means. I have just expressed doubts about that. Yet, intending to kill as an end might still be just a bit worse, enough to render an otherwise merely permissible action impermissible.

However, the reason why there is only a sufficient reason to redirect in *Bystander* is the bystander's claim not to, as it were, have blood on his hands by redirecting the trolley. He may refuse to become a killer, even a justified one (cf. Williams 1973: 108, Walen and Wasserman 2012: 554). If the agent's claim were not valid, the claims of the five to be saved would outweigh the claim of the one to be spared, so he would be morally obliged to redirect. The claim is not absolute and there might be disagreement about its strength but I shall assume, as most people, that its strength is comparable to the strength of the claims of the five. Notice, however, that if I decide to redirect the trolley, my claim against becoming a killer is no longer present. I face only a conclusive reason to redirect.

³⁶ Thomson (2008) now believes that redirecting is impermissible. For a good rebuttal see Kamm 2015: Ch. 1., cf. FitzPatrick 2009, Walen and Wasserman 2012.

In *Bad Bystander*, I am willing to redirect; indeed, I am eager to do it since I intend to kill the one. I therefore have no claim against having blood on my hands. I am left with a conclusive reason to redirect and a reason not to do it due to my intention to kill. So even the view under consideration must imply that the former reason defeats the latter. Redirecting the trolley with intention to kill the one as an end is, therefore, permissible.

The reasoning applies to any killing which is merely permissible due to such agent-centered reasons. If there are merely permissible killings which do not fall under this scheme, they are to be considered on their own. I conclude that, in all the cases around which the debate about the DDE concentrates, reasons against intending harm do not outweigh the competing reasons.

1.4 Objections and Replies

I shall now consider four objections to the arguments offered. The first states that, since I argued that reasons against intending harm as a means and against harming as a means are at best very weak, I should either accept that *Terror Bombing*, *Transplant*, and *Bridge* are permissibility cases or provide an alternative explanation of their impermissibility. I take the second horn but I shall postpone giving an alternative account until the last section.

Another worry is that my target was only the absolutist DDE. I considered cases where the constraints against intending harm and against harming as a means are supposedly as strong as to prohibit one to act but I did not consider the possibility that the constraints make the actions

only harder to justify, or somewhat worse than the identical actions without those properties, but that they do not make them impermissible.

First, whether intending harm is a wrong *per se*, i.e whether the reason against intending harm is a *pro tanto* reason against acting, is to be settled by the best theory about why intending harm is a wrong. However, even if the reason against intending harm is a *pro tanto* reason, it is interesting only insofar as it at least sometimes decisively counts against an action. If it makes it a bit more wrong but never impermissible, the DDE loses its action-guiding and explanatory potentials. At best, its victory is Pyrrhic.

Second, the non-absolutist DDE, as usually understood, allows for variations in strength of the reason not to intend harm so that it sometimes renders actions impermissible, although not necessarily in the same cases the absolutist DDE tells us it does. I examined a variety of cases where that might happen, finding them all wanting. I therefore think that the non-absolutist DDE, understood in the usual way, does not fare much better than the absolutist one.

The third objection is that my arguments miss the target since the DDE does not prohibit acting, *simpliciter*, but only acting with intention to harm. This is sometimes put as a charge of confusion of different act-types, thick and thin, the difference being in the thickness of their description: the former comprise intentions and the latter do not, e.g. ‘redirecting the trolley with the intention to kill the one’ and ‘redirecting the trolley’ (Tadros 2011b: 156-7, Wedgwood 2011a: 389-91, 2011b: 468-9). The DDE prohibits thick act-types of the sort; a thin-act type is impermissible if and only if every thick act-type of which that thin act-type is an instance is impermissible (Wedgwood 2001a: 391). A similar way to make the same point is to say that the

DDE need not prohibit what the agent who intends harm does but only his doing of it (cf. Parfit 2011a: 290-1).³⁷

This implies that, in *Bad Bystander*, I act impermissibly, in the sense of redirecting the trolley with intention to kill the one, and permissibly, in the sense of redirecting the trolley, as doing it is permissible, even for me, as long as I could have acted but not with that intention. The same is true of other such cases. However, on this view, the DDE still entails that, if I could not have redirected with a different intention, I ought to have let the trolley kill the five. Suppose that my chances of not intending to kill the one were low. I was indoctrinated to think that he is my enemy and I did not have an opportunity to reflect on that belief. Some might argue that I could not have intended otherwise. This version of the DDE would then implausibly imply that I am not permitted to redirect. The reason against redirecting with the intention to kill the one does not outweigh the reason to save the five.

It might still be true that I should have redirected the trolley with a different intention if I could have done so. But suppose that I deliberate whether to redirect it intending to kill the one. I care about acting morally but I also hate the one. I know that I would need time and effort to get rid of the hatred, none of which I have at the moment. I decide to let the trolley kill the five. But although it is permissible not to redirect, I cannot justify my choice saying that I have a fairly strong reason not to act with intention to kill. The risk of dropping that intention at the expense

³⁷ This corresponds to Hanser's (2005: 444-6, 449-51) distinction between adjectival and adverbial permissibility judgments. That distinction overlaps roughly with the distinction between permissibility judgments of thin and thick act-types. I say roughly because adverbial judgements concern concrete instantiations of act-types as performed by particular agents (e.g. 'X acted impermissibly in redirecting the trolley') while judgments about thick act-types are, in a sense, wider since they lack reference to a concrete agent and, in a sense, narrower since, as defined here, they involve only descriptions of act-types with particular intentions. For the purposes of this objection, however, these subtleties are not important. They boil down to the question whether descriptions of actions such as 'redirecting the trolley with the intention to kill the one' are impermissible. If yes, anyone acts impermissibly performing actions under those descriptions.

of not saving the five is justified only if the likelihood of succeeding is not low and the reason to do it is weighty enough. Since changing intentions for most people in most cases is not an on-off process but a time-consuming and effortful endeavor, the likelihood is indeed low. And even if it were greater, even a coin-toss likelihood, the reason to stop and try to change the intention would not, I believe, be as strong as to justify doing it in the face of the reason to save five people.³⁸

2. The Closeness Objection

I now turn to the Closeness Objection. The objection starts with the supposition that intentions can be fine-grained. I intend to push the large man in order that his body causes the trolley come to a halt but I do not intend to kill him, even if I foresee with certainty that he will not survive the impact. I might intend to make non-combatants appear dead by bombing them in order to force the enemy to surrender without intending to kill them, even if I foresee their deaths as inevitable. If, against all odds, they were to get away unscathed and my ends were to be achieved, I will be delighted. On any plausible version of the DDE, these actions must be impermissible.

As Foot (2002a: 21) has famously remarked, making someone's body stop the trolley and killing him, or making someone appear dead by bombing them and killing them, even if different events, are "too close" to be treated morally differently. The defenders of the DDE therefore need to spell out a theoretically and intuitively satisfying explanation of that closeness which does not contradict the DDE's verdicts about which actions are permissible and which are not (at least in

³⁸ Suppose finally that I can change intentions like TV channels, or like Joshua, I can freeze time and start it again once I change my intention. Would the requirement to do it stand in those unrealistic scenarios? The answer depends in part on whether the reason not to intend harm is a *pro tanto* reason against acting. I say in part since, even if it is, I might be permitted to proceed with that intention if the other reasons at play are stronger. If I risked death or grave injury in redirecting the trolley, my intention to kill the one would not, I believe, suffice to make it true that I ought to have done it with a different intention (cf. Parfit 2011a: 291).

the paradigm cases it purports to explain). In general, they need to tell a story why foreseeably harming without intending harm but an acceptable description of the harm or whatever leads to it is impermissible when identical harming with intention to harm is impermissible but relevantly similar merely foreseen harming is not. The Closeness Objection is nothing but a doubt that such a story could be told.

In this section, I examine two groups of proposals and find them both wanting. The idea of the first group is that intending an alternative description of harm or whatever leads to harm, due to some relation between it and the harm, is in fact intending harm. Call these the Relational Views. The other views state that, although intending a different description of harm or whatever leads to it is not the same as intending harm, it is morally on a par with it. It is impermissible to act intending to involve or use a person in a way that foreseeably harms him if it is impermissible to act in that way intending to harm that person. Call these the Parity Views (cf. Liao 2014: 4).

2.1 The Relational Views

Since much has been written about problems with the Relational Views (see e.g. Bennett 1995: 205-13, Nelkin and Rickless 2015), I shall briefly repeat the problems with some older proposals and focus on the most recent ones.

Five simple views can be ruled out from the start. The relation, one might think, is either that of identity, of events or acts, or of entailment, causal or logical, or a social convention.

There are two problems with the identity views. First, stating that pushing the large man and killing him are one and the same event, they rely on coarse-grained accounts of event and act

individuation, according to which, as long as an event or an act is described in different ways, even if the descriptions are of causally or temporally distant things, they are one and the same event or act which just has all these descriptions. So if I press a switch which activates a robot, which then activates another robot, and so on down a hundred of them, ending with the last robot pushing the large man, the event and the act of pressing the switch and pushing the man are one and the same. Many would go for more fine-grained accounts, according to which events or acts are different if they exhibit different properties, most clearly, temporal properties (see Goldman 1970: Ch. 11).

Second, even if event or act individuation is coarse-grained, intention individuation need not be. For the distinction between intention and mere foresight to be possible, we cannot intend whatever is expectably caused by what we intend even if these are numerically the same. This is why simple causal entailment will not work either. It is also hard to believe that we must intend all descriptions of what we intend. I intend to save the five in *Bystander* but not to save a prime number of people, even though I know that five is a prime number. This also shows why logical entailment cannot be that of logical necessity.

Another reason why logical necessity is not the right relation is that it would make the DDE permit too much. If I must intend a harm caused by what I intend only if the harm is caused in all possible worlds where I intend so, then I need not intend to kill the man in *Bridge* because there are worlds where miracles happen, God intervenes, etc.

Suppose that we understand the entailment in terms of comparative similarity of possible worlds (we might understand causal entailment in this way if we accept a counterfactual view of causation) (cf. Tadros 2015a: 57). On this proposal, I must intend a harm caused by what I intend

if there is no nearby possible world where I intend so and the harm is not caused. Distant worlds where miracles occur do not matter. I must therefore intend to kill in *Bridge* and *Terror Bombing* because the worlds where I push the large man or bomb the non-combatants and they survive are too distant from the actual world.

However, the same seems to be true in *Bystander* and *Strategic Bombing*. A world where I redirect the trolley towards the one but I do not kill him, assuming he is trapped on the tracks, is far away from the actual world. And in all nearby worlds where I bomb a munitions factory, I kill some non-combatants, because munitions factories are usually populated by workers. Moreover, since in all nearby worlds descriptions of what I intend remain fixed, the view implies that I must intend them too. But, as we have seen, that is rather implausible.³⁹

Finally, we might go for the social conventionalist account, according to which I intend a harm caused by what I intend if its nonappearance, given my act, is what “the plain man would call *inconceivable*” (Bennett 1995: 213). Although it is possible that the large man gets unscathed after being hit by the trolley or that the non-combatants survive the raid, no sane person would consider those possibilities seriously.

However, I do not see why the plain man would think it conceivable that the one trapped on the side track survives being hit by the trolley in *Bystander*. The conventionalism also brings in a dose of relativism. It is unclear how to think of “the plain man”: as someone I have just meet on a street, an average person of one or another society, etc. Similarly, the epistemic capacities of “the plain man”, even if he is epistemically virtuous, are heavily context-dependent. What I might find inconceivable, you might not. Before the discovery of penicillin, surviving gangrene

³⁹ The same objections apply to probabilistic version of that view.

in one part was virtually inconceivable. It might be a matter of time when we find it conceivable that a person hit by a vehicle running in full speed gets away unscathed.

Let us now turn to more refined proposals. On one view (FitzPatrick 2006), the relevant relation is that of natural constitution, where natural is opposed to conventional and one state of affairs constitutes another naturally when it neither causally necessitates it nor logically entails it but only makes it happen as a matter of physical laws operating under normal conditions (i.e. without God's intervention, etc.).⁴⁰ Examples include the large man being thrown off the bridge and him being hit and killed by the trolley, the non-combatants being blown up and them being killed, etc. Receiving an "F" in an exam is constitutive of losing the right to financial aid but the constitution is conventional; the rules of financial assistance could be set in other ways.

The view then states that the agent must intend harm if one intends that which is naturally constitutive of it and the constitution is known to the agent (FitzPatrick 2006: 593). I must intend to kill the large man if I push him onto a track where a runaway trolley in high speed will hit him but I need not intend a student to become ineligible for financial aid if I intend to give him an "F", knowing what will happen as a result of it.

An obvious worry here is that, if the natural constitution is the operation of physical laws under normal conditions, it is but a mere causal relation. Redirecting the trolley is constitutive of killing the one on the side track and bombing munition factories is constitutive of killing non-combatants. It might be replied that these are not constitutive relations since it is possible, under normal conditions, for one state of affairs to obtain without the other. Side tracks usually do not have people trapped on them and some munitions factories have no personnel or people living in

⁴⁰ Truth be told, FitzPatrick does not offer necessary and sufficient conditions for natural constitution but only gives a set of examples instantiating it. However, I believe that the above is the most natural reading of his view.

their vicinity. However, absent miracles, being hit by runaway vehicles or blown into bits in raids does result in deaths of those affected.

The problem is that the normality at issue cannot be taken in isolation from the essential features of a case. An essential feature of *Bystander* is that the one is trapped on a side track. If we are allowed to abstract from that feature, why are we not allowed to suppose that pushing the large man in *Bridge* does not kill him but rather causes him to harmlessly fall on a switch which stops the trolley (nothing miraculous about it)? Likewise, if we suppose that some munition factories are deserted, why not suppose that some bombs explode above non-combatants, making the enemy surrender falsely believing that they died? There is no reason for differential treatment apart from modal or probabilistic grounds, which were rejected before.

Furthermore, even if pushing the large man were constitutive of killing him (and so on for other cases), it does not follow, without an additional argument, that if I intend the former I must intend the latter. As we saw above, there are intuitive grounds to think that we need not intend different descriptions of one and the same event. This is clear when descriptions are of permissible actions, such as saving five people and saving a prime number of people. I see no reason why the situation would be different in cases of impermissible actions.

Another (Wedgwood 2011a, 2014) view tries to revitalize the “plain man” view, stating that I must intend harm or, more generally, a bad state of affairs if and only if the state of affairs I intend to bring about are considered (extrinsically) bad on the basis of expectations about them that a virtuous person knowing a range of true *ceteris paribus* moral generalizations (e.g. that, *ceteris paribus*, and under normal conditions, it is harmful to be hit by a trolley running in full speed) but being ignorant of particular facts about the case would form, and those expectations

are borne out. The plain man is here epistemically and morally virtuous; it is an ideal type that spells out the badness of a certain state of affairs.

So if I intend to make the large man stop the trolley and foresee that he will thereby die, I intend to kill him because being hit by a trolley is extrinsically bad. An appropriate reaction to hearing that someone was hit by a trolley is “Oh no, horrible news!” Although such a reaction is appropriate if a virtuous person justifiably believes that the man will be hit by a trolley but he miraculously survives, on this view, I would not intend to kill the man since the expectation is not borne out. I find it very strange that what one intends depends on what one actually succeeds in doing. A murderer who cares only about killing the large man in *Bridge* intends to kill him even if he survives. Otherwise, no incomplete attempted harm is intended, which cannot be true.

But there is a bigger problem. The view has different implications about the badness of a state of affairs depending on its description. Suppose first I tell a virtuous person that someone redirected a runaway trolley. His reactions might be: “That’s an interesting info”, “So what, very boring”, etc. Suppose then that I tell him that someone redirected a trolley towards one trapped person. His reaction would be the same as if he heard that a man was thrown off a bridge into the path of a runaway trolley: “Horrible!”. Suppose, finally, that I tell him that someone redirected the trolley to save the five. He might say: “It’s horrible that one person died, but it is a good news that the five were saved”.

Which of these descriptions is relevant to deciding whether the bystander intended to kill the one? Since, on this view, we are to assume that a virtuous person is unaware of peculiarities of a case, it cannot be the third. I see no reason why the first or the second would have a priority. Depending on how a virtuous person’s ignorance of particular facts about a case is understood,

the view has contrary implications. Since it is completely arbitrary to choose one over the other, we cannot decide whether the intended state of affairs was bad or not. The same is true of *Bridge* or any other case.

Suppose that one suggests that the proviso of ignorance of particular facts is dropped and we are to assume that a virtuous person has knowledge of all the relevant facts. This would not solve the problem as the virtuous person's reactions would then merely reflect ordinary intuitions about the permissibility of actions in the relevant cases. *Bystander* would be characterized as a bad but overall acceptable state of affairs and *Bridge* as just bad. We could not non-circularly appeal to that criterion to decide whether what is intended is bad and an action is impermissible in virtue of that intention. Either way, the view is flawed.⁴¹

2.2 The Parity Views

I now turn to views that intending harm and intending what leads to it or a different description of it, although numerically distinct, are morally on a par. Two such views are relevant. On the Involvement View (Quinn 1989: 343, Tadros 2015a: 57, Nelkin and Rickless 2015: 404), if one intends to involve or to affect a person without his consent in a way that foreseeably leads to harming him, one acts as wrongly as if he were to intend to harm that person, other things being

⁴¹ Delaney (2008: 349) pessimistically writes that “if we do not offer some admittedly rough and ready rationale for blocking extremely fine-grained ascriptions of direct intention (...) DDE will simply issue absurd pronouncements about the moral permissibility of patently unacceptable courses of conduct”. This need not follow if the Parity Views are true.

equal.⁴² On the Using View (Tadros 2015a: 57), to act with intention to use a person without his consent in a way that foreseeably leads to harming him, is to act as wrongly as with intention to harm that person. The former view is broader; we might affect or involve someone without using him but not vice versa. If I intend to remove someone who presents an obstacle to achieving my end, I do not intend to use him. His presence in no way helps me to achieve my end; indeed, it prevents me from achieving it.

Although these views give better answers to the Closeness Objection than the Relational Views, I believe they are also mistaken. Let us start with the Using View. As I have just said, in some cases, we remove others but do not use them. Some removals are impermissible since they kill or gravely injure people in ways that cannot be justified. In *Driver*, rushing to the hospital to save the five, I run over a person trapped on the only road that gets me to the hospital (Kamm 2007: 22). In another famous case, *Cave*, a large man is stuck in the mouth of the cave, trapping five persons behind him who will be killed by a rising flood. I save the five by blowing up the large man with a stick of dynamite (Foot 2002a: 21). Since the Using View prohibits only acting with intention to use others in foreseeably harmful ways, it does not prohibit those killings.

Another objection is more worrisome. In *Bystander*, I might intend to use the switch, i.e., to press it, foreseeing that this will cause the trolley to get redirected on the side track and kill the one. In *Bridge*, I might intend to use a switch, i.e. to press it, foreseeing that that will cause the large man to fall into the path of the trolley, stop the trolley, and die. In both cases, I intend to use that which foreseeably harms a person. The Using View implies that such intentions are morally

⁴² That acting is often called harmful direct agency, a term coined by Quinn (1989: 433) to contrast acting with intent to involve a person in a foreseeably harmful manner without his consent to acting in the same way either without that intent or with it but without the harm resulting from it (harmful indirect agency). It is then claimed that there is a bigger moral burden on justifying the former type of agency (Quinn 1989: 344, Nelkin and Rickless 2014: 129-30). I argue against that claim below.

on a par with intentions to harm. We then have to distinguish them in order to avoid prohibiting both actions (cf. Mapel 2001: 263-5).

The causal chain cannot stop at different points in the two cases. That would be ad hoc. It must be that there is something in the latter using which makes it especially hard to justify. One explanation (Tadros 2015a: 65-7) is that, by using the large man to save the five, I either compel him against his will to act in the service of bringing about what is impartially the best, namely saving the five, or use his body for that purpose against his will, and no one has a duty to serve others's ends against his will, even if those ends are impartially the best.

Why do I not compel the one on the side track in *Bystander* to serve what is impartially the best? Since I do not rely on his agency or body to achieve my end in *Bystander* (but I do so with the large man in *Bridge*), the reply (Tadros 2015a: 67) goes, I do not compel him to serve my end. But the fact that one's agency or body is not *needed* for the achievement of an end makes it true that he lacks the right not to be compelled to serve that end only if relying on one's agency or body as a causally contributory means to an end is a wrong in the first place. And it cannot be a wrong because, qua autonomous beings, we are entitled to determine for ourselves which ends to pursue. If we have that entitlement, we can legitimately refuse to be sacrificed as a side effect too. If the one in *Bystander* does not have a right against being killed as a side effect of saving the five but the large man in *Bridge* has the right not to be killed as a means of saving the five, that cannot be because both of them can legitimately refuse to be compelled against their will.

A similar problem haunts the Involvement View. As we saw, all using is involving and affecting, so the critique just given translates to those cases. However, some involving, like that

in *Cave* or *Driver*, is not using. Can it be that the reason why it is impermissible to blow up the large man in *Cave* is that I intend to involve him in a way that foreseeably harms him but the reason why it is permissible to redirect the trolley in *Bystander* is that I do not intend the same? To answer that question suppose that, in a variant of *Cave*, I intend to affect a switch, by pressing it, which activates the bomb that foreseeably kills the large man. If that intention is morally on a par with the intention to kill the man, why is not the intention to affect the switch with a foreseen side effect of redirecting the trolley and killing the one in *Bystander* on a par with the intention to kill him?

One reply (Tadros 2015a: 73-4) could be that, because I do not *need* the one in *Bystander* to achieve my goal of saving the five but I need the large man to achieve the same goal in *Cave*, I do not impose an end of saving the five on the former. To repeat, it is question-begging to claim this. And if one has a right against being involved in the way that foreseeably harms him in the course of achieving an end because one can refuse to adopt that end, one has that right whether one's presence is a causally necessary means to the end or its foreseen and regrettable side effect.

The Involvement View, therefore, can neither serve as a basis for the DDE.

3. The Explanatory Objection

Finally, I turn to the Explanatory Objection, according to which the DDE's explanatory power is null. The cases the DDE is often said to explain are neither explained in virtue of the distinction between intending and merely foreseeing harm nor in virtue of the distinction between harming as a means and harming as a side effect.

This is an important objection since some defenders of the DDE believe that, even if the DDE is not the most perfect principle to tell us when an action is permissible, it is the best thing we have when it comes to the explanation of hard cases. For instance, Nagel (1984: 179) writes:

I believe that the traditional principle of double effect, despite problems of application, provides a rough guide to the extension and character of deontological constraints, and that even after the volumes that have been written on the subject in recent years, this remains the right point of convergence for efforts to capture our intuitions.

Similarly, McMahan (2009a: 370, cf. Nelkin and Rickless 2014: 131) writes:

[M]any of our firmly held moral beliefs – such as that there is a significant moral difference between just war and terrorism, that it is permissible to divert the trolley in the original case (*Bystander*) but not to push the man into its path in the means-variant (*Bridge*), and so on – are explained in a simple and seemingly plausible way by the principle that one's intention can affect the permissibility of one's action.

Examining them in turn on some relevant cases, I show the explanatory futility of the distinction between intending and foreseeing harm and the distinction between harming as a means and as a side effect. I then sketch alternative explanations of those cases.

3.1 The Method of Isolation

What is the best way to examine the explanatory relevance of some consideration? I believe it is what we might call *the Method of Isolation*. It tells us to take a description of an action which is

said to be impermissible because of some consideration and contrast it with identical description of that action but without that consideration. If the judgment about its impermissibility remains, we have a reason to assume that the consideration plays no role in the explanation of the former description; that is, that the action is not impermissible in virtue of it.

Something like the Method of Isolation is familiar and widely used. However, as current disagreement about the DDE shows, it fails to produce results. There are two reasons for it. First, there is often an intuitive disagreement about the permissibility of the descriptions in which the questionable considerations are isolated. Our task should then be to consider only descriptions of actions everyone agrees are impermissible. Second, the contrasted descriptions are rarely fully identical. They include other morally relevant considerations which confound our assessment of those cases.

Both mistakes are present in the debate about the DDE. Take first the distinction between intending and foreseeing harm. It is often said that it is impermissible to act in *Bridge*, *Terror Bombing*, or *Transplant* because I would intend to kill the victims as a means. We saw that I need not intend harm in those cases but also that the defenders of the DDE deny this or claim that my intentions would be nonetheless morally problematic.

However, there is a simpler and relatively uncontroversial way to prove the explanatory irrelevance of intentions in those cases. Since considering identical actions without intention to harm causes disagreement, we should consider identical actions without intentions at all. We can do so with the help of intentional but unintended actions. It is widely thought (e.g. Anscombe 2000, Harman 1976, Bratman 1987, Mele 1992) that such actions are possible and indeed quite common. Examples include habitual, spontaneous, and actions done for no reason. When I shift

from one to another gear or raise my voice unexpectedly, I do so intentionally but I do not intend anything. We can then also intentionally cause or allow harm without intending it.

It is easy to construct versions of *Bridge*, *Terror Bombing*, or *Transplant* identical in all respects with the exception that I do not intend anything. For example, suppose that, like in the movie *Groundhog Day* I am trapped in a time loop, and am the only person aware of it. In the first scenario, I find myself repeatedly in *Bridge*, in the second, in *Terror Bombing*, and in third, in *Transplant*. I am convinced that I ought to save the greater number in all cases, so I push the man off the bridge to save the five, bomb the non-combatants to end the war, and cut up the healthy patient to save five others. Even after the *n*th time, I do the same. My actions become habitual, just like shifting from second to third gear is to an experienced driver. It is plausible to suggest that I cause these harms intentionally but that I do not intend anything. Yet, it is obvious that I nonetheless do it impermissibly. This gives us a reason to doubt that intending harm as a means (or intending to involve or use one in such a way) is what explains the impermissibility of acting in the original cases.

Consider now the harming as a means/as a side effect distinction. We need cases identical to *Bridge*, *Terror Bombing*, and *Transplant* except for the fact that the harms are brought not as a means but as a side effect. Some might say that we already have such cases: *Bystander*, *Strategic Bombing*, and *Drug Shortage*. But these are not, I believe, even sufficiently similar to the former. There are important moral differences between them, which I shall turn to below. It is best to try to imagine descriptively identical cases minus harming as a means.

Here are some such cases. In *Bridge II*, I cannot push the large man off the bridge to save the five but can push him onto the switch, which will then become activated due to his weight

and cause the trolley to stop just before it hits the five. Unfortunately, if I push the man onto the switch, he will fall off the bridge onto the track and the trolley will hit and kill him. Since it is the switch which causes the trolley to come to halt and not the man's body, his falling off from the bridge is not a causally necessary (nor sufficient) means of saving the five but its side effect. Therefore, if I push him, I kill him not as a means but as a side effect (Parfit 2015b: 16). Yet, it seems impermissible to do it.

In *Terror Bombing II*, I drop the bomb on non-combatants foreseeing that the enemy will hear the blast and that that will cause it to surrender. The blast will not kill the non-combatants directly but, unfortunately, it will cause a chasm to open in which they will fall and get killed. I therefore do not kill them as a means but as a side effect of ending the war (Tadros 2015a: 63, cf. Quinn 1989: 343, n. 16, FitzPatrick 2006: 612). Yet again, it seems that I act impermissibly.

In *Transplant II*, I cannot save the five patients in need for organs by taking the organs of the one healthy patient because they do not match perfectly. However, the five could be saved by artificial organ matches made on the models of the one's organs. Unfortunately, the only way to make the artificial organs is to cut up the one, take out the needed organs, and scan them on the machine that will make the models for artificial matches. Since the scanning takes time, although I would then return the organs to the one, he would foreseeably die in the meantime. I would not kill him as a means of saving the five but as a side effect, though I would kill him impermissibly.

If these cases are identical in all morally relevant respects to *Bridge*, *Terror Bombing*, and *Transplant* except for the fact that killing in them is a side effect and not a means and if they are

impermissibility cases, then we have reason to think that killing as a means is not what explains the impermissibility of the original cases.⁴³

What does explain it then? A legitimate answer is: “I do not know”. Lack of alternative explanation does not count in favor of the DDE nor does it undermine the objections against it I gave. Nonetheless, in what follows, I shall sketch what I see as the right path to think about the hard cases.

3.2 Explaining the Hard Cases

Consider two notable explanations first. On one view (Quinn 1989: 191, Tadros 2015a: 64-5), it is impermissible to cause or allow a good by exploiting an opportunity that exists only in virtue of some person in the way that kills or gravely injures that person. The view seems nicely to explain the impermissibility of acting in *Bridge II*, *Terror Bombing II*, and *Transplant II*. In these cases, I exploit the situations that would not exist were the victims not there: I could not save the five if the large man were not in front of the switch, I could not force the enemy to surrender if they did not believe that the non-combatants are dead (presumably, they could not believe so if the non-combatants were not in fact there), etc.

Notice that this view explains why it is impermissible to act in the original pairs of these cases too. However, since there are cases where a person is impermissibly harmed although he does not contribute to an opportunity for the agent to achieve a good, the view prohibits too little.

⁴³ Regarding a case akin to *Terror Bombing II*, Nelkin and Rickless (2014: 140) say that it features an impermissible and as bad action as the one in *Terror Bombing* but that the former is morally different because it does not harmfully use non-combatants as a means. But, as said above, if the victims have a right not to be used in such a way because they can rationally refuse to serve one’s ends, they have an equally strong right not to be sacrificed as a side effect of achieving those ends.

As we have seen, in some such cases, like *Driver* or *Cave*, the victim creates an obstacle rather than an opportunity to achieve a good. Similarly to the Using View, the Exploitation View gives a wrong verdict in such cases.

The Exploitation View also prohibits too much. Some instances of harmfully exploiting an opportunity are not impermissible. In *Bad Bystander*, for instance, I abuse the opportunity to save the five intending to kill the one but I am nonetheless permitted to redirect the trolley.

On another view (Kamm 2007: Ch. 5), roughly, causing or allowing harm or involving a person in a way that foreseeably causes or allows harm to him in pursuit of a good is permissible if the harm results from the good itself or that which constitutes it but impermissible if it is the causal, in the sense of causally necessary, means to the good or if it results from the causal means to it, even as its side effect. In *Bystander*, the redirection of the trolley is not a causal means of saving the five but the same event under a different description; redirecting the trolley is identical to, or at least constitutive of, the five being saved, which is why it is permissible to do it (Kamm 2007: 141). However, in *Bridge*, the death of the large man—that is, his collision with the trolley—is the causal means of saving the five, which is why it is impermissible to push him.

Leaving aside the questions of event individuation and constitution underlying this view as well as the question of justifying the moral relevance of its causal distinctions, I shall focus on its implications. In *Bridge II*, pushing the large man on the switch is the causal means of pressing it. His falling onto the track and being killed by the trolley is a side effect of saving the five. But pressing the switch is identical or at least constitutes the events of the trolley coming to halt and the five being saved, no more than pressing the switch is identical or constitutes the events of the trolley being redirected and the five being saved in *Bystander*. Therefore, since in *Bridge II* the

harm of the man being killed is a side effect of the good itself or of what constitutes it rather than of the causal means to it, the view in question implausibly implies that it is permissible to act in *Bridge II*. The mistake reappears in *Terror Bombing II* and *Transplant II* as both are cases where the harms are not the result of the causal means to the goods.

Moreover, since this view deems impermissible any causing or allowing harm by means to a good when the means do not constitute the good, it must distinguish morally between killing as a side effect in a version of *Bystander*—where in the absence of a redirecting switch, I throw a bomb on the trolley, which causes it to get redirected on the side track and then explode, killing the one before it would otherwise hit him—from a version where everything is the same but the one is killed by a shrapnel rather than a part of the trolley. This is because, only in the former version, the harm of the one being killed is caused by what constitutes the good. But intuitively, there is no moral difference between the two; both actions are either permissible or they are not (cf. Hurka 2015: 138-9, 141-2, *pace* Kamm 2015: 222-4).

Here is a story that is simpler and, I think, closer to the truth. Recall that one reason why our disagreements pertain is that we compare cases that differ in more than one morally relevant respect, hoping to find principles that explain the moral differences between them. Although this hope shaped much of nonconsequentialist ethical theory for the last couple of decades, we made very little progress. It is then not unreasonable to ask whether our goals are set correctly. It seems to me that they are not.

Consider first the Trolley Problem. When Foot (2002a: 23-4) introduced it, she argued that, despite some of its theoretical difficulties, the DDE explains the moral difference between *Trolley Driver*—in which, being the driver of the runaway trolley, I redirect it towards the one in

order to save the five—and *Judge*—in which, acting as a judge unable to find a real culprit for some horrendous crime, I order killing of an innocent person in order to save five others from being killed by an aggressive mob demanding the culprit to be found.⁴⁴ She also compared *Drug Shortage*, in which I withhold a drug from a patient in need of the whole supply of it to give it to five others in need of one fifth of it, with *Transplant*. The first action in each pair is permissible and the second impermissible despite that, in all of them, one person is killed (or allowed to die, in *Drug Shortage*) to save five others. Similarly, when Thomson (1976) introduced the problem under its current name, she sought to find the difference between *Trolley Driver* and *Transplant*. In her later, better known, article (1985), she contrasted *Bystander* with *Transplant*, and only then *Bystander* with *Bridge*.

Most people nowadays think of the Trolley Problem in terms of the last pair. But as some (Kamm 2015: 12-3, 172-3) have suggested, we should understand it more broadly, as a search for the principle that accounts for all structurally similar cases. Suppose this is the case. What gives us reason to think that any of the traditionally contrasted pairs are such that there is a principle that explains the moral differences between them? The answer to that question will suggest why all structurally similar cases will be explained in such a way.

One answer might be that we search for the supreme principle of morality, the principle that would tell us when and why any action is impermissible. However, this is not the goal of those who try to solve the Trolley Problem. For example, Kamm's (2007) search for the essence of nonconsequentialist constraints on harming is limited to constraints on harming innocent, non-threatening persons, for whom further life is a good, and who do not consent to be harmed.

⁴⁴ Foot (2002a: 25, 27-9) later writes that the DDA does a better job in explaining these cases.

Moreover, even if we find the supreme principle of morality, it would likely be general and, thus, unable to explain the contrastive pairs at the level of detail we search for. For example, suppose that Parfit's (2011a: 412-3) Triple Theory is the supreme principle. According to it, an action is wrong just when it is disallowed by optimific, universally willable, and not reasonably rejectable principle. Suppose also that, contrary to what Parfit thinks, this theory tells us why any action is impermissible. Even if pushing the large man in *Bridge* or killing the healthy patient in *Transplant* are impermissible because they would be disallowed by a principle that satisfies the three properties but redirecting the trolley in *Bystander* or withholding the drug in *Drug Shortage* would not, this would show why these are permissible or impermissible on the deepest level of explanation. But in comparing those cases, we hope to find much more specific explanations. We search for mid-range principles, which are themselves grounded in the supreme principle, if there is any.

Why then think that there are mid-range principles that account for our contrastive pairs? Why not suppose that at least some permissibility cases are permitted as exceptions to principles against harming innocent persons whereas some impermissibility cases are forbidden by *different* principles? Recall Foot's comparison of *Trolley Driver* and *Judge*. The two might differ morally not only in how the victims are harmed but in who harms them and who they are. Trolley drivers might have different role-based obligations than judges and, as citizens, we might have stringent claims against being framed but less stringent claims against being killed by redirected runaway vehicles. Or it might be that role-based obligations and rights play some role in *Judge* but not in *Trolley Driver*. Either way, we should not expect the difference between the two to be explained by a single mid-range principle.

The same seems to me to be true of other controversial pairs. *Bystander* is very different from *Transplant*, primarily in who harms whom rather than in how the harm is brought about. As patients, we have stringent rights against being cut up by our doctors and, as doctors, we have stringent role-based duties not to kill or gravely injure patients or use them for medical research without their consent for the sake of saving a greater number of other patients. This is true whether we intend or merely foresee doing it, as a means or as a side effect (cf. *Transplant II*). Nothing of the sort is true in *Bystander*.

Drug Shortage might be apparently similar to *Transplant* but operate under a completely different principle of distributing scarce resources. We might not be allowed to distribute organs in the same way as drugs, and, when drugs are not a patient's property, we might be allowed to allocate them so as to save the greater number of patients (Scanlon 2008: 33-6).

In some other hard cases, like *Bridge*, it might not matter how someone is harmed nor who harms whom but in which context that person is harmed. We might have a stringent claim not to be caused to fall from bridges or other tall structures, whether as an end, or as means, or as a side effect of saving several other people (cf. *Bridge II*) (Scanlon 2008: 120-1, Parfit 2015b: 17).

Finally, *Terror Bombing* might be impermissible because killing non-combatants in war is not a legitimate military advantage, even if they are killed as a side effect (cf. *Terror Bombing II*) (Scanlon 2008: 28-9). Legitimate military targets, we might say, are munition factories or other military facilities but not non-combatants. Since it is inevitable that some non-combatants will be killed as collaterals in the course of aiming at legitimate targets, the general prohibition should

have an exception for collateral killing as long as some kind of proportionality is satisfied. But this has nothing to do with the harm being brought merely foreseeably or as a side effect.

One might ask why is it that we have more stringent claims against being cut up by our doctors, thrown off of bridges, and bombed in war as non-combatants? I doubt that the appeal to any mid-range principle will give us a unifying answer to such diverse cases. For that we need principles of greater generality, like the Triple Theory.

My point was not to give a complete explanation of these cases but to point to a different direction in explaining them. If I am right that they differ in so many morally relevant respects, the project of explaining them by a single mid-range principle (such as the DDE) was misguided from the start.⁴⁵

Conclusion

In this part of the two-part piece on the DDE, I defended three objections to the interpretation of the DDE I gave in the previous part. The objections are that the DDE's prohibitions on intending harm and on harming as a means are indefensibly strong, that the former cannot be spelled out in a way that prohibits intending that which is close to harm but not harm but that does not prohibit merely foreseeing it, and that they have no role in explaining the relevant cases. Taken together,

⁴⁵ Kagan (2015: 153-6) has recently pursued a similar deflationist strategy. He argues that, even if we find a mid-range principle that captures everyone's intuitions about the relevant cases, the likelihood that it will rest on features that are independently morally relevant significantly decreases with its complexity. Kagan (2015: 164) adds that "our intuitions about trolley problems respond to factors that simply do not have any genuine moral significance". My claim is weaker: complexity of proposed principles is but a manifestation of a deeper problem having to do with the cases in question being too far from each other in morally relevant features to be successfully explained by any principle of that range. So even if our intuitions respond to factors of genuine moral significance, which I think they sometimes do, a unifying rationale requires that the factors are about the same set of phenomena, which, as I argued, we have no reason to believe is the case.

these objections imply that the DDE neither spells out when nor why an action is permissible or impermissible.

Appendix: Triple Effect Defects

Here I briefly consider the weakness of the so-called Doctrine of Triple Effect (DTE), a principle which, like the DDE, takes intending harm to non-derivatively render actions impermissible, but which restricts the scope of cases where that happens more than the DDE. We owe the DTE to Kamm (2007). She no longer accepts it, for some of the reasons I shall mention below, although I shall focus on what I see as the crucial problems with it.

One of Kamm's (2007: 133) definitions of the DTE reads: "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 the expectation of which we may have as a condition of action."

Several clarifications are in order. First, the DTE is an absolutist principle. That a greater good may justify a lesser evil means that it might make causing it permissible rather than merely less wrong (Kamm 2007: 92-3). Second, like the DDE, the DTE prohibits causing and allowing harm, given the other conditions (Kamm 2007: 119-20). Third, evil* is a shorthand for harming others and/or involving them in ways that foreseeably harm them (Kamm 2007: 93). I shall refer to it as harm. Fourth, the DTE's application is restricted to harming innocent, non-culpable, and non-threatening persons, who do not consent to be harmed, and for whom further life is a good (Kamm 2007: 130-1). Sixth, the DTE takes the distinction between acting with intention to harm

and acting because of harm, or on condition of it, as morally relevant in the sense that the former is impermissible and the latter permissible (given the other conditions).

I shall focus on the objections raised against the DDE and see how the DTE fares against them. Consider first the Strength-Of-Reasons Objection. The DTE prohibits intending harm, as a means or as an end, regardless of other considerations, as long as the victims are innocent, non-threatening, etc. It would therefore prohibit redirecting in *Bad Bystander*, which, as I argued, is a mistake (cf. Kamm 2007: 135). Interestingly, it would not prohibit redirecting in *Loop*. Kamm believes that *Loop* is a permissibility case because I do not intend to kill or even to hit the one or involve him in the way that foreseeably harms him but I redirect the trolley only because of (or on condition of) the five being thereby saved. She gives two arguments for that conclusion, both of which I find wanting.

The first argument is by analogy. Kamm appeals to *Party*, a case where I intend to give a party for me and my friends to have fun but, knowing that a mess will result and that my friends will feel indebted and help me clean up, I give the party only on the condition that they help me (my pleasure of partying does not outweigh the hardship of cleaning) but not intending that they feel or act in those ways. Kamm (2007: 95) claims that *Party* shows that it is possible to take a harm, assuming a feeling of indebtedness is a harm, as a reason for acting, i.e. as something one acts only because of, but not to intend it. Admitting that *Party* is not fully analogous to *Loop*, since in the former the harm is not needed for the good, she (2007: 96) introduces *Party II*, a case identical to *Party* with the exception that I throw the party only on the condition that my friends feel indebted (the party will not be a success unless it lasts as long for the mess to accumulate)

but not intending that they do. Since *Party II* is sufficiently similar to *Loop*, Kamm believes, we might safely assume that in *Loop* I also do not intend harm.

The argument just sketched might not only show that I do not intend to hit the one in *Loop* but that I act permissibly, if *Party II* is a permissibility case and if it is sufficiently similar to *Loop*. However, the latter is not true. There are several disparities but the following is obvious: although a feeling of indebtedness is a causally necessary side effect of the good of throwing the party (and having fun), it is not a causally necessary means for it, in the ordinary, narrow, sense of ‘means’. It is such a means to a different end, namely cleaning the mess, and a motivational *sine qua non* for me throwing the party. However, hitting the one is a causally necessary means for the good of saving the five in *Loop*, in the ordinary sense of ‘means’.

A better analogy would be *Party III*, where indebtedness is a causally necessary means for giving the party because the party’s theme and aim is teaching reciprocity. But it is no longer clear that just because I would give such a party on the condition that my guests feel indebted, I would not intend that they feel that way. It will neither help to know my attitude towards them. I might think it is instrumentally good that they feel that way but fail to intend that they do, like in *Chemotherapy*. To check whether I intend so or not, we must appeal to the Counterfactual Test or some test of that kind (a full story would require giving sufficient conditions for intending too).⁴⁶ But even if I do not intend harm in *Party III*, those thinking of *Loop* as an impermissibility case, most probably think the same of *Party III*. The analogy argument, therefore, fails.

The second argument builds on Kamm’s (2007: 96-101) account of intending. Drawing on Bratman’s (1987: Ch. 10) insights, Kamm sets three conditions as necessary and sufficient for

⁴⁶ Kamm (2007: 96) rejects the Counterfactual Tests because she wrongly takes passing it as sufficient for intending.

intending: that we must seek means to accomplish an intended end, that we must adopt additional means to achieve the end if the initial ones fail, and that we must, insofar as we are rational, have non-contradictory intentions, revising or abandoning the ones that conflict with the intended end. She then argues that, because none of these conditions is met in *Loop*, I do not intend to hit the one. The argument shows that redirecting in *Loop* is permissible only if intending/acting because of distinction has moral relevance, cutting across the impermissibility/permissibility distinction, other things being equal. Let us grant that for a moment.

Kamm spends much time showing that the three conditions are not met in *Loop*. Even if she is right about it, I find the whole endeavor futile. The conditions are not jointly sufficient for intending. In *Chemotherapy*, the doctor might seek different means (i.e. types of chemotherapy) to damage the organs in order to save the patient and he might try his best to minimize the harm but not intend the damage. Moreover, as argued before, the second condition is not necessary as some intentions are conditional in nature. Kamm, therefore, did not show that it is permissible to redirect in *Loop*.

Consider next the Closeness Objection. Notice that the DTE does not prohibit harming as a means independently of intending harm as a means. To reach the proper verdicts, it must then suppose that, in impermissibility cases like *Bridge* or *Transplant*, I must intend harm (understood as evil*) as a means (cf. Kamm 2007: 122-3, n. 5). But if I must at least intend to involve the victims in those cases in ways that foreseeably harm them, then the DTE faces the objections to the Involvement Views.

Consider finally the Explanatory Objection. Since the DTE explains the impermissibility of acting in *Bridge*, *Transplant*, and similar cases by the fact that the agent intends harm in them,

its explanation is mistaken for the same reasons as the DDE's. Kamm (2007: 136) acknowledges that the DTE fails in *Driver* and *Gas* because it permits any merely foreseen harming as a side effect but she does not realize that, on the same basis, cases analogous to *Bridge* and *Transplant* could be constructed (e.g. *Bridge II*, *Transplant II*).

It is interesting that Kamm (2007: 119-20) argues that there is a permissible version of *Transplant*, in which I can save five people from an illness leaving one person die from the same illness. The treatment of the five is peculiar since I would first infect them fatally and then save them by transplanting the one's organs into them once he dies. Kamm thinks that I may refuse to save the one expecting to profit from his demise because I would save them only because the one will die but I would not intend him to die, as an end or as a means.

I share the intuition that I would be permitted to do this but I doubt that the explanation Kamm offers is correct. In original *Transplant*, I am not permitted to let one *healthy* patient die so that I could use his organs to save the five. As *Transplant II* shows, it is irrelevant whether I kill him as a means or as a side effect. We could even imagine that, in *Transplant II*, cutting him up is the only reason why I act but that I would return his organs after the copying and try to get him back to life with all available means. The action would still be impermissible. Kamm's case is special in the sense that it is closer to *Drug Shortage* than to the original *Transplant*. It is an instance of the problem of distributing scarce drugs to the deadly ill; absent other grounds, the rule seem to be here to save the greater number. If we add to it that there is no good reason not to cut up a corpse and use its healthy parts to save those in need of them, we get a more plausible explanation for the permissibility of acting in Kamm's version of *Transplant*.

The DTE, therefore, inherits most of the DDE's defects.

3. Intentions and Overdetermined Wrongdoing

Introduction

It is a common belief that good intentions, such as the intention to bring good or to prevent or reduce harm, do not render otherwise morally impermissible actions permissible (but that they do render blameworthy ones less blameworthy). Presumably, the belief is the main reason why the debate about the relevance of intentions to permissibility is engrossed by the questions whether and why bad intentions, such as the intention to harm, affect permissibility and why the modes of influence of good intentions are glossed over.⁴⁷

A noteworthy exception is Tadros (2011b, 2014b), who argues that good intentions affect permissibility in some cases of overdetermination. On Tadros's view, an action which, along with some other actions, (otherwise impermissibly) overdetermines a wrong is permissible just when and because it is done with a good intention. To take an example that is not his own, a member of a firing squad which wrongfully executes a prisoner of war (he is innocent and there is no other reason to kill him) acts permissibly just when and because he shoots the prisoner only in order to make him die a less harmful and a swifter death than he would die were he not to be hit by that

⁴⁷ The intuitive asymmetry between the influence of good and bad intentions on permissibility is problematized by Kagan (1989: 177-81) and Quong (2013). The asymmetry is puzzling but I do not think that it is impossible to give some theoretical basis for it, for instance, by defending its derivation from the general asymmetry in stringency of reasons to benefit and those not to harm.

bullet (holding the others's shootings fixed). But if he shoots the prisoner intending to kill him, out of hatred, revenge, or a sense of loyalty to his co-combatants, he acts impermissibly, like any other member of the firing squad. Call this the *Overdetermination View* (OV). If the OV is true, the common belief that good intentions are irrelevant to permissibility is false or at best true in a more limited range of cases.⁴⁸

In this chapter I argue against the OV and in favor of a version of the view I shall call the *Counterfactual View* (CV). On the CV, roughly, the permissibility of an action which, along with some others, (otherwise impermissibly) overdetermines a wrong depends on whether it would be performed if the other actions had not been performed; if it would, it is impermissible, if it would not, it is permissible.⁴⁹ The reason why the executioner acts permissibly, therefore, lays not in his intention to alleviate the prisoner's suffering but in the fact that, given that intention, he would not shot the prisoner were it not for the others's shooting. That counterfactual fact, I shall argue, is not morally relevant in itself but only to the extent that it shows a more fundamental fact about the executioner's action, namely that it responds to the normative reason to reduce the prisoner's suffering.

The chapter has the following structure. In the first section I deal with conceptual issues. I distinguish different kinds of overdetermination and explain how to think about the notions of cause and harm in light of overdetermination cases. In the second section I discuss cases which motivate Tadros to accept the OV and which he appeals to in order to justify it. Examining their variants, I raise skepticism that the OV gets the things right. In the third section I defend the CV.

⁴⁸ It should be noted that, when Tadros defends the OV, he does not examine its implications for that belief but rather takes it to undermine the thesis that intentions, good or bad, are never non-derivatively relevant to permissibility.

⁴⁹ The original formulation of the CV comes from Parfit (1984: 70-1), who did not defend it apart from showing how it solves certain hard cases.

I first show how the CV accounts for the cases the OV is meant to account for and then revise it to account for some additional problematic cases. I then argue that the reason to accept the CV is that it is often a useful tool to tell us whether one acted upon or against some relevant normative reason for acting. In the final section I discuss another argument of Tadros, which states that, in overdetermination cases, actions that differ only in the intentions they are done with differ in liabilities and that, since permissibility depends on liability, intentions are relevant to the former. I question the first assumption appealing to a version of the CV.

1. Overdetermination and Harm: Conceptual Issues

Before we proceed to Tadros's arguments for the OV, a few conceptual clarifications are in order. We can distinguish different types of overdetermination by asking these three questions. First, do causes that overdetermine an event occur at the same time? Second, does each cause make the same causal contribution to producing the event? Third, do some causes interrupt the others from producing the event? For our current purposes, the event is some bad effect for a person, such as the harm of death, injury, property damage, and alike, and the causes are voluntary or intentional actions of other persons (rather than voluntary omissions, natural causes, or mere happenings).

Depending on the answer to the first question, we have simultaneous or non-simultaneous overdetermination. In the former, an event is a result of multiple causes occurring at the same time. Standard examples are of a person killed by a firing squad whose members shot him in chorus, a property destroyed by a fire started by arsonists who concurrently burn some gasoline, etc. In the latter, causes are temporally ordered; some are initiated first, others join them later, and

they all overdetermine the event. For example, a person is shot and, just before he would die, he is shot dead by another person.

The second question helps to distinguish symmetric from asymmetric overdetermination. In the former, the causal contributions are the same, not only in the sense that any cause suffices to produce the event but that they do it to an equal degree. The asymmetric overdetermination is much more frequent. However, for it to count as overdetermination rather than just a joint action, it must be true that each or most causes, though they differ in causal contributions, are sufficient to produce the event without the help of others. If at least two out of many shots of a firing squad would not, each by itself, have killed some prisoner but would have killed him only in tandem or together with other shots, that would be a joint but not an overdetermined killing.

Finally, the third question indicates the need for isolating the phenomenon of preemption. In preemption, an event need not be overdetermined since it can be the result of only one cause. That cause usually, though not necessarily, occurs before the others and it prevents them from producing the event. The main feature of the preempted causes is that each would have produced the event if the preemptory cause had not. In a classic case, A is killed by B who, seeing that C is about to kill A preparing to fire at him, shoots and kills A first. C would have killed A if B were not to kill him.⁵⁰ Clearly, more than one cause can prevent the others, overdetermining the event, e.g. B and D could trump C and kill A.

⁵⁰ This is sometimes called late preemption, as opposed to early preemption where B sees that C is about to kill A and shoots and kills A before C has even fired a bullet at A. Another case is that of the so-called trumping preemption, such as when B, a major, and C, a sergeant, simultaneously give orders to soldiers to kill A, which they do because B and not C tells them to do it – the major's order trumps, i.e. preempts, the sergeant's in virtue of military hierarchy. Finally, we should distinguish preemptive harming of victims from preemptive self- and other-defense, where those who initiate the harm suffer it in the end, and which are not instances of overdetermination. I shall discuss neither of these cases.

Simultaneous and non-simultaneous overdetermination can be symmetric or asymmetric. Although non-simultaneous preemption is much more common, simultaneous preemption is not impossible. But symmetric preemption is. Because the preempted events are causally ineffective, i.e. their causal contribution is null, preemption can only be symmetrical. In what follows, I shall sometimes refer to overdetermination in the broadest sense that covers any preemption.

Ordinary language use of the terms ‘cause’ and ‘harm’ suggest that, in overdetermination cases, each agent by himself causes harm and harms the victim, besides doing it together with others. On one influential, although revisionist, view of causation, neither agent independently causes harm (Lewis 1986). When A and B overdetermine killing of C, the most we can say is that they together kill C but not that A kills C and that B kills C. Similarly, on one revisionist view of harm, neither A nor B harms C; they only do it together (Parfit 1984: 69-70). These views spring from the counterfactual analyses of causation and harm, according to which something is a cause if the event would not have occurred if the cause did not occur and something is a harm only if the person harmed is worse off than he would have been had it not for that harming.

The consequences of these views might seem especially counterintuitive in preemption cases. The account of causation entails that, if A preemptively shoots B, A’s shooting is not the cause of killing of B since B would have been killed by C if A did not shoot. Strictly speaking, A does not kill B. Similarly, the account of harm entails that A does not harm B by preemptively killing him as long as he does not thereby make B worse off (e.g. C would have tortured and then killed B if A had not intervened).

It might be said that, even if a technical notion is very counterintuitive, we have sufficient reason to hold to it as long as it allows us to ask important metaphysical or moral questions more

clearly than its ordinary language analogue. But if fails in that respect or creates more confusion, we have decisive reason to stick to ordinary language.

To illustrate this, consider moral questions about preemption. The revisionary account of causation is immediately discredited. We should be able to ask whether A acts impermissibly by killing B, i.e. by preemptively causing his death. We cannot do this without appealing to causal notions.⁵¹ Tadros (2014a: 186, 188) argues that the same is true of the revisionary account of harm since it makes it difficult to ask who owes compensation in preemptive cases – the one who did the harm or the one who would have done it had he not been preempted? The difficulty reappears for questions about the former's liability to compensation and the relevance of his knowledge and intention to the permissibility of his action.

We could in principle ask these questions without appealing to the notion of harm. We could ask, for example, whether the preemptor wrongs the victim when he kills her even though he does not thereby make her worse off. We could ask whether such killing is impermissible, even if it is in one respect beneficial for the victim (the same goes for questions about liability). Tadros (2014a: 186) agrees. However, he adds that we should use the notion of harm for the sake of clarity. He (2014b: 13-4) proposes to distinguish harming in particular from harming, overall, in order to articulate our intuitions in preemption cases. For we should be able to say that, if A preemptively kills B, thereby saving him from being tortured and then killed by C, A harms B (relative to the world where A preempts C but not by killing B or not by making him worse off than he would ordinarily be) as well as that A benefits B (relative to the world where A does not intervene to preempt C). In short, A harms B in particular but does not harm him, overall.

⁵¹ This motivated Lewis (1986) to revise his views in favor of the notion of quasi-dependence, which he eventually also rejected. For the purposes of our discussion, it is non needed to describe these subtitles.

We might prefer to ask whether A acts impermissibly by harming B in this respect rather than whether A acts impermissibly by killing B but not by making him worse off, although these are essentially one and the same question. The former is closer to what we would ordinarily say but that does not matter much. As long as we do not talk past each other, the conceptual debate about harm is not important for the moral issues we strive to resolve. Unraveling the conceptual issues is neither necessary for understanding nor for defending the OV or the CV.⁵²

That being said, I shall follow ordinary language in saying that the preemptor harms the victim even if he makes her no worse off. I shall also say that each agent comprising a group that overdetermines a harm harms the victim by himself too.

We can now proceed to Tadros's arguments for the OV.

2. The Case-Based Argument

Tadros has two arguments. The first one is motivated by the contrastive pairs of overdetermined wrongdoing differing only in the agents's intentions. It states that, because these cases intuitively

⁵² I should stress two points to avoid misinterpretation. First, I am not saying that the conceptual debate about harm is irrelevant for moral issues whatsoever. I leave that as a hypothesis. Tadros (2014b: 171-2) argues that, since the magnitude of compensation for wrongdoing depends on the magnitude of harm done, we have first to understand what harm is. However, those believing that an agent wrongs the victim when he preemptively kills her but does not make her worse off and, therefore, does not harm her, can say that the victim is owed compensation due to wrongful but harmless killing. Tadros (2014b: 188) writes that they cannot make such claims when the preemptor is unaware of preemption. I fail to see why. The revisionists about harm could say that an agent is harming someone not if he is making that person worse off but if he justifiably believes that he is doing so. Even if they do not tailor their claim in this way they could say that, given his knowledge, the preemptor commits a murder and thus owes compensation for the same reason any murderer does. Similar agreements can be reached regarding different moral issues despite the conceptual disagreements about the notion of harm.

Second, I am not saying that the counterfactual analysis of the notion of harm should be rejected altogether. As a matter of fact, the reconstruction I offered is in line with Tadros's (2014b: 188-91) strategy to capture ordinary intuitions about harm in preemption cases by considering a more broad set of counterfactuals where the victim fares better or worse as a result of preemption.

differ in permissibility, intentions are the reasons for the difference. More precisely, the argument seems to be:

- (1) An action which, together with others, (otherwise impermissibly) overdetermines a wrong is permissible just when it is done with a good intention.
- (2) If an action which, together with others, (otherwise impermissibly) overdetermines a wrong is permissible just when it is done with a good intention, then it is permissible because it is done with a good intention.
- (3) Therefore, an action which, together with others, (otherwise impermissibly) overdetermines a wrong is permissible (when it is permissible) because it is done with a good intention.

I shall call this argument *the Case-Based Argument* because its premises are justified solely by case-based intuitions. My strategy to oppose it will also be mostly case-based. The relevant cases are of two sorts: overdetermination (including preemption) and overdetermination under duress. I examine them in turn.

But first, two clarifications are needed. First, the relevant cases are of an overdetermined wrong, rather than simply harm. The victims have *prima facie* claims against being harmed. This rules out the cases where the victims are liable to be harmed and where, therefore, harming them is not wrongdoing. Second, the cases are of otherwise (i.e. intention-independently) impermissible overdetermined wrongdoing, or, simply, overdetermined wrongdoing. This is in contrast with cases in which it is otherwise permissible to wrong someone. For example, in *Bystander*, I redirect a runaway trolley from the main track with five innocent people trapped on it towards a side track

with one innocent person trapped on it. It might be argued that the one has a claim against being killed but that it is overridden by the claims of the five. It is easy to imagine overdetermination versions of *Bystander* (e.g. you and me redirect the trolley), but we should put those aside and focus on the relevant cases.

2.1 Overdetermination

In *Poisoned Pipe* (Tadros 2011b: 159-60), a boss offers a financial reward to anyone who kills V. If more than one person kills V, the reward will be split between them. Motivated by the reward, A and B put poison in different water pipes leading to V's home. They do it independently of each other though aware of each other's action. Each puts in enough poison to kill V. A's poison alone or B's poison alone would have led to a very slow and painful death for V. However, their poison together kills V swiftly.

Tadros (2011b: 160) claims that, since A's and B's actions are identical and they act with identical intentions, they either both act permissibly or both impermissibly. Intuitively, they both act impermissibly, even though, if either of them had not acted, V would be even worse off.

Consider now *Benevolent Variation*. Everything is the same as in *Poisoned Pipe* except that, seeing that A is going to put the poison to kill V, B wishes to help V. He cannot prevent V from being killed so he puts exactly the same amount of poison as he puts in *Poisoned Pipe* in order to help V die a swift death. V is killed swiftly.

The intuition that moves the argument is that B acts permissibly. However, if that is true and if B's action is identical to his action in *Poisoned Pipe* except for the intentions they are done

with and if the latter action is impermissible, then what makes the difference in permissibility is B's good intention, or so Tadros (2011b: 160) claims.

Tadros (2014a: 12, 17-8) makes the same point in a preemption case, *Preemptive Killer* (cf. Parfit 1984: 70). Here X tricks V into drinking poison that causes a painful death. Before the poison has any effect, Y kills V painlessly because he hates V and wants to be the one who kills him. Y acts impermissibly even though (in one respect) he makes V better off.

Tadros contrasts this with *Pain-relief Variation*, where Y kills V intending not to kill him but only to make him avoid suffering. There is no other way for V to be made better off other than to be killed painlessly. Tadros's (2014a: 18) intuition is that Y acts permissibly. The reason for the difference, on his view, is in Y's intentions.

There are three ways to respond to this line of reasoning: first, to deny the difference in permissibility between *Poisoned Pipe* and *Preemptive Killer*, on the one hand, and *Benevolent Variation* and *Pain-relief Variation*, on the other; second, to accept the difference but show that it pertains even if the agent's intentions are similar; third, to accept the difference but try to explain it by factors other than intentions. I find the first reply *prima facie* unappealing and the other two tenable. However, as I shall argue, it is the former reply that we have most reason to accept.

It is true that the actions in question are permissible when done with good intentions but it is not true, I believe, that they are permissible only when that is the case. If so, the first premise

of the Case-Based Argument is false.⁵³ Consider first *Benevolent Variation II*, where B puts the poison in V's pipe intending to make V die a swifter and less painful death but only for the sake of others praising his deed.⁵⁴ Some might say that B's intention is bad. Although he intends to benefit the victim, the intention is not good since it stems from a vainglorious motive. And yet it is very intuitive that B acts permissibly.

It can be replied that it is not B's intention but his motive that is bad. On that view, an intention to benefit cannot be bad, no matter the motive behind it. This is a legitimate point but it would take us towards complex questions of when and why an intention is bad, which cannot be answered here. It might be said that, even if the point is correct, Tadros cannot appeal to it since, in *Benevolent Variation*, both A and B intend to kill V. A does it out of a desire to get the reward and B out of a desire to help V. However, Tadros could reply that B does not intend to kill V but intends only to alleviate V's suffering, though he foresees with certainty that he will kill V. Here one might press Tadros with a version of the case where B intends only to get praised and merely foresees alleviating V's suffering as its side effect. The intention seems bad but the action does not seem impermissible. For the same reason as above, that would move the debate too far.

Be that as it may, *Benevolent Variation II* at least shows that the OV fails as a view about the relevance of good motives to the permissibility of overdetermined wrongs.

⁵³ There is another objection to the first premise. Consider a reverse prisoner's dilemma: several persons intend to minimize harm to the victim justifiably believing that others intend to harm him but the victim ends up worse off than if no one had acted. Did they act impermissibly? I believe they did not. They thought that they were minimizing harm and they could not reasonably expect that that would lead to a worse outcome. But those who hold that not only intentions but beliefs are irrelevant to permissibility would say that, despite their good intentions and the fact that they could not know better, these people acted impermissibly (cf. Thomson 1986: 173-91, 1990: 171-3, 229-34). If so, the first premise is false. I cannot go into the debate about the relevance of beliefs to permissibility but I think that, in its action-guiding sense, permissibility must depend on what one can reasonably be expected to believe is true, whether or not that is in fact true (cf. Scanlon 2008: 47-52, Zimmerman 2008). If that sense of permissibility interests us here, as I think it does, this objection does not succeed.

⁵⁴ A similar version can be constructed from *Pain-relief Variation*. The same is true of preemptive analogues of other versions of *Benevolent Variation* I discuss below.

The critics of the relevance of intentions to permissibility might be drawn to cases like *Benevolent Variation III*, where everything is as in *Benevolent Variation* but B kills V upon his request secretly intending only to kill him.⁵⁵ B hates V so much that he prefers to kill him even if V will suffer less as a result of it.

That B acts permissibly in *Benevolent Variation* is indisputable; that he acts permissibly in *Benevolent Variation III* is not. The latter is the point of departure between the critics and most defenders of the relevance of intentions to permissibility. I, for one, am strongly inclined to think that he acts permissibly. I also think that that intuition can be defended but this is not a place for it. To show that some overdetermined wrongdoings done with bad intentions seem permissible to many people, it is sufficient to grant that the disagreement is reasonable and that the competing intuitions are at least equally compelling. Therefore, the first premise is, *pace* Tadros, far from intuitively acceptable.

Moreover, it seems to me that both sides of the debate might reach agreement that the first premise is false. The discussion about the relevance of intentions to permissibility can be avoided considering an intentional action done without any intention which, together with others, overdetermines some wrong. For example, in *Benevolent Variation IV*, everything is just like in *Benevolent Variation* but B kills V not intending anything. B found himself in situations like *Benevolent Variation* so many times in his life that they became habitual to him, just like shifting from one to another gear is to an experienced driver. In the first couple of hundreds of killings he intended to help the victims but in the later ones he acted habitually, not intending anything. Still, his later actions seem permissible.

⁵⁵ The secrecy factor is to exclude potential additional harm or offense that V might have by coming to know that he will be killed out of malice (cf. Kamm 2014: 482-3).

It can be replied first that, since B's initial intentions were good, his initial actions were permissible in virtue of them, and only later on he started acting impermissibly. This is very hard to believe. If there was a clear point where his actions turned habitual and lacked any intent, it is implausible that it was the point where they stopped being permissible. It can be said that his later actions are permissible in virtue of his earlier intentions. This must suppose some kind of transfer of the moral relevance of good intentions from past to future actions. I see no way to make sense of it. The remaining reply is to deny the possibility of intentional but unintended actions. This does not work either and for several reasons, one of which is that it implies that all foreseen side effects intentionally caused or allowed are intended (Anscombe 2000: 1, Harman 1976: 433-4, Ginet 1990: 76, Bratman 1987: 123). Those who, like Tadros, rely on the conceptual distinction between intention and mere foresight cannot embrace that inference without undermining their own position.

We can now turn to case-based reasons to doubt the truth of the second premise. Suppose that, in *Benevolent Variation V*, everything is the same as in *Benevolent Variation III* except that V requests B to kill him and B does it intending to fulfill V's request to ease his suffering, merely foreseeing that he will kill V. The reasons why B's actions is permissible is not his good intention but the fact that he fulfills V's justified request. It would be one thing too many to postulate B's good intention in conjunction with him fulfilling V's request as necessary for permissibility.

However, if acting with a good intention is not what accounts for permissibility in request cases, why is it what accounts for it in request-free cases, such as *Benevolent Variation*? We need an explanation why good intentions are an irrelevant consideration in the former but crucial in the latter since *prima facie* there is no reason for the difference.

The defenders of the relevance of intentions might contemplate the following reply. They might say that, even if *Benevolent Variation III* is a permissibility case, it does not refute the first premise; it only shows its limits. The premise needs to be qualified, stating that, absent morally weightier factors, an action which, together with others, overdetermines a wrong is permissible just when it is done with a good intention. One such factor, they might add, is a request of a sort that V gives in *Benevolent Variation III*. Assuming that the request is not conditional on acting with a good intention, it would override the claim against acting with a bad intention. B's action would then be permissible because V's request overrides the supposed wrong of that action being done with a bad intention.

This might lead them to claim further that the second premise needs to be revised too and that the actions in question are permissible due to good intentions except in cases where morally weightier factors are present. Although B's action in *Benevolent Variation V* is permissible due to V's request, that does not show that the second premise needs to be abandoned but that it needs to be revised, excluding request cases from its scope. This would explain the difference between the relevance of good intention in *Benevolent Variation V* and *Benevolent Variation*.

However, the analogy is incomplete in two relevant respects. On this proposal, the reason not to act with a bad intention in *Benevolent Variation III* competes against and is outweighed by the reason to fulfill V's request to kill him (granting that the former enters the weighing at all). But in *Benevolent Variation V* the reason to act with a good intention does not even count on a *pro tanto* level in favor of the permissibility of B's action. And even if it did, it would not weigh against the reason to fulfill V's request but rather add up to it. But as said above, the conjunction of the two is superfluous.

It might be said that it is not impossible that a consideration plays a crucial role in one context but no role in a slightly different context. This is true but because there is no reason to suppose that this is what happens between *Benevolent Variation* and *Benevolent Variation V*, we need a story that goes beyond the case-based judgments. In its absence, we have good reasons to doubt the truth of the second premise of the Case-Based Argument.

I shall now discuss another kind of examples Tadros invokes to support this argument.

2.2 Overdetermination under Duress

The following examples concern actions which overdetermine a wrong and are performed under duress. They have two relevant features. First, some agents are motivated by and intend to avert credible threats while others are motivated in part or only by wrong reasons and intend harm or similar bad effects. Second, the actions non-simultaneously and non-preemptively overdetermine the wrongs. When Tadros (2011b: 158) considers such cases, he takes them exclusively as duress cases. But since they have the standard properties of overdetermination, that term is appropriate as well.

Consider first *Nazi Propaganda* (Tadros 2011b: 158). It is Nazi Germany. The only way to avoid having a family member sent to a concentration camp is to distribute Nazi propaganda. Suppose two persons distribute the propaganda, call them A and B. In the first case, they both do it as committed Nazis, intending to support the cause. Intuitively, they both act impermissibly.⁵⁶

⁵⁶ It is unclear whether A's and B's contribution is independently sufficient for the Nazi cause as we do not know what the cause is. If not, *Nazi Propaganda* is an instance of joint action rather than overdetermination, and it then cannot support the OV. For the sake of simplicity, I shall assume that their contributions are independently sufficient.

In the second case, A does it intending to advance the Nazi cause; B does it intending to save his family from death. Tadros (2011b: 158, cf. n. 24) claims that, assuming that distributing the propaganda made a trivial difference to the Nazi cause, the fact of duress not only excused those who did it but also rendered their actions permissible. But if distributing it was done with the intent to help the Nazi cause, that was not only inexcusable but also impermissible. Even if A and B distributed the propaganda by the same means and to the same effect, B acted permissibly but A did not. On Tadros's view, the difference is explained by the fact that B acted with a good intention.⁵⁷

Similarly, in *Gang Robbery* (Tadros 2011b: 158), a gang of three (call them A, B, and C) robs the post office. The gang is ruthless; their members threaten to execute the families of each other if someone pulls out. In the first scenario, no one pulls out because they are all committed to robbing the post office. Intuitively, they all act impermissibly.

In the second scenario, again no one pulls out. A and B are committed to robbing the post office but C does it unwillingly, only in order to avert the threat to his family. C realizes he did the wrong thing in joining the gang and is reluctant to go ahead. Faced with a dilemma between committing two wrongs, robbing the post office and letting his family get killed, C opts for the former intending to prevent the latter. No doubt he is largely responsible for the dilemma but, as Tadros (2011b: 159) notes, it must be that at least one of his options is permissible, and, clearly, it is the latter. Since C acts identically to A and B and the only difference between their actions is

⁵⁷ *Nazi Propaganda* is probably modeled on a real case, *R v. Steane* 1947 KB 997, where the defendant's intention to save the life of his family sufficed for his acquittal for distributing Nazi propaganda. About it, Duff (1990: 93) claims that the defense should have been duress rather than the lack of treasonous intent. If intent was crucial, Duff argues, we could by parity of reasoning acquit any person who did the same but not under duress and in order to earn money, saying that he did not intend to distribute the propaganda. (On a different note, Duff thinks that both of them must intend to distribute the propaganda, which is false.) Some of my criticism of Tadros's appeal to *Nazi Propaganda* to justify the OV follows a similar line of reasoning.

in the intentions, C's intention explains why his action differs in permissibility from A's and B's, or so Tadros concludes.

As before, several replies are open. The first group of replies state that it is permissible to distribute Nazi propaganda or to rob the post office even if one's intentions were bad in other ways, being, say, to be seen as a savior of one's family or to frighten one's family for exposing oneself to such a risk (Kamm 2012: 116-7), or even if one acts habitually, with no accompanying intention. The reasoning is the same as above.⁵⁸

The second group questions whether these cases support the second premise. It states that appealing to good intentions is not the best way to explain the permissibility of acting in such duress cases.

As Tadros (2011b: 158, 24) acknowledges, giving in to threats and harming innocents is permissible, and perhaps even morally required, if the threats are credible and the alternative is a considerably greater harm befalling these or other innocents. Duress is normatively mitigating; it silences the reasons not to harm. For the sake of the argument, we can assume that this is true of *Nazi Propaganda* and *Gang Robbery*. It is then *prima facie* permissible to act in those cases due to the duress factor. This is so independently of whether one acts with the intention to avert the harm to one's family. Acting with that intention seems permissible not because of the intention but because of the duress factor. To add the intention as necessary for permissibility is, as before, superfluous.

⁵⁸ *Nazi Propaganda* and *Gang Robbery* are different from pure overdetermination cases because the agents cause harm that is not beneficial for some but not all the victims, namely only for their families. For that reason, there are no request versions of these cases similar to those of pure overdetermination. The agents cannot act on the requests of their families to benefit them but cause harm to them only. Other requests would not suffice for permissibility.

This gives us another reason to doubt the truth of the second premise. However, we must now explain the impermissibility of acting with bad intentions in *Nazi Propaganda* and *Gang Robbery*. If the duress factor accounts for the permissibility of such actions with good intentions and if the factor is present when the actions are done with bad intentions, why are not they also permissible? In the next section, I argue that, in one sense, they are.

3. The Counterfactual Alternative

3.1 How the CV Explains the Cases

According to the CV, an action which, together with other actions, (otherwise impermissibly) overdetermines a wrong is impermissible just when and because it would be performed if the other actions had not been performed (and permissible just when and because it would not be performed if the others had not been performed). The CV is implied by a more general view of what it is for a member of a group of people that together harm a victim to act impermissibly – that that person is a member of the smallest group of people of whom it is true that, if they had all acted differently, the victim would not have been harmed (Parfit 1984: 71).

Group is used here as a technical term; it need not consist of more than one person (see Parfit 1984: 72). If X tricks V into drinking poison that causes a painful death but before the poison has any effect Y kills V painlessly to save him from suffering, X is the smallest group since, if he (and no one else) had acted differently, Y would not have killed V, so V would not have been harmed. The same cannot be said about Y unless he would have killed V even if X did

not poison V. It would then be true that V would not have been harmed only if X and Y both acted differently but not if only one of them did. In that case X and Y are the smallest group.

The attribute smallest is needed to specify the relevant counterfactuals as we do not want to say that one acts impermissibly in overdetermination cases if one is a member of *any* group of which it is true that, had its members acted differently, the victim would not have been harmed, for that is true of all people. In the last case, V would not have been harmed not only if X and Y had acted differently but if any other person in the world had acted differently. But certainly, the rest of the world's population has nothing to do with harming V (cf. Parfit 1984: 71-2).

The CV explains our intuitions about Tadros's original pairs as well as their variants. The CV tells us that, in *Poisoned Pipe*, both A and B act impermissibly since each would kill V were the other not to kill him, but that, in *Benevolent Variation*, B acts permissibly since he would not kill V were A not to do it. In the former, both A and B are the smallest group which harms V but in the latter, it is only A who is that group. It might be tempting to say that it is B's intention to alleviate V's suffering which explains why he is permitted to kill him but that intention is only derivatively relevant, pointing towards the fact that B would not have killed V if it were not for A poisoning him. I shall explain later why that counterfactual fact matters morally. Similar claims apply to *Benevolent Variation IV* and *V*.

The CV also explains the intuition that B acts permissibly in *Benevolent Variation II*. B would not put the poison in V's pipe if A had not done the same. Motivated *only* by vainglory, he would not aim to act impermissibly as he would not then receive praise and admiration. No one would think that B deserved praise in poisoning V had it not been for A trying to kill V (no one would think that if they knew the real reason behind his action but that is beside the point).

Regarding *Benevolent Variation III*, I claimed that there is intuitive appeal in assuming that B acts permissibly but acknowledged that the critics of the relevance of intentions will deny this. Since B kills V only after V's request to be killed, he would not give V the poison if it were not for the occasion to realize his malicious intent through an otherwise permissible acting. The CV then implies that B acts permissibly.

Tadros would reject this implication. I shall return to such cases below but, for now, note that Tadros accepts the OV because he thinks that it explains the intuitive difference between *Poisoned Pipe* and *Benevolent Variation*. As there is no such clear difference between *Poisoned Pipe* and *Benevolent Variation III*, Tadros cannot appeal to the OV. To resolve the disagreement about whether *Benevolent Variation III* is a case of permissibility, it is not enough to appeal to the OV or the CV. In assessing either view in their ability to explain some intuitions, we should focus on cases about which our intuitions concur.

Consider now duress cases. I claimed that duress explains why acting with intention to avert the threats is permissible and that that intention is explanatorily unnecessary. Since those who act with that intention in most cases would not act were they not under duress, to find out whether one acts permissibly, we need only to ask what one would do were he not under duress.

However, in some duress cases, the fact that others act in the same way is relevant for one's motivation. Suppose that, seeing that A is going to put in the poison to kill V, B puts the poison in V's pipe but he does it motivated not only by a desire to help V suffer a less painful death but by C's threat to kill his family if he does not help V or, in case A's poison turns out to be ineffective, to kill his family if he does not kill V with a larger dose. In that case, B would put the poison if C were not to threaten him but also if he were but A were not to act. Since B would

act permissibly either way, the relevant counterfactual is the one where B is not under duress *and* A does not aim to kill V.

We can now turn to *Nazi Propaganda* and *Gang Robbery*. The former is an easy case as the threat is external. If A were to do it because he shared Nazi values, most probably he would do it even if he was not under duress (it seems to me that the case is best understood so that A does not even act under duress), so the duress version of CV entails that A acts impermissibly despite being under duress. Nothing of that sort if true of B.

However, in *Gang Robbery* the threat is not external in this sense. A, B, and C threaten to kill each other's families unless they commit the robbery. Because C realized that the threat was wrong, he would have not committed the robbery if A and B did not threaten him. The CV then implies that he acts permissibly. Since A and B intend only to get the financial gain, they would most probably proceed even if C did not threaten them, so they acted impermissibly.

The CV explains all the cases the OV explains and more. However, in some other cases, it clearly fails. I shall examine them now.

3.2 Problems with the CV

There are two types of problematic cases. The first shows that the counterfactual independence of one's action from others's is insufficient for impermissibility. Consider a variant of *Preemptive Killer – Corruption Variation* (Tadros 2014b: 24). X sets out to poison V. Y is influenced by X's behavior and, in order to impress X, he poisons V swiftly and painlessly. Had Y not poisoned V, X would have done it in a way that would make V suffer a more painful death. The important

novelty in *Corruption Variation* is that, if X had not prepared the poison, Y would not have been influenced by it and would have never killed V. Y is not the smallest group of whom it is true that, had they all acted differently, V would not be harmed. The CV then implies that Y acts permissibly, which Tadros (2014b: 25) finds intuitively implausible.

Regarding such an interpretation of *Benevolent Variation III*, I said that there is an intuitive appeal in claiming that one acts permissibly with the intention to harm a victim if one prevents greater harm to occur and one would not do it were it not for the occasion. The defenders of the relevance of intentions disagree. As I have said, the intuitive disagreement about such cases is part of a wider controversy about the relevance of bad intentions to permissibility and it does not affect the debate about their relevance in overdetermination cases. For the latter debate to start, we must agree that there is a difference in permissibility between *Preemptive Killer* and *Pain-relief Variation* and other similar pairs. But we do not need to agree about *Corruption Variation* and similar cases.⁵⁹

That being said, the CV has implausible implications when we turn to acting with bad intentions which would not prevent harm or benefit the victim in other relevant ways. Suppose that, in *Corruption Variation II*, to impress X, Y poisons V in a way that causes V to suffer a much more painful death than if X were to poison him. Y would have never harmed V if X had

⁵⁹ It pays to compare them with *Bad Bystander* (see Kamm 2007: 131-2, Liao 2012: 716, Tadros 2011b: 156-7). Like in *Benevolent Variation III* and *Corruption Variation*, in *Bad Bystander* the agent abuses the opportunity by acting in an otherwise permissible way but intending only to kill the victim. He redirects a runaway trolley away from five innocent persons trapped on the main track towards one innocent person trapped on the side track but because he hates the one and not because he wishes to save the five, something he would not do if the one were not on the side track. Since there is an intuitive disagreement about whether he acts permissibly, to proceed we must engage in further theoretical arguments rather than insist that our views should explain one of the positions. Similarly, the CV need not explain why one acts impermissibly in *Benevolent Variation III* and *Corruption Variation* for there is no good reason to suppose that he does.

not made the poison. It is then true that, if X had not acted, Y would not have acted. According to the CV, Y acted permissibly, which cannot be true.

The second type of cases shows that the counterfactual independence is not necessary for impermissibility. Suppose that, in *Benevolent Variation VI*, everything is just like in *Benevolent Variation* but if A were not to put the poison to kill V, C would do it, causing V to suffer much more than if A were to kill him. B wishes to help V anyway and would do the same if C were to kill V. But since B would help V even if A had not acted, according to CV, he acts impermissibly, which is clearly false.

To avoid such cases, we should revise the CV. According to the CV*, an action which, together with other actions, (otherwise impermissibly) overdetermines a wrong is impermissible when it would be performed if the other actions had not been performed *or* (only) when it would be performed even if the other actions are performed but there would be no sufficient reason for its performance, and an action is permissible when it would not be performed if the other actions had not been performed *or* (only) when it would be performed even if the other actions had not been performed but there would be a sufficient reason for its performance.

The CV* accounts for all the cases the CV accounts for and more. For example, there is, intention-independently, a sufficient reason for Y's action in *Corruption Variation*, namely that it prevents V's greater suffering. But there is no such reason in *Corruption Variation II* because Y's action harms V more than X would harm him otherwise. Indeed, there is a decisive reason for Y not to act. The CV* correctly implies that the former but not the latter is permissible. Similarly, in *Benevolent Variation VI*, there is a sufficient, perhaps a decisive, reason for B to harm V if A were not to harm him, since V would be then harmed more by C.

At least three objections can be brought against the CV*. First, it can be said that, since the principle leaves open the question which reasons are sufficient, it risks being trivial. But this is not so. The CV* is not meant to be the final word on when and why an action is permissible or impermissible in overdetermination cases. Since such cases are diverse, the CV* has to retain some level of abstraction. I doubt that the gap regarding when such actions are permissible or impermissible can be filled in with a less abstract principle than the CV*. But I do think that we can provide a unifying explanation about why such actions are permissible and impermissible. I shall say more about it in a moment.

Second, it can be said that, the CV*, like the CV, excludes bad intentions as the relevant sufficient reasons against an action's permissibility. This is true but only tentatively. As already said, I leave it open whether bad intentions affect permissibility.

Third, it can be said that the revision is ad hoc. The CV* fits well with our intuitions but this is not a good enough reason to accept it. This would be true if there was no rationale for the revisions. But I think there is.

3.3 A Rationale for the CV: The Normative Reasons View

Why does the counterfactual dependence or independence of an action matter morally? How is it relevant for explaining why that action is permissible or impermissible?

There are two questions here. First, why does the fact that, in *Benevolent Variation* (or *Poisoned Pipe*), A would kill V in the world where B does not kill V explain the impermissibility of A killing V in the actual world? Second, why does the fact that, in *Benevolent Variation*, B

would not kill V in the world where A does not kill V explain the permissibility of B killing V in the actual world? (The same questions can be asked of other such cases.)

Counterfactual independence *does not* matter morally in itself. It is useful only because it often reveals morally important facts about an action in the actual world. In *Benevolent Variation* (if we first take it as a case of non-simultaneous overdetermination), these facts are that A does not act on the normative reason to ease V's suffering but is the cause of that reason appearing, and that B does act on the normative reason to ease V's suffering and it is due to that reason that his action is permissible and not due to his good intention. B gets the normative reason to ease V's suffering only after A puts the poison.

Since, in all our cases, victims are innocent, non-threatening people, who do not consent to be harmed, and for whom further life is a good, *prima facie* there is a decisive reason not to harm them. The reason to harm them appears only when harming would be beneficial for them, i.e. when it would minimize another, greater, harming.⁶⁰ We have that reason in the same way we have a reason to prevent unnecessary suffering, other things being equal. Since A is the one who initiates harming V, he does what he has a decisive normative reason not to do. But B acts only on condition of helping V. B does what he has a sufficient, and perhaps a decisive, normative reason to do. His action is then permissible not due to his good intention but due to the fact that

⁶⁰ I assumed that the reason to minimize harm appears when victims would be harmed voluntarily and intentionally by some people. But they can appear even if harm would be a result of mere happenings, such as muscle spasms, or natural events, such as earthquakes or killer bee stings. It would not make a difference to the impermissibility of Y's action if he killed V out of spite after V had been stung by a bee whose poison would cause him to suffer a painful death (cf. Tadros 2014b: 22).

Tadros (2014b: 22-3) believes that the CV cannot make room for the last claim since it has to distinguish the preemption of harm which is a result of someone's action from the preemption of harm which is a result of some natural event. Presumably, he thinks this because the CV's conclusion is premised on counterfactual actions of other people. Since most relevant cases concern group behavior, this is the most appropriate way to formulate the view but it does not exhaust all there is to it. Nothing prevents us from widening the CV so that it states that one acts impermissibly in overdetermining a harm together with a natural event or in preempting a harm caused by it if and only if one would act in those ways had it not been for the natural event. What is more, since acts are a species of events, this is a straightforward widening of the CV.

he acts upon the normative reason to ease V's suffering. B's intention to act in that way is good because his action is responsive to that normative reason and not the other way around.

Similarly to A in *Benevolent Variation*, those who would harm the victims only when others would harm them too but who cause needless harm (e.g. Y in *Corruption Variation II*) act impermissibly because they do what they have decisive reason not to do, namely cause needless harm. And similarly to B in *Benevolent Variation*, those who would harm the victims even if the initiator of harming would not harm them but someone else would instead (e.g. B in *Benevolent Variation VI*) act permissibly because they do what they have decisive reason to do, namely ease suffering.

Call the view that the (im)permissibility of harming in overdetermination cases is due to normative reasons (not) to harm the Normative Reasons View (NRV). The NRV is a rationale for the CV. It explains why the CV gets the verdicts about permissibility right (when it gets it right, as it does in many, though not all, cases). However, the NRV is not without controversy.

Suppose now that, as Tadros meant it to be, *Poisoned Pipe* and *Benevolent Variation* are cases of simultaneous overdetermination: in both cases, B puts the poison at the same time as A. They then each have the normative reason not to harm V but also the normative reason to put the poison (provided by the fact of easing V's suffering), given that the other will do the same. Now if it is only normative reasons that matter for permissibility (but not intentions), we are lead to a striking conclusion: *Poisoned Pipe* and *Benevolent Variation* must come out the same in terms of permissibility, depending on which of the two normative reasons is stronger. Since it is plausible that each A and B, once they decided to act against the normative reason not to harm V, i.e. once they decided to put the poison in V's pipe, have a strong normative reason to put the poison,

given that the other does the same. They each have a strong normative reason to put the poison provided by the fact that that will ease V's suffering. We must then conclude that A and B each act permissibly in *Poisoned Pipe* and *Benevolent Variation*. I am aware that this goes against the intuition of many people. And I acknowledge it as a difficulty in my view. However, let me now tentatively suggest why this intuition might not be of fundamental moral importance.

We can assess the permissibility of group actions, not only individual actions. A's and B's actions can be considered individually or together, as one complex action of putting sufficient poison to kill V swiftly. While each A and B have both the normative reason not to harm V and the normative reason to do it (once each decided to do it), they *together* have only the normative reason not to harm V. They together do not have the reason to ease V's suffering; it is only A and only B who have that reason since A's and B's individual actions do not suffice to kill V swiftly. But their complex action suffices to kill V swiftly, so there is no reason for them as group to ease V's suffering. If that is true, A and B together act impermissibly when they both put the poison in V's pipe, in both *Poisoned Pipe* and *Benevolent Variation*. I find this intuitively right. It seems to me that the intuition that some wrongdoing occurs in *Poisoned Pipe* might be a consequence of conflating the view that A and B together act impermissibly with the view that each of them acts impermissibly. But only the former is true, or so it seems to me.

Duress cases might be similarly explained. Consider *Gang Robbery*. Recall that, initially, C threatens to kill A's and B's children but later on he changes his mind and commits the robbery only in order to avert A's and B's equally serious threats. As before, I shall simply assume that in *Gang Robbery* (as in *Nazi Propaganda*) duress is a mitigating factor and that it renders actions of A, B, and C *prima facie* permissible (I doubt that duress can play that role in most such cases). I

shall also assume that A, B, and C each have the normative reason not to rob the post office and that, once they decided to do it, they each have the normative reason to rob it and that that reason is provided by the fact that each thereby averts the threats to his family. Now if it is exclusively that reason which determines the permissibility of their actions, A, B, and C each act permissibly in robbing the post office (even if A and B do it only to get the financial gain). However, it does not follow that, as a group, they act permissibly. They together have no normative reason to rob the post office provided by the fact that they together thereby avert the threats to their families. The threats are not external; A, B, and C is each a source of the threats. So they together do not have the reason to avert them. What they, as a group, have the normative reason to do is not to rob the post office. To some extent this might explain the powerful but likely misguided intuition that A and B each act impermissibly.

If the NRV is correct, the relevance of intentions in overdetermination cases is predictive and, therefore, derivative (cf. Scanlon 2008: Chs. 1-2, Moore 2005: 96-7). Because we get that prediction by asking what one would do in certain counterfactuals, the CV is often a useful tool for finding out whether one does what one has normative reason not to do or one does what one has normative reason to do. The CV* is just a more accurate tool since, in some cases, what one would counterfactually do is not the sole indicator of the permissibility of what one does. But the NRV is what fundamentally explains the (im)permissibility of actions in such cases.

The second premise of the Case-Based Argument is therefore false. Tadros (2014b: 30-7) has another, theory-driven, argument. In the next section, I discuss and reject it.

4. The Argument from Liability

Since this argument stipulates dependence of moral permissibility on moral liability, I shall call it *the Argument from Liability*. I believe that the following is the best reading of it:

- (1) One is liable just when one acts impermissibly.
- (2) In overdetermination cases of a kind considered, one is liable because one acts with a bad intention.
- (3) Therefore, in overdetermination cases of a kind considered, one acts impermissibly because one acts with a bad intention.

According to the OV, what explains the permissibility of an action which, together with others, overdetermines a wrong is a good intention behind it. However, the conclusion of the Argument from Liability is a stronger claim. It states that what explains the impermissibility of acting with a bad intention in overdetermination cases is that intention. The two need not come as a package. An action which overdetermines a wrong might be permissible just when and because it is done with a good intention but an action done with a bad intention, in identical circumstances, might be impermissible because of factors other than or in addition to the bad intention. It seems to me that Tadros (2014b: 36-7) defends the package view. If so, the Case-Based Argument is a defense of the part about the relevance of good intentions and the Argument from Liability is a defense of the part about the relevance of bad intentions.

The first thing to note is that the Argument from Liability is invalid. Even if one is liable just when one acts impermissibly and if the liability is due to one's bad intention, it need not be

that the impermissibility is also due to bad intention. Correlation of liability and impermissibility does not entail them being explained by the same phenomenon. However, I think that we should understand Tadros as claiming that, absent good reasons to the contrary, if the premises are true, the conclusion is true too. It is not that the conclusion must follow but that it is likely that it does. Let us then see whether the premises are true.

Two points need clarification. On Tadros's (2014b: 47-51, cf. McMahan 2009b: Ch. 1.1.) notion of liability, the fact that one is liable to harm means that he is not wronged by being harmed, not even *pro tanto*, even if he did not consent to it. Particular liabilities are the liability to be preventively harmed, to pay compensation for the harm done, and to be punished for it. Presumably, Tadros believes that the liabilities are not necessarily legally enforced (the punitive liability is then also moral, in the sense of liability to be blamed, condemned, etc.).

Second, Tadros (2014b: 31) claims that one can be liable for actions and, specifically, for wrongdoing or for threatening to wrong or attempting to do so. Having a bad intention is typically insufficient for liability. He (2014b: 36-7) then claims that, in *Preemptive Killer* (and, similarly, *Poisoned Pipe*), Y is liable to be harmed, so he must be acting impermissibly. But in *Pain-relief Variation* (and *Benevolent Variation*), Y is not liable to harm, so he must be acting permissibly. Since the cases differ only in Y's intention, the difference in liability must turn on intentions and so must the difference in permissibility. The critics of the relevance of intentions, Tadros (2014b: 37) claims, cannot explain the difference in liability between these cases.

Consider now the first premise. Even if the correlation of impermissibility and liability is typical, it does not follow that it is impossible to be liable for a permissible acting with a bad intention. One might make an analogy with moral culpability. It is common to hold one morally

culpable for a permissible acting with a bad intention, so it can be argued that, in the same way, one can be liable to preventive, compensatory, and punitive harm even for permissible acting if one's intention is bad.

The reply fails for two reasons. First, culpability is not necessary for liability. The main illustration is the case where one is liable to defensive harming even if one acts blamelessly but causes harm due to very bad luck. For example, a driver keeps his car well maintained and drives cautiously but, due to a freak accident, the car goes out of control and veers towards a pedestrian whose only way of saving his life is to bomb the car, thereby killing the driver (McMahan 2005: 393). Whether such cases succeed is a complex issue which I shall not discuss here.

But there is another problem. If liability depends on culpability and culpability depends on bad intentions, bad intentions are relevant to impermissibility via their relevance to liability via their relevance to culpability. The critics of the relevance of intentions cannot take that route without compromising their main thesis.

A different strategy appears more promising. We can grant the first premise and deny the second. For even if one can be liable to harm just when one acts impermissibly, one need not be liable to it because one acts with a bad intention. Assuming the CV is a true about permissibility in overdetermination cases, there is no reason why an analogous view would not be true about liability in such cases.

On that view, one would be liable for an overdetermined wrong just when and because one would harm the victim even if others were not to. The justification would be the same as the justification for the CV. The counterfactual independence of one's action would show whether one does what one has a decisive reason not to do. Moreover, this view could be revised in line

with the CV*. If the critics of the relevance of intentions accept the CV*, they can successfully explain the difference in liability between *Preemptive Killer* and *Pain-relief Variation* and similar pairs appealing to its liability analogue.

Conclusion

I argued that good intentions are irrelevant to permissibility in overdetermination cases. It might seem that they are relevant because of an intuitive difference in permissibility between actions which causally overdetermine a harm, some of which are done in order to cause the harm and other only in order to benefit the harmed person by such harming. I argued that the intuitive difference can often be accounted by the fact that the latter action would not be done if it were not for the former. The counterfactual absence of other actions is morally relevant to the extent that it often reveals the fact that one acts upon the normative reason not to harm the victim for no need or upon the normative reason to ease harm. I argued that it is such normative reasons that determine the (im)permissibility of actions in overdetermination cases. I admit that my view has counterintuitive implications in some cases. But I suggested that the permissibility of actions in such cases differs on the basis of their assessment as being done by individuals or by groups and that that might explain the intuition my view goes against. I conclude that the common belief that benevolence does not make a wrongdoing morally acceptable remains unshattered.

4. Intentions and Inchoate Wrongdoing

Introduction

An act is an inchoate wrong just when it causally contributes to an independently wrong act (not necessarily as a causally necessary or sufficient condition for it). Inchoate wrongs are not wrongs by themselves but only in virtue of their relation with independent wrongs. For example, there is nothing wrong in buying rat poison. But if one buys it in order to poison his wife, then not only is poisoning his wife a wrongdoing but so is buying the poison itself (*Poison*, Scanlon 2008: 41).

Criminal law recognizes different types of inchoate crimes, the most familiar of which are attempted crimes, such as attempted murder, and the crimes of facilitation, such as knowingly loaning a car to someone for criminal purposes. Similarly, there are morally wrong attempts and facilitations. While the non-derivative relevance of intentions to inchoate criminal wrongdoing is recognized in statutes and by most criminal law theorists (but see Alexander and Ferzan 2009, Walen 2009, Chiao 2010), their relevance to moral wrongdoing is a matter of dispute. I shall argue that attempts and facilitations which are impermissible are not even in part impermissible due to the agents's intentions to commit the independent wrongs they lead to. I shall also offer an alternative account of their wrongness, defending and developing Scanlon's (2008) account.

The chapter is structured in the following way. I first present two direct arguments against the relevance of intentions to the impermissibility of attempts and facilitations. I then show the defects of the risk-based explanation of the impermissibility of inchoate acts. I conclude with interpretation and defense of Scanlon's account, arguing that the reason why some inchoate acts are impermissible is the reason why the independently wrong acts they lead to are impermissible. The inchoate acts are causally facilitating the latter.

1. The Intentions View

1.1 The No-Intention Argument

According to what might be called the Intentions View, the impermissibility of inchoate wrongs is non-derivatively due to the fact that they are done with intentions to do the impermissible acts they lead to (see Tadros 2014c). My first argument against this view is that, since some inchoate wrongs are done with no intentions, their impermissibility cannot be due to that reason. Call this *the No-Intention Argument*. The argument requires an analysis of the notions of intending, trying, and attempting. I shall start from trying.

Trying can be understood both as a mental and as a physical activity. If I try to remember the name of my favorite childhood cartoon, I need not do anything but think hard about it. But if I try to poison you, I shall perform some bodily movements, such as putting poison in your food or drink. Inchoate wrongs are wrongs of performing such physical acts. Thinking about the best way to poison you might, in itself, be blameworthy and even impermissible, though I shall not be concerned with it.

On some views (Armstrong 1973, Hornsby 1980), one always tries to do what one does intentionally. If that is true, inchoate wrongs are wrongs of trying to commit a wrong (cf. Hanser 2014). This seems to be true in the case of attempts since trying to act seems to be necessary for attempting to act.⁶¹ It is sometimes (Schroeder 2001) said that trying is not a part of behavior but an indication of uncertainty of its success. But suppose that I connect wires that activate a bomb. Although I am confident of success, the wiring is a time-consuming maneuver. It is warranted to say that the series of bodily movements which would result in activating the bomb are the acts of trying to activate it (in the sense of setting out to do it). As long as I do all that I can to activate it, I try to do it, even if my success is guaranteed (cf. Hornsby 2010: 19).

A better objection is that we often just act, exerting no effort. We do not try to act unless there is a difficulty in succeeding. This might be what Wittgenstein (1953, sec. 622) meant when he famously said: “When I raise my arm I do not usually *try* to raise it.” The same might be true of effortless facilitation of a wrong, e.g. pulling the trigger in order to kill. However, the reason why it might be thought odd to ask one who just shot a person whether he tried to pull the trigger is not that he could try to pull it only if it took him some effort but that asking that question *post factum* makes no sense. It is trivially true that he tried in the same sense it is trivially true that he began to pull it (cf. Grice 1989: Chs. 1-3). This is visible when one does a series of effortless acts that certainly end up in someone getting killed, e.g. splashing poison onto one’s food, drink, etc. It would be sensible to ask “Are you trying to poison him?” (in the sense of “Are you set out to poison him?”). But it would not be sensible to ask “Did you try to poison him?”, once the victim

⁶¹ It might not be sufficient for it. Criminal law distinguishes mere preparations from attempts. Purchase of balaclava with intent of robbing a bank is not an attempted robbery but entering a bank with it is, even if one is arrested just after one put it on his face. However, in both cases one tries to rob a bank in the sense that he sets out to do it. This is only to say that the distinction between merely preparatory acts and attempts is intuitive but not that it always has moral relevance. I believe that it does not and I shall say more about it later (cf. Duff 2012: 412-5).

got poisoned. That does not mean that he was not trying to poison him but that the question is misleading. I shall therefore assume that, whenever one commits an inchoate wrong, one does try in this sense to achieve an independently wrong aim.

Since trying is an action, it is something that can be intended (Hanser 2014: 138). Does trying necessarily entail intending (conceptually or at least as a matter of practical rationality)? Intuitively, it does not. One might try to do what one believes one hardly will do and one cannot intend, but only hope, to do it. When I buy a national lottery ticket, with one in a million chance of winning, I try but do not intend to win, at least as long as I am not overconfident about it. I might intend to try to win, try to win, but I cannot intend to win. Intending entails believing that there is good chance of success or at least not believing that failure is certain (Audi 1973: 387-9, Harman 1976: 434-7, Bratman 1987: 37-41, Mele 1989, 1992: 129-30, Brunero 2014: 35-8).⁶²

Another reason to think that trying does not imply intending are habitual attempts, which, like other habitual actions, are intentional but unintended. Imagine a basketball player coming every day to the mid-court and intending to try to make a basket but continuously missing. After some time, this becomes his habit; he no longer intends to try but tries. He does it intentionally, just like an experienced driver shifts from second to third gear intentionally but does not intend to do it each time he does it.

Finally, one can try to do inconsistent things at the same time so that, by doing them, he boosts up his chances for achieving one. But one cannot intend to do all of them (see Bratman 1987: 114-5, 1997: 183-5). I might try to undertake several jobs simultaneously—by filling out

⁶² Anscombe (2000: 93) says that, in some cases, one can intend not to do what one foresees with certainty one will do: “A man could be as certain as possible that he will break down under torture, and yet determined not to break down.” What the man intends here, I think, is not to break down as long as he can, but not not to break down, period (cf. Brunero 2014: 36). His determination not to break down, period, is at best a desire or hope.

the application forms, requesting references, going for interviews, etc.—but I cannot intend to undertake them all simultaneously. This is a straightforward entailment of the above thesis that one cannot intend what one takes to be impossible.

One might object that, since several agents can intend things they cannot simultaneously achieve, a single agent can intend them too (Yaffe 2010: 61). But I do not believe that each and every applicant on the job market can intend to undertake a job he believes he cannot undertake simultaneously with other applicants. The explanation for it seems to be some form of the Ought Implies Can principle, ranging over acting and intending but not over desiring and trying. The latter two are out of scope of that principle because they do not have a commitment to success component inbuilt in them, which is why it is not odd to desire or try to do the impossible.

It might be replied that trying does have such a component for its function is to “translate our intentions into behavior” (Yaffe 2010: 68). The argument behind this is that trying is a means to acting in the same way intending is a means to coordinating behavior, so trying and intending must always go together. One must intend to do what one tries to do. Granting that mental states are means to some effects, strange as that may sound, I do not believe that the argument is valid. Coordination is possible on the basis of others’s unintended intentional behavior. For instance, we rely on others’s habitual actions which are typically unintended. The argument does show that we often intend to do what we try to do, but it does not go further than that.

We therefore have good reason to believe that one can commit an inchoate wrong without intending an independent wrong. This seems obvious in case of facilitatory wrongs. Gun vendors do not typically intend to assist their customers in murder although they might justifiably foresee that some particular customers will commit murder with guns they sold them. In such cases, they

might be morally obliged to refuse to sell guns. Similarly, tobacco companies sometimes justify their business saying that they do not intend to get people addicted or to facilitate their premature deaths but to profit offering an enjoyable but dangerous substance at their own risk. Even if that is no good justification, it is plausible that the intent behind the tobacco business is not to harm people.

Consider next attempted wrongs. It is customary to distinguish complete from incomplete (or last-act from non-last-act) criminal attempts on the basis of whether the defendant did all that he believed to be necessary to commit of a crime (see Duff 1996: 119, Yaffe 2010: 25-6). We can make the same conceptual distinction in attempted wrongs. One might do both without intending to do the independent wrong they would lead to if they were successful.

For instance, in one actual case (*People v. Smith*, Calif. 2005), a defendant fired a single bullet into a moving vehicle, missing a mother and her baby in it. The defendant could not have killed both the mother and the baby since the bullet would have stopped once it killed the mother. However, he was convicted of two counts of attempted murder. The defendant objected that he intended to kill only the mother but the appeal was rejected. Suppose that he did try to kill them both. Given what is said above, he could not have intended to kill them both. He could have only intended to try to kill them. But that does not change the fact that he committed two complete attempted murders. The same would be true if he committed an incomplete attempted murder, pulling his gun out but being stopped by the police before he fired a shot.

The impermissibility of such actions cannot therefore be due to the fact that they are done with intention to do the independent wrong since they are not done with that intention. Therefore, the Intentions View cannot explain such inchoate wrongs.

1.2 The Comparative Argument

My second argument against the Intentions View is that, comparing impermissible inchoate act types with intention to do an independent wrong with act-types identical in everything except for being done with a mere foresight of the same event, we see that the latter remain impermissible. This gives us good reason to doubt that the reason of the former's impermissibility is the fact that they are done with intention to do the independent wrong. Call this *the Comparative Argument*. I shall illustrate it with two examples.

Return to *Poison* but complicate it in the following way. A man buys rat poison intending to kill rats and merely foreseeing, with high probability, that he will use some of it to kill his wife too (*Poison II*). In *Poison*, the act type is 'buying the poison in order to kill wife' and, in *Poison II*, it is 'buying the poison in order to kill rats foreseeing using it to kill wife'. Intuitively, both act types are impermissible. There is no reason to think that their impermissibility is due to different reasons. If so and if the first act type contains the intention to kill wife but the second one does not, the intention to kill wife is not the reason why the former is impermissible. A different way to put it is to take 'buying the poison' relative to the intention to kill wife and relative to the mere foresight to do it, holding everything else fixed. Either way, it is impermissible to buy the poison, so the conclusion is as before.

The same would be true in thwarted attempted wrongs. Suppose that, in *Poison III*, all is as in *Poison I* but police discovers the man's intent and arrests him before he gives the poison to his wife. In *Poison IV*, all is as in *Poison II* but the police predicts his behavior and arrests him before he gives the poison to his wife. Intuitively, in both *Poison III* and *IV*, the man attempts a

murder. Since he only intends to kill his wife in *Poison III*, the intention is not the reason why he acts impermissibly.

It might be objected that in *Poison II* and *IV* the man must intend to kill his wife. But we often do not intend what we foresee we are going to cause, even if we foresee it with certainty. I might foresee that, by going to run a marathon, I will wear down my sneakers. But certainly, I do not intend to wear down my sneakers; I intend only to go run a marathon. Similarly, as we saw, to commit an attempted wrong, one need not intend to bring that wrong about.

Another objection would be that actions in *Poison II* and *IV* are more wrong than actions in *Poison I* and *II*, even if they are all impermissible. The presence of intention to kill wife in the former pair, one might argue, accounts for the greater wrongness. For the objection to work, the purported extra wrongness cannot be due to the greater blameworthiness of actions in the former pair; there must be more reason against their permissibility. Whether there is more reason of that sort cannot be settled intuitively since many people (myself included) do not spot the difference in wrongness between these actions. We are then left with assessment of different justifications of the wrongness of intending harm *per se*. That would require intricate discussion that cannot be afforded at this place. However, if the line of reasoning above is plausible, then even if inchoate wrongs with the intention are more wrong than their analogues with a mere foresight, intentions are neither necessary nor sufficient for the former's impermissibility. This should suffice to make us seriously reconsider the Intentions View.

Finally, notice the difference between the Comparative Argument and the No-Intention Argument. The scope of the latter is limited. It proves that the Intentions View does not work for inchoate wrongs which lack the intention to commit the independent wrong. The Comparative

Argument, on the other hand, proves that the Intentions View does not work even for inchoate wrongs that are done with intention to commit the independent wrong. What does then explain the impermissibility of inchoate wrongs? I shall try to find the answer in what follows.

2. The Risk-Based View

The explanation that first comes to mind to most people is that impermissible inchoate acts are impermissible because they risk the completion of independently impermissible acts. Buying rat poison in order to kill one's wife is impermissible because it risks killing her with it. Call this the Risk-Based View.

There are different understandings of risk. On some views, risk is a species of bad effects. On others, it is the (increased) probability of bad effects occurring. And on some views, risk is the probability of bad effects occurring multiplied by their badness. Another question concerns the badness of being subject to risk. Some people think that imposing risk is itself a harm while others think that it is an unfair burden (see Hayenhjelm and Wolf 2011: 29-32). No matter which of these, if any, is the correct understanding of risk and its badness, the truth of the Risk-Based View turns on whether any such understanding accounts for the impermissibility of inchoate acts.

The Risk-Based View can explain *Poison-Poison IV*. In all these cases, the risk of killing one's wife is the same. The conditional probability of her getting killed by rat poison, given the man's act of buying the poison, is stipulated as constant regardless of whether he buys it with the intent of killing her or with a mere foresight of doing it, and it is clearly significantly higher than the prior probability of her being killed by the poison. Moreover, because in real life it is more

likely that the one who buys the poison intending to kill his wife in fact kills her with it than the one who merely foresees that happening, the Risk-Based View explains the derivative relevance of the former person's intention to kill his wife by the fact that having that intention increases the probability of murder. This partly explains the appeal of the Intentions View. However, there are cases where the Risk-Based view fails.

Three types of cases feature prominently in the literature (Yaffe 2010: 28-9, Hanser 2014: 140-1, Tadros 2014c: 14). The first are cases where one has a false belief, justified or unjustified, that he will cause harm but in fact he cannot cause it. For instance, one might try to kill his foe sticking pins in a voodoo doll which represents him. Or one might try to kill his foe shooting at his chest, not knowing that he wears a bulletproof vest. Some people would be reluctant to say that, morally, the former person attempts a murder although most would not deny that the latter does.⁶³ But this is not due to the difference in risk. The intuition remains if we assume that neither attempt increases the probability of murder. There must be something else in virtue of which the latter attempt is impermissible.

The second are cases where one is preempted to cause harm by an internal hindrance or a third party intervention. For instance, just before one pulls a trigger in order to shoot and kill his enemy he gets a heart attack or the police stops him. Either way, he attempts murder but the risk of killing the enemy remains the same as if he had not tried to pull the trigger (assuming that the heart attack or the police intervention were causally inevitable once he tried to do it).

⁶³ The reason for it might be that the former person's belief is unjustified and that there is no belief-relative sense of impermissibility but only justified belief- and fact-relative sense. It is also clear that that person is open to blame but much less clear whether that suffices for criminal culpability.

The third are overdetermination cases. The man who buys rat poison in order to kill his wife does not increase the probability of him killing her with it if, were he not to buy the poison, he would certainly kill her with a different poison he already had prepared at home. Nonetheless, his act of buying the poison with the intent of poisoning his wife remains impermissible.

The Risk-Based View is at best of limited scope, accounting only for the impermissibility of some inchoate wrongs on some occasions. But even that might turn out to be a false if there is a deeper explanation of their impermissibility. The Risk-Based View looks appealing because its explanation for the impermissibility of inchoate wrongs is derivative from their relation with the independent wrongs. As Hanser (2014: 143) rightly notes, the desideratum of a plausible account of inchoate wrongs is that the *pro tanto* moral objection to them is the *pro tanto* objection to the independent wrongs they lead to. What makes inchoate acts impermissible should be the fact that the acts they lead to are impermissible.

How does the wrongness of independent acts, as it were, transfer onto the inchoate acts? I shall now discuss an answer that appears between the lines of Scanlon's (2008: 41-6) account of inchoate wrongdoing. This account is often misunderstood although it seems to me to be correct. My aim is therefore twofold: to offer an interpretation and a defense of it.

3. Scanlon on Inchoate Wrongdoing

3.1 The Facilitation View

Scanlon gives two important but seemingly incompatible explanations. He (2008: 41-2) starts by saying that buying the poison with the intent of killing one's wife is impermissible for the same

reason buying the poison with the intent of giving it to one's neighbor who would use it to kill his wife is impermissible. In both cases, by buying the poison, one knowingly facilitates murder. Call this *the Facilitation View*.

What Scanlon says just after he introduces this view appears as a further defense of it. He (2008: 42-3) claims that buying the poison with the intent of killing one's wife is impermissible because that intent changes the probable consequences of buying the poison in the sense that it predictively leads to murder. This might get one to think that Scanlon claims that the Facilitation View is true because the Risk-Based View is true (Kolodny 2011b: 61). Alternatively, one might think that Scanlon grounds the former on what might be called the Prediction View, according to which buying the poison is impermissible because the man can predict that he will use it to kill his wife, whether or not buying the poison actually increases the probability of him killing his wife (Tadros 2014c: 3, 14).

The problem with those interpretations is that they seem incompatible with other claims that Scanlon (2008: 43) makes. He introduces the following objection to the Facilitation View. If what explains the impermissibility of inchoate acts is that they facilitate independently wrong acts, then any act that causally contributes to an independently wrong act, no matter how remote or innocuous, is impermissible. For instance, if it is causally necessary for a man to wake up very early in the morning in order to go to buy the poison to kill his wife, then waking up very early in the morning is itself impermissible (*Wake Up*). But that seems absurd.

Scanlon (2008: 43-4) shares the intuition and argues that the reason why waking up early in *Wake Up* is permissible is that the man could abandon his murderous plan. He could wake up

but not intend to kill his wife because it is in his full control to drop that intention. Call this the Control View.

Unfortunately, the Prediction View and the Control View are incompatible. Not only their explanations but their implications differ. The problem arises because it is often the case that one can predict that his inchoate act would facilitate an independently impermissible act and yet he could perform the former without intending the latter (Kolodny 2011b: 62, Tadros 2014c: 13). If the man could wake up but not intend to kill his wife, then he could buy the poison but not intend to kill her with it. Therefore, on the Control View, buying the poison is also permissible, and that seems implausible. On the other hand, as we saw, the Prediction View rightly implies that buying the poison is impermissible but it incorrectly censors acts like waking up early.

At this point, one might choose sides, denying that it is impermissible to buy the poison in *Poison* or that it is permissible to wake up early in *Wake Up*. The Prediction View might seem appealing because, like the Risk-Based View, it shows that one's intentions have only derivative relevance. It also gives the correct verdicts in *Poison II* and *IV*, in which the man merely predicts that buying rat poison will lead him to kill his wife. However, the Prediction View fails whenever one is in control of changing his intention but foresees that he will not do it and yet he should. In *Poison V*, a man has a duty to get rid of the rats because his daughter is afraid of them. Buying the poison is the only means to do it but he buys it predicting that he will use it to kill his wife.

The Prediction View implies that buying the poison is impermissible but it is not; in fact, it is obligatory (Tadros 2014c: 19-20).⁶⁴

On the other hand, the Control View is appealing because it avoids the absurd implication in *Wake Up*. It might even have resources to justify buying the poison in *Poison*, distinguishing between different act types on the basis of their description. What the man does in *Poison* can be described both as ‘buying the poison’ and ‘buying the poison in order to kill his wife’. Since he can perform the former but not with intention to kill his wife, it is permissible for him to ‘buy the poison’. But he cannot ‘buy the poison in order to kill his wife’ without the intention to kill his wife since that intention is necessary for that act type. If it were done without that intention, it would be a different act type. Therefore, on this version of the Control View, buying the poison with intention to kill one’s wife is impermissible but buying it, *simpliciter*, is not.

This revision might be also resisted. For it to be permissible to buy the poison with the intent of killing one’s wife, we only need to stipulate that this is necessary for preventing even greater harm, say, for saving five people who would be murdered by one’s wife. What the man could do might then also be described as ‘buying the poison in order to kill his wife in order to save five people’. He will of course use the poison only in order to kill his wife, though he might foresee the side effect of saving the five. But it is implausible that just because he preempts a

⁶⁴ Tadros (2014c: 22-3) also argues that the Prediction View has counterintuitive implications when one predicts that his inchoate act will lead to a good outcome but he initially plans to do it as a means to an independently wrong act. In one of his cases, a man lies to his wife that he is going for a business trip when he is in fact planning to go for a sex holiday while predicting with high probability that he will change his mind and organize a surprise party for his wife. I do not think that the man can plan to go for a sex holiday if he predicts, with high probability, that he will not do it; he can at best hope for it. For that reason, I do not find lying here impermissible. Similarly, I do not think that it would be impermissible to buy the poison, hoping to kill one’s wife with it, but predicting, with certainty, that one will use it to kill the rats.

murder, he does not commit one himself. Had his wife not tried to kill the five, he would have still tried to kill her.

This seems to leave us with no foundation for the Facilitation View. We can neither base it on the Prediction View nor on the Control View. That leads some people (e.g. Tadros 2014c: 24) to reject the Facilitation View. However, that might be too quick. I shall now argue that the Facilitation View is in fact more basic than the others.

3.2 Rescuing the Facilitation View

The Facilitation View satisfies the above desideratum that the reasons against an inchoate act are (transferred from) the reasons against the independently wrong end act. If it is impermissible to kill one's wife by poisoning because she is an innocent, non-consenting, non-threatening person, for whom further life is a good, then buying the poison intending to kill her is impermissible for the same reason. Buying the poison is a causally necessary (or contributory) means to killing her. In short, it facilitates her murder (cf. Kolodny 2011b: 60-1).

On the interpretation offered above, the Facilitation View is in need of further support. It is unclear whether that is Scanlon's view. It seems to me that the issue gets complicated for two reasons. First, the question of permissibility, that is, the question whether it is permissible to act in some way, can be asked from the perspective of a deliberating agent but also independently of anyone's particular perspective (impersonally). Second, since an action can be described in many ways, the question of permissibility can be about different act types.

For instance, we can recognize four questions regarding *Poison*: (1) “Is it permissible to buy the poison?”, (2) “Is it permissible to buy the poison in order to kill one’s wife with it?”, (3) “Is it permissible for me to buy the poison? (asked by the man deliberating whether to buy it)?”, (4) “Is it permissible for me to buy the poison in order to kill my wife with it?”.

Answers to (1) and (3) are “Yes”, and to (2) and (4) are “No”. The former are explained by the Control View: any person, presumably including the man himself, can buy the poison but not in order to kill one’s wife. As long as that is a possibility, buying the poison, *simpliciter*, is permissible.

(2) is explained by the Facilitation View: anyone who, under those conditions, buys the poison in order to kill his wife facilitates her murder. There is nothing more to say to explain the impermissibility of that inchoate act. The reason against buying the poison in order to kill one’s wife is the reason against killing one’s wife.

Finally, (4) is explained by the Prediction View: when deliberating whether to buy the poison, the man can predict that, if he buys it to kill his wife, he would probably kill her, and that is sufficient to make him realize that he should abandon that intent or abandon buying the poison altogether. However, that prediction is of no use when asking the question of permissibility from the impersonal perspective. Importantly, the prediction itself is only a derivative explanation of the impermissibility of buying the poison with that intention; the basic explanation is still given by the Facilitation View.

If this interpretation is correct, no conflict between these views exist. They serve different purposes. The Facilitation View has the purpose to explain why an inchoate act is impermissible as such. The Prediction View explains it for a deliberating agent; its purpose is action-guiding.

Finally, the Control View's purpose is to show that even if performing an inchoate act with intent to harm is impermissible, performing it, *simpliciter*, need not be.

A reason to think that Scanlon appeals to these views in this way is his note on a case of inchoate wrongdoing we find in Mill (1998). Mill (1998: 645, n. 2) cites his opponent saying: "Surely the rightness or wrongness of saving a man from drowning does depend very much upon the motive with which it is done. Suppose that a tyrant, when his enemy jumped into the sea to escape from him, saved him from drowning simply in order that he might inflict upon him more exquisite tortures, would it tend to clearness to speak of that action as a 'morally right action?'". Mill responds: "I submit that he who saves another from drowning in order to kill him by torture afterwards does not differ only in motive from him who does the same thing from duty or benevolence; the act itself is different. The rescue of the man is, in the case supposed, only the necessary first step of an act far more atrocious than leaving him to drown would have been."

Scanlon (2008: 219, n. 3) claims that the tyrant acts impermissibly in saving the man in order to torture him later because he is going to torture him later. The reason against torturing people, whatever it may be, is also the reason against the means to facilitate it. Moreover, since the tyrant intends to torture the man, he can predict that he will torture him and that he will then counter the reason against torturing him, which should make him realize that he should abandon that plan. But the reason against torturing, Scanlon also claims, it is not the reason for the tyrant to abandon saving the man. Since he could abandon the intent to torture him, what he should do is save the man but not in order to torture him later.⁶⁵

⁶⁵ Notice that this does not assume possibilism about moral obligations, namely the view that, because one could act in some way, one ought to act in that way as long as that would be better than not acting at all even if one will not in fact act in that way (Wedgwood 2009c, Ross 2012). The tyrant will in fact save the man just to torture him later. He could save him but not to torture him. But it would not be better if he saved the man and tortured him latter than if he let him drown.

I shall now turn to some objections to the account of inchoate wrongdoing just presented.

3.3 Objections and Replies

Let us first turn to hard cases. Recall *Poison V*, where the man is obliged to get rid of the rats but he buys the poison predicting that he will kill his wife with it. The Facilitation View states that, since the man has decisive reason not to kill his wife, he has the same decisive reason not to buy the poison predicting that he will kill her. But that does not mean that it is impermissible for him to buy the poison, *simpliciter*. He should buy it to do his duty to kill rats. (If he has an irresistible urge to kill his wife, he might be obliged not to buy it at all, but this is no such case.)

A more interesting analysis is that of *Wake Up*. The Facilitation View does not state that it is impermissible to wake up early but that it is impermissible to do it in order to kill one's wife. But even that may sound strange. It may be objected that the act of waking up early in order to kill one's wife is so causally remote from killing her that it cannot be impermissible. But causal remoteness is, I believe, in itself morally irrelevant. Suppose I flip a switch that foreseeably leads to a million events, each causing the other, resulting in a nuclear catastrophe. The act of flipping the switch is no doubt impermissible. The same is true of temporal remoteness. Suppose I dig a land mine in a distant field, being aware that someone might step on it. Several decades later, a reaper steps on it and gets killed. Surely, I acted impermissibly in digging the land mine.

The intuition that it is permissible to wake up early in order to kill one's wife might be a result of the influence of the criminal law distinction between mere preparations and attempts. We would not criminalize or punish the man for waking up with that intention in the same way

we would not criminalize or punish a robber just for buying a balaclava, even if he buys it with intent of using it to rob a bank. More is needed for such acts to count as criminal attempts. The distinction makes sense for practical reasons. As Tadros (2014c: 28) notes, criminal law tries to avoid punishing the innocent and criminalizing mere preparations increases the risk of doing it. Those concerns, however, do not justify drawing a moral distinction on the same lines.

The last objection I turn to comes from Scanlon (2008: 43) himself, though he does not respond to it (at least not directly). Scanlon notices that most people would judge buying poison in order to kill one's own wife worse than buying it in order to give it to the neighbor who would kill his wife with it. However, on the Facilitation View, both acts are impermissible because they facilitate murders. It might then be thought that the Facilitation View cannot explain the intuitive difference between them.

However, this is a problem for any such view. The Intentions View cannot account for the difference because the man's intention to kill his wife is not worse than his intention to assist his neighbor in killing his wife. We can assume that the risk of murder is the same in either case but that does not affect the intuition. The Risk-Based View is therefore also incapable to account for the difference.

The most plausible explanation why we believe that it is, other things being equal, worse to do harm oneself than to assist in doing harm is because we believe that there are agent-relative reasons against doing harm and, more generally, agent-relative duties against it. It is worse to kill my wife *myself* than to help others to do it. This belief might be mistaken because agent-relative negative duties might not exist (see Parfit 2015a: Ch. 18). But that does not matter since many people seem to think that they do and this probably gives rise to their intuition.

Conclusion

I argued that intentions are non-derivatively irrelevant to the impermissibility of inchoate acts. I showed that there are inchoate wrongs done without intention to do the independent wrong and that even those done with that intention are not impermissible in virtue of it since they remain so if done without it. I also argued that the impermissibility of inchoate acts cannot be explained by the fact that they risk the completion of independently wrong acts, as some do not, but by the fact that they facilitate the latter and that, thus, the same reasons that count against the permissibility of the latter count against their permissibility too.

PART TWO

Defenses of the Irrelevance Thesis

5. Intentions and Permissibility: A Confusion of Moral Categories?

Introduction

A common objection to the view that an agent's intentions are non-derivatively relevant to the moral permissibility of actions is that it confuses permissibility with other moral categories: moral character assessment or blameworthiness. The objection states that a failure to distinguish what one is permitted to do from what kind of a person one is, or from what one can be held blameworthy for, leads one to believe that intentions are relevant to permissibility when, in fact, they are only relevant to blameworthiness or to character assessment (Mill 1998: 65, n. 2, Moore 2005: 95-8, Bennett 1981: 97-8, Rachels 1986: 93-5, Thomson 1991: 295-6, 1999: 517, Norcross 1999: 121-2, McCarthy 2002, Enoch 2007: 81, Scanlon 2008: 20-36, 52-6).

Call this objection *the Objection from Confusion* (OFC).⁶⁶ The OFC is not the most important objection to the relevance of intentions to permissibility, but it is by far the most often cited one. The aim of this chapter is to show that the OFC is mistaken. I shall defend two claims: first, that a confusion of moral categories is not the source of the view that intentions are relevant to permissibility and, second, that, in conjunction with some other premises, a confusion does not undermine that view. The OFC neither serves as an error theory, explaining why intentions have seemed to many people to be relevant to permissibility, nor as an argument proper, questioning that belief.

I focus on three strategies used to defend the OFC. I first discuss the inductive strategy, according to which the paradigm cases where intentions supposedly matter to permissibility upon closer reflection reveal that they only matter to blameworthiness or to character assessment and it is safe to generalize from these. I then turn to two notable deductive strategies: the first, defended by Thomson (1991), according to which the position that intentions are relevant to permissibility rests on a false assumption that impermissibility depends on blameworthiness, and the second, defended by Scanlon (2008), according to which the problem arises due to a failure to distinguish deliberative from critical moral assessments.

⁶⁶ As defined here, the OFC presupposes the disjunction that bad intentions, such as the intention to harm, matter to character assessment or that they matter to the blameworthiness of actions. Whether these are just different sides of the same coin is a matter of dispute. Bennett (1981: 99, 1995: 221-4) argues that, although intending harm as an end is often sufficient to display a bad character and a blameworthy action, intending harm as a means is sufficient for neither. Scanlon (2008: 21) agrees about the former, although adding that intending harm as a means shows a defect in one's deliberation about the considerations at stake, and, thus, might sometimes be relevant to blameworthiness (2008: 26-8). Which of these views, if any, is correct will turn on the best theories of character and blameworthiness and on the best explanation of what makes an intention bad. Since these are complex questions which need not be addressed here, I formulated the OFC to cover all possibilities.

1. The Inductive Strategy

The debate about the relevance of intentions to permissibility often concentrates around some paradigm cases from the ethics of war, palliative care, trolleyology, etc.⁶⁷ Although they do not spell it out in such a form, one gets the impression that the critics of the relevance of intentions to permissibility (henceforth: the critics) think that they have a case if, upon closer reflection, these paradigm cases reveal intentions to be relevant to blameworthiness or character assessment but not to permissibility. The argument lurking behind seems to be:

- (1) Certain paradigm cases where intentions are supposedly relevant to permissibility, upon closer reflection, reveal intentions to be only relevant to blameworthiness or to character assessment.
- (2) If intentions are not relevant to permissibility in these cases, they are never relevant to permissibility but only to blameworthiness or to character assessment.
- (3) Therefore, intentions are never relevant to permissibility but only to blameworthiness or to character assessment.

The soundness of the OFC stated in this form depends on the success of generalization from the paradigm to all, or nearly all, cases. It could be said that the argument is fallacious because of the inductive leap. But this is not the best way to reject it. If the paradigm cases do reveal intentions

⁶⁷ Historically, most of these cases were used to support the Doctrine of Double Effect (DDE), according to which it is permissible to bring about harm in pursuit of a good provided that there is (intention-independently) a sufficient reason to pursue the good and that the harm is not intended but merely foreseen (see Boyle 1980, Cavanaugh 2006). The DDE is not the only principle specifying how intentions might affect permissibility but it is the most famous one. The inductive strategy is often directed against the DDE but its conclusion is universal. If the OFC succeeds against the DDE, I shall take it to succeed against any such view.

as relevant only to blameworthiness or to character assessment then, even if the conclusion does not follow straightforwardly, there is a compelling reason to subscribe to it, absent reasons to the contrary. If the argument is sound, it explains the source of the delusion about the relevance of intentions to permissibility and, depending on the justification of the first premise, it provides a reason to accept the contrary view.

The problems with it are twofold. The intuitions about the relevance of intentions to permissibility in the paradigm cases are heavily divided, which halts the appeal to the OFC as an error theory. Moreover, a reasonable theoretical disagreement about the relevance of intentions undermines its force as an argument proper.

To illustrate these claims, consider a contrasting pair from the ethics of war.⁶⁸ In the first case, a bomber engaged in a just war, following the orders of his superior, bombs the enemy's munitions factory intending to end the war though he foresees, but does not intend, the deaths of non-combatants living in the vicinity. In the second case, the circumstances are the same except that a bomber, who acts identically and brings the same effects as the previous one, bombs the factory secretly intending only to kill the non-combatants, something he would not do were it not for the opportunity.⁶⁹

Call the first person the tactical bomber and the second the punitive bomber (cf. Bennett 1995: 215, Cavanaugh 1999: 182-3). Suppose that there is, intention-independently, a sufficient

⁶⁸ Whatever I say about this example applies to the other structurally identical pairs which I do not discuss for the reasons of space, e.g. terminal sedation in order to alleviate the patient's pain/in order to kill him (see, e.g., Thomson 1999), redirecting the trolley in order to save the five/in order to kill the one (see, e.g., Kamm 2007: 131-2), etc. The opinions of some of the authors I refer to later on are actually about these cases.

⁶⁹ The secrecy of intending to kill precludes the predictive significance of that intention and equalizes the effects in the two cases as the victims and the general public might be harmed or offended more if they were aware that they have been targeted deliberately than as a side effect (cf. Kamm 2014: 482-3). Similarly, that the bomber would not act were it not for the opportunity is a detail that precludes the counterfactual relevance of his intention.

reason to bomb the factory, which both bombers are well aware of. Independently of intentions, both bombers would act permissibly. However, as the punitive bomber's intention is bad,⁷⁰ those who believe that bad intentions bear on the permissibility of actions (henceforth: the defenders) think that he acts impermissibly. I should emphasize that they typically claim only that it is impermissible to bomb the factory intending to kill the non-combatants but not that it is impermissible to bomb it, *simpliciter*. Most such views would censor bombing the factory if the punitive bomber cannot act with a different, good, intention and/or there is no better motivated bomber to take over the raid (Tadros 2011b: 156-7, Wedgwood 2011a: 388-91, 2011b: 468-9, Liao 2012: 716).^{71, 72}

The critics disagree, claiming that, holding fixed the consequences and the causal path of the raid, its permissibility cannot differ because of the difference in the bombers's intentions. If it is true that the tactical bomber acts permissibly, it must be true that the punitive bomber does too (Bennett 1995: 218, Thomson 1991: 297-8, 1999: 515-6, Kamm 2004: 658-62, 2008: 168-9,

⁷⁰ It is bad in the minimal sense: its propositional content is an event we all can agree is bad. It might also be bad if it sprang from a bad motive, e.g. hatred or revenge, as such intentions usually do, but we should be careful to keep the things equal and not to bring motives into the picture (the debate about their relevance to permissibility, for the most part, follows this one).

⁷¹ Wedgwood's view is disjunctivist. He (2009b, 2011a: 390-1) would claim that, if the punitive bomber dropped the bombs intending to kill the non-combatants, it is true either that he acted impermissibly or that he ought not to be such that, if he dropped the bombs, he would do it intending to kill the non-combatants. However, if he could not but behave as he did (he could not get rid of the hatred towards the non-combatants in a short timespan), by elimination (and Ought Implies Can), he acted impermissibly.

⁷² FitzPatrick (2003, 2012) believes that permissibility is primarily, if not exclusively, about acting, *simpliciter*. He comes to this conclusion holding that intentions are relevant to permissibility but that actual intentions cannot make a difference, except in rare cases where the one who would act with a bad intention does not let others act identically with good intentions (Fitzpatrick 2012: 112-8). As an alternative, Fitzpatrick suggests that acting, *simpliciter*, is impermissible if there is a justification of it such that the one who would act because of that justification would act with a bad intention. Even if true, this gives up on the idea that intentions are relevant to permissibility. It says that permissible actions are typically not done by malicious agents but not that intentions feature in their justification. Moreover, permissibility judgments of acting, *simpliciter*, are definable in terms of acting with an intention such that the former is impermissible if and only if every possible instance of the latter is impermissible (Wedgwood 2011a: 391). This is true even if intentions do not matter to permissibility. There is then no reason not to talk about the permissibility of acting with an intention besides the permissibility of acting, *simpliciter*.

Scanlon 2008: 20, 31).⁷³ This is followed by the error theory that the defenders confuse the moral categories. The punitive bomber's intention, it is said, does not render his action impermissible but does render it blameworthy or shows that he is a bad person.

However, the strategy misses the target. The defenders do not deny that intentions are relevant to blameworthiness or to character assessment. They only deny that intentions are not relevant to permissibility. Some of them report the intuition that intentions make a difference to permissibility in the paradigm cases (Liao 2012: 716, Sverdlik 2011: 15-6). That intuition does not hinge on the category mistake but amounts to a different perception of the importance of bad intentions as considerations counting against the permissibility of actions.

The defenders might develop different justifications of that perception. They might claim that bad intentions outweigh all the reasons counting in favor of actions or that they function as exclusionary reasons, excluding or silencing the others. These people might further appeal to the ways in which bad intentions affect one's agency for the worse (see Nagel 1986: 181, Duff 1990: 113, Moore 1997: 409-10, Wedgwood 2011a: 392-3, 397), or in which acting with them wrongs the victims by exhibiting disrespect (see Quinn 1989: 348-51, Cavanaugh 2006: 147-58, Tadros 2011b: 130), or in which such intentions give rise to impersonal badness (cf. McMahan 2009: 357-8), holding that permissibility is grounded in one of those features or some combination of them.

⁷³ Some of those authors leave open the possibility that the bomber who intends to harm non-combatants as a means to a bad end *and* harms them as a means to it, e.g. kills them in order to demoralize the enemy (commonly called the terror bomber), acts impermissibly. But again, this is not due to his intention but other considerations, e.g. the death of non-combatants not being a proper military advantage (Scanlon 2008: 28-9). McMahan (2009a: 367-9) ascribes to Scanlon the view that terror bombing is impermissible due to the morally objectionable causal path the bomber pursues his goal through. That might read too much into what Scanlon says as the only additional claim he (2008: 28-32) makes is that, in real life, because the terror bomber intends to kill the non-combatants, he is more likely to produce graver consequences than the tactical bomber.

These views are united by the belief that bad intentions are as morally important as to oblige one to abandon them and act with good intentions or, if that is not an option, to abandon acting at all. The belief cannot be countered in a non-question-begging way by appealing to the opposite intuition, in tandem with the error theory about the confusion of moral categories. What gives rise to that opposite intuition is not the detection of the supposed confusion but the implicit assumption that bad intentions do not or, perhaps, cannot count against an action's permissibility. This must be what motivates the first premise for otherwise there would be no opposition to the claim that the punitive bomber acts permissibly once everyone agrees that he is a bad person or merits blame. The OFC therefore fails as an error theory.

The critics might reply that one person's modus ponens is another's modus tollens. The fact that the defenders conclude that the punitive bomber acts impermissibly suffices to discredit their intuitions as peripheral and theory-driven. No theory can justify the result which many find jarringly counterintuitive.

To be taken seriously, the remark needs to be backed up by a study of the prevalence of those intuitions in a wider population. Such a study would prove nothing about the relevance of intentions to permissibility but, if it would demonstrate an intuitive stalemate, that would, at the minimum, constrain too quick dismissals of the defenders' intuition.⁷⁴

⁷⁴ In an online survey I conducted, majority of subjects reported the intuition that the punitive bomber's action is impermissible (72.00%, $\chi^2 [1, N=50] = 5.086$, $p < 0.001$, two-tailed). I repeated the survey, this time telling another group of subjects to pay attention not to confuse character evaluation and blameworthiness with permissibility. The results differed but not excessively (54.00%, $\chi^2 [1, N=50] = 0.16$, $p < 0.001$, two-tailed). Since I used convenience sampling, the results are not conclusive but they are at least indicative of the non-peripheral nature of the belief that intentions are relevant to permissibility. At best, the confusion of moral categories accounts for the intuition of some but not all people. (Future surveys should also examine whether, and to what extent, intuitions about such cases are influenced by the existing rules of international humanitarian law.)

A similar result is reached by qualitative research. A recent study of Douglas, Kerridge, and Ankeny (2013) brings in-depth interviews with 18 Australian palliative care specialists, majority of who are not religious. It reports a wide agreement that administering lethal drugs, when the conditions of terminal illness conclusively favor doing it, is permissible only with the intention to relieve a patient's anguish.

However, even if the intuition would be shown to be peripheral, the justifications of the relevance of intentions to permissibility would deserve to be addressed on their own. There is a reasonable theoretical disagreement between the defenders and the critics; both are engaged with the question seriously and advance convincing arguments. There is then no reason to assume that the defender's view is mistaken just because it entails what, pre-theoretically, many would find counterintuitive. The defenders might bite the bullet but the bullet might not be as detrimental as it seems to some people.

The inductive strategy also fails as an argument proper. As I have just said, the belief that bad intentions are as important as to render otherwise permissible actions impermissible is not without foundations. Most critics do not engage with the justifications but rest the case on what they think is an obvious inference from the paradigm cases, namely that there is no difference in permissibility between actions differing only in the intentions of those performing them.⁷⁵ Kamm (2008: 168, cf. 2014: 380, Rachels 1986: 92), for instance, writes:

When an agent has a wrong intention, he does not do an act merely for the sake of the properties that make the act permissible or required—he acts for a bad reason—but that does not make the act itself lose any of the properties that make it permissible or required. This is because, as Thomas Scanlon has argued, in most cases an agent can decide whether an act is permissible by considering properties and effects the act would have independently of considering with what intentions she would do the act.

⁷⁵ An exception is Kagan 1989: 177-82. However, since Kagan does not appeal to the OFC in any form discussed here, I leave his criticism aside.

But this clearly begs the question against the defenders. Holding fixed the intention-independent properties of a permissible action, its right-making properties remain, but the real question is not whether those properties disappear if one acts with a bad intention but whether they are the only properties that can affect the permissibility of actions. In other words, we need to ask whether the property of acting with a bad intention is among the wrong-making properties that override the other, right-making, ones. The defenders believe that it is; indeed, they think that bad intentions are considerations that render otherwise permissible actions impermissible.

Consider two other claims the critics often make. As Kamm suggests in the above quote, appealing to earlier observations of Thomson (1991: 293) and Scanlon (2008: 19-20, 30), moral decision making is exhausted by considering the properties and consequences of an action. To examine further the content of one's own intention in acting is to make one thought too many. As Thomson (1991: 293, cf. 1999: 514-5) notably puts it, it would be "queer" to consider intentions in decision making. "Can anyone really think", she asks, "that the pilot should decide whether he may drop the bombs by looking inward for the intention with which he would be dropping them if he dropped them?"

The critics find this question rhetorical but it need not be. The defenders could reply with a counter-question: "Why would the performance be queer?". It cannot be because intentions are irrelevant to permissibility, as the perception of queerness would then just reiterate the intuitions of the critics. To the ears of the defenders, there might be nothing queer in the performance (cf. Kolodny 2011a: 119). To avoid begging the question, the critics need a theoretical justification of the queerness. I am not saying that no justification is promising but that such observations, taken alone, are futile in justifying the first premise.

Similarly, the critics often claim that, if bad intentions were relevant to permissibility, an agent aware of the badness of his intention and motivated to act morally would either need to abandon his action or to change his intention and then proceed. Since changing intentions is not an on-off process but an effortful and tedious engagement in reassessing one's reasons for action, the requirement to act but not with a bad intention would, in most, if not all cases, be practically equivalent to the requirement not to act, *simpliciter*. But when the stakes are high, a failure to act has grave consequences. In the punitive bomber example, the war and suffering would prolong if the raid is aborted. Sacrifices of that kind just for the sake of avoiding acting with a bad intention make no sense, or so the critics argue (Thomson 1991: 297-8, 1999: 514-6, Bennett 1995: 218, Enoch 2007: 81, Scanlon 2008: 20, 31).

The only thing I can make of this line of reasoning is a restatement of the critics's thesis. Consider an analogy with the constraint on harming as a means to a good end. Many people believe that that constraint has absolute or near-absolute importance. They would argue that a bomber would act impermissibly if, unable to kill non-combatants as a side effect, he kills them as a means of destroying the munitions factory, which then causes the war to end. Now no one, apart from act-consequentialists, would seriously think that that view is questioned by the observation that the war and suffering would continue were the bomber not to kill the non-combatants as a means, unless, perhaps, the consequences counted in thousands of victims. Consequences matter morally but, for non-consequentialists, they are not all that matters. The constraint on harming as a means might be philosophers's fiction but we need a story to tell why. By the same token, we need a story to tell why the constraint on intending harm is a fiction, if we intuit it is. Again, this puts the ball squarely in the court of the critics.

The critics could reply that they did enough theoretical work to establish the irrelevance of intentions to permissibility. Several such arguments are offered and they stand on their own, so there is no need to use them to back up the first premise.⁷⁶ Therefore, if the arguments against the relevance of intentions are conclusive (which I disagree with but cannot argue for here for the reasons of space), the inductive strategy, in itself, is powerless.

The rationale for the first premise can instead be provided by a deductive variation of the OFC, such as Thomson's (1999) or Scanlon's (2008). Those variations tend to be error theories of their own too. I discuss them in turn.

2. The Deductive Strategy: Thomson

Thomson (1991: 295) argues that the reason why some people believe that intentions are relevant to permissibility is that they are confused about the entailment relation between blameworthiness and impermissibility. They think that an action cannot be blameworthy and not be impermissible. But that is certainly possible, Thomson claims, so an action's blameworthiness is irrelevant to its impermissibility and so is its being done with a bad intention. The argument is:

- (1) If intentions are relevant to permissibility, the blameworthiness of an action done with a bad intention is relevant to its impermissibility.
- (2) The blameworthiness of an action is irrelevant to its permissibility.

⁷⁶ According to one of them, intentions cannot be relevant to permissibility because, if that were the case, some actions which should be encouraged would not be permissible because of being done with bad intentions and it cannot be permissible to encourage impermissible actions (see Thomson 1999: 516). It has been also argued that it is impossible to act for some reason rather than another, so it cannot be morally required, permissible, or impermissible to do so (see Scanlon 2008: 57-62). See n. 19 for a brief critique of the latter.

(3) Therefore, intentions are not relevant to permissibility (but are relevant to blameworthiness).

Some defenders do claim that, although impermissibility need not entail blameworthiness, blameworthiness does entail impermissibility. On their view, an action is impermissible when it is blameworthy or, more precisely, an action which is blameworthy in virtue of being done with a bad intention must be impermissible under the same description (considered with the intention and not *simpliciter*) (Wedgwood 2011b: 471). However, no such view states that the reason for an action's impermissibility is its blameworthiness.

So while some defenders think that blameworthiness entails impermissibility, they do not think of the entailment as explanatory. But for Thomson's argument to succeed, it is precisely the relation of explanatory dependence (or grounding) that must obtain. Blameworthiness needs to account for impermissibility and not only to correlate with it. I find this the only sensible way to interpret Thomson's talk of the relevance of blameworthiness to permissibility. A more accurate formulation of the argument is, therefore, the following:

- (1) If intentions are relevant to permissibility, an action done with a bad intention is impermissible because it is blameworthy.
- (2) No action is impermissible because it is blameworthy.
- (3) Therefore, intentions are not relevant to permissibility (but are relevant to blameworthiness).

It is now easy to see that the defenders need not commit to the first premise. They usually think that an action is impermissible not because it is blameworthy but because it is done with a

bad intention. They have no reason to stipulate the explanatory chain between impermissibility and intentions going through blameworthiness or the existence of such a chain at all (this would, perhaps, be unattractive because it would take intentions to be, in a way, indirectly relevant to permissibility).⁷⁷ Thomson's OFC therefore fails as an error theory.

It is also beside the point that some defenders think that an action cannot be blameworthy but not be impermissible. For if they are right about it, that is because impermissibility and blameworthiness are both affected by the presence of a bad intention and not because the former category is grounded in the latter. And if they are wrong, that an action done with a bad intention is blameworthy but not impermissible shows only that a bad intention, while being sufficiently important to render an otherwise blameless action blameworthy is not sufficiently important to render an otherwise permissible action impermissible. Which of these options is true, we need not settle here; what matters is that either is immaterial to the soundness of Thomson's OFC as an argument proper.⁷⁸

⁷⁷ Interestingly, Scanlon allows permissibility to depend on intentions via depending on blameworthiness in what he calls the expression cases. In such cases (Scanlon 2008: 39-40), an action presents one as being moved by a certain reason (e.g. calling a sick relative out of concern or out of interest for inheriting his wealth) and, when one's reasons are wrong, they not only render an action blameworthy but count against it being done, albeit rarely decisively. Scanlon (2011a: 460) is explicit that this is not because both categories are independently affected by one's wrong reason but because permissibility is first affected by blameworthiness. However, if intentions are indirectly relevant to permissibility when permissibility depends on blameworthiness, there is not much difference between that kind of relevance and their direct relevance, as usually understood (unless blameworthiness is ultimately due to factors other than intentions, which, at least in the sick relative example, is not the case). This is especially problematic if the relevance of intentions in the expression cases is generalizable to any case where one acts for a wrong reason, and nothing Scanlon says about the issue precludes generalizing.

⁷⁸ Hanser (2005: 455) attacks the second premise, claiming that Thomson can appeal to it only if she proves first that intentions are irrelevant to permissibility. Essentially, then, the argument is question-begging. Hanser may think this because he reasons that, if no action is impermissible because done with a bad intention and if all such actions are blameworthy because of the intentions, no action is impermissible because it is blameworthy. But that is an invalid inference. As we have seen, impermissibility can be independent of blameworthiness even if intentions are relevant to both categories. If the second premise is true, it is not because intentions are irrelevant to permissibility.

3. The Deductive Strategy: Scanlon

Scanlon claims that the defenders fail to distinguish between two kinds of moral assessment or, as he puts it, two usages of moral principles: ‘deliberative’ and ‘critical’ (Scanlon 2008: 22).⁷⁹ Moral principles are employed as guides to deliberation in order to respond to the question of permissibility - ‘May one act in some particular way?’ To answer that question is to specify the considerations counting in favor or against acting; it is to identify normative reasons for or against acting.

On the other hand, employed as standards of criticism, moral principles assess the ways in which one went about deciding to act. We ask why one acted, which considerations one took into account and whether one weighed them in the right way. To use principles in this sense is to identify and assess one’s operative reasons for acting and, by extension, one’s intentions (as, on Scanlon’s view, operative reasons are necessary, but not sufficient, condition for intentions, see Scanlon 1998: 33-4, 2000: 306-7, 2008: 10, cf. Bennett 1995: 194).^{80, 81}

Since an action is rendered impermissible by the considerations that the relevant principle identifies as counting decisively against it, and since the deliberative use of principles is the specification of such considerations, deliberative questions concern judgments of permissibility

⁷⁹ By a moral principle Scanlon (2008: 22) understands a general moral requirement specifying which considerations decisively count in favor or against an action.

⁸⁰ It is easy to conflate the deliberative and the critical uses with similar distinctions. Scanlon himself used to claim that the two uses can be distinguished in virtue of their temporality. The deliberative employment is, Scanlon (2000, 307, cf. Suikkanen 2011: 562) wrote, prospective whereas the critical is retrospective. But it is certainly possible for the deliberative use to be retrospective (Scanlon 2008: 9). We can ask whether one would have been permitted to act as one did. Likewise, the distinction does not correspond to that between the first-person and the third-person evaluative perspectives (Scanlon 2008: 218, n. 9, cf. Kolodny 2011a: 102-3). Both questions can be asked from both standpoints. We can ask ‘May I act?’ as well as ‘May she act?’ Similarly, we can ask ‘What does/did she take as reasons for acting?’ as well as ‘What are/were my reasons for acting?’.

⁸¹ I use ‘taking as/believing to be a reason’ and ‘operative reason’ interchangeably.

(Scanlon 2008: 27). However, the critical use of principles, Scanlon (2008: 52-6) claims, has nothing to do with permissibility. Rather, it is connected to a different moral category he calls ‘meaning’. Meaning is a property of an action, and it concerns its significance for the agent and for those affected by it in the sense that what the agent takes as reasons for acting determines the judgment about the meaning of the action (Scanlon 2008: 52). For example, an altruistic action has a different meaning if the one performing it is motivated by compassion for the needy than if motivated by vainglory.⁸² In the latter case, there is a reason to be disappointed and to blame that person. Blameworthiness is, therefore, a species of meaning (Scanlon 2008: 7).

The argument can be spelled out as follows:

- (1) The permissibility of an action is determined by the existence of reasons counting in favor or against it and not by what one takes these reasons to be. It is the meaning of an action which is determined by the latter.
- (2) If the permissibility of an action is determined by the existence of reasons counting in favor or against it and not by what one takes these reasons to be, then what one takes these reasons to be and, by extension, what one intends are irrelevant to permissibility.
- (3) Therefore, what one takes as reasons for acting and, by extension, what one intends are irrelevant to permissibility.

As said before, the defenders do not deny that intentions are relevant to blameworthiness. They also need not fail to distinguish deliberative from critical questions nor normative from operative

⁸² Operative reasons are often motivating reasons to act. On Scanlon’s (1998: 33-41) view, they are the sole source of moral motivation.

reasons. They might believe that what one takes as a reason for acting is not itself the normative reason for acting but that there are normative reasons not to act (or to act only) with a certain intention, that is, normative reasons to act but not for some reason (or for some reason only). These normative reasons are reasons to act for one or another operative reason but they are not themselves operative reasons; they are only based on them. For normative reasons of that kind to exist, the operative reasons they require abstaining from or exclusively acting for might need to be relevant to permissibility to begin with. They will be relevant only if one of the justifications of the relevance of intentions stated above is true.

To put it differently, Scanlon assumes that the answer to the question ‘May one act?’ must take the ‘Yes’ or ‘No’ form. But the defenders of the relevance of intentions claim that it might sometimes take the ‘Yes for R_1 but no for R_2 ’ form, where R_1, R_2, \dots, R_n are reasons one acts for. For example, they claim that the answer to the question whether the bomber may drop the bombs is not a simple ‘Yes’ or ‘No’ but rather ‘Yes, in order to end the war’ and ‘No, in order to kill the non-combatants’, even if the consequences of the two are predictably the same. Put in terms of reasons, this means that there is a conclusive reason against bombing the factory intending to kill the non-combatants but a sufficient reason in favor of bombing it intending to end the war. It is these normative reasons which, if the defenders are right, would determine the permissibility of bombing. These reasons are not identical with the operative reasons of the two bombers but that does not mean that the latter cannot be relevant to the permissibility of bombing. The normative reason to bomb the factory but not in order to kill the non-combatants might not be able to get off the ground without the intention to kill being relevant to the permissibility of bombing in the first place. Whether that or some other operative reason is relevant to permissibility will turn on

the best justifications of their relevance. But there is nothing on a conceptual level that would preclude that possibility.⁸³

It might be replied that these more filled-out claims specify which considerations do the work and which do not, and that operative reasons are neither. “Yes, you may drop the bombs in order to end the war” means that ending the war is a sufficient reason for bombing and “No, you may not drop the bombs in order to kill the non-combatants” is way of pointing out that killing non-combatants is not a good enough reason for bombing. Either way, it is normative and not operative reasons and intentions that feature as relevant for permissibility.

The problem with this reply is that, at the same time, it prohibits too much and excuses the inexcusable. If killing non-combatants is not a good enough reason for bombing, then even the strategic bomber acts impermissibly. And if ending the war is a sufficient reason, any kind of bombing is permissible. Holding the actions of the strategic and the punitive bomber fixed, this will not do as an explanation of the difference in permissibility between them. Most critics deny that there is a difference to start with but those who share the intuition of the defenders will need to accept the framework of normative reasons to act for some but not another operative reasons, even if they dispute that the operative reasons are what ultimately induces the normative reasons.

The defenders, therefore, need not confuse normative with operative reasons; they might believe that normative reasons for or against acting, *simpliciter*, are not the only game in town. To deny this Scanlon cannot say that deliberative questions ask us to specify normative reasons

⁸³ I said above that the defenders of the relevance of intentions can accept the weighing model or take intentions as exclusionary reasons. It can be objected that, if an action’s permissibility is determined by the normative reason not to act for a wrong reason, there is no logical space for the weighing model. However, although the normative reason not to act for a wrong reason is a second-order reason, that does not mean that it must be exclusionary. There is no reason why second-order reasons of this kind could not weigh against the other, first-order, reasons counting in favor or against an action.

to act and not normative reasons to act for some reason rather than another without begging the question. He provides an independent argument for that conclusion relying on what he thinks is a conceptual connection between the impossibility of choosing one's operative reasons and the impossibility of acting for them and, accordingly, of being morally permitted or forbidden to act for a reason (Scanlon 2008: 57-62).⁸⁴ That argument aside, the appeal to the confusion between normative and operative reasons is unhelpful for settling the issue of the relevance of intentions to permissibility. Scanlon's OFC, then, is neither good as an error theory nor an argument proper.

Conclusion

I have argued that appealing to the confusion of permissibility with blameworthiness or character assessment cannot explain why some people believe that intentions are relevant to permissibility nor can it, accompanied with some additional premises, undermine that belief.

I first argued against the inductive variation of the OFC, showing that it is useless in the face of the intuitive and theoretical disagreements about the moral importance of intentions as considerations counting against an action in the paradigm cases. I then turned to two deductive variations of the OFC. *Pace* Thomson, I argued that the defenders believe that acting with a bad intention is impermissible because they believe that bad intentions are as important morally as to render it impermissible and not because such acting is blameworthy. *Pace* Scanlon, I argued that, although the defenders claim that permissibility is determined by operative reasons, they do not

⁸⁴ Even if we cannot choose what to see as a reason for an action in the same way we cannot choose what to believe is true, we can be responsible for what we believe and intend because we have the ability to believe and to intend otherwise. Any normally functioning agent who can look for and assess reasons favoring or disfavoring an action has that ability. It is that ability, and not, as Scanlon thinks, the ability to believe at will that something is or is not a reason for acting, which is necessary for the possibility of the moral requirement to act for one reason rather than another, and with one intention rather than another.

confuse them with normative reasons; they might rather think that permissibility is determined by normative reasons to act but not for a reason (or to act for a reason only), which would exist just when and because those operative reasons are relevant to permissibility.

The failure of the OFC is not the failure of the view that intentions are irrelevant. The debate continues as there are other arguments in favor of this and the opposite view. However, to make progress in any direction, we should discard the OFC as a non-starter.

6. Intentions, Permissibility, and Choice

Introduction

Scanlon (2008) has argued that the intentions with which one acts, or more specifically, one's reasons for acting, are non-derivatively irrelevant to the moral permissibility of one's actions.⁸⁵ According to one of his arguments in favor of that thesis, it can be permissible to act for one reason rather than another only if one can choose to act for a reason but, since that is impossible, one can be permitted to act but one cannot be permitted to act for a reason. Call this argument *the Argument from Choice* (AFC). It can be spelled out as follows:

- (1) If one cannot choose to act for some reason rather than another, then one cannot be morally required, permitted, or forbidden to act for some reason rather than another.
- (2) One cannot choose to believe that something is the case.
- (3) (From P2) Therefore, one cannot choose to believe that something is a reason for action.
- (4) If one cannot choose to believe that something is a reason for action, then one cannot choose to act for some reason (rather than another).

⁸⁵ On Scanlon's (2008: 11) view, one's reasons for acting, i.e. considerations one takes to be relevant, are necessary though, perhaps, not sufficient for one's intention to act in that way (cf. Bennett 1995: 193-4, Kamm 2007: 95-6). I shall assume this relation between one's reasons for acting and intentions throughout the chapter.

- (5) (From P3 and P4) Therefore, one cannot choose to act for some reason rather than another.
- (6) (From P1 and P5) Therefore, one cannot be morally required, permitted, or forbidden to act for some reason rather than another.

P1 states that an action can be permissible or impermissible only if one can choose whether to pursue it or not (Scanlon 2008: 58). Without such a choice, no moral requirement is applicable. This presumably follows from a version of the Ought Implies Can principle. P2 states that beliefs are not objects of choice. One can choose to act in one way or another but one cannot choose to believe that something is or is not the case, so one cannot choose to believe that something is a reason for acting (Scanlon 2008: 59-60). P2, then, entails P3. These premises presumably follow from the truth of doxastic involuntarism, the thesis that believing at will is impossible. Finally, P4 states that the impossibility of choosing to believe that something is a reason for action entails the impossibility of choosing to act for some reason rather than another (Scanlon 2008: 61). Each premise is controversial and requires elaboration, which I shall provide in the following sections.

The AFC matters because of examples such as the following. Imagine a selfish rescuer who pulls a drowning child out of a pond not because he cares about the child but only for the sake of getting praised. It is intuitive to say that, despite his wrong reason, he acts permissibly. Nevertheless, it might be argued that, while saving the child, *simpliciter*, is permissible, saving the child in order to get praised is not (cf. Tadros 2011b: 157, Wedgwood 2011a: 388-91, 2011b: 468-69). It does not seem so implausible to claim that one is sometimes morally required to act for right reasons or, at least, to act but not for wrong reasons (this need not be true in this case but I take it simply as an illustration). However, if the AFC is sound, no obligation, permission,

or prohibition of that kind can ever exist. Moreover, as the Doctrines of Double and Triple Effect and similar moral principles state that the permissibility of an action is, under certain conditions, determined by one's reasons for acting, they presuppose that it is possible to act for some reason rather than another.⁸⁶ Accordingly, if the AFC is sound, these principles are all false.

The aim of the chapter is to criticize the AFC. I start by arguing that P1 is best understood to follow from Ought Implies Can and that resisting a familiar objection to that principle comes at the cost of internal consistency for Scanlon's view. I proceed to argue that P3 and P4 are false because one is able to believe otherwise, by looking for and evaluating other reasons for action. I conclude with the AFC's absurd implications.

1. Permissibility and the Ought Implies Can

1.1 'Ought' and 'Can'

To see whether and how P1 is derived from the Ought Implies Can, we need to look closer at that principle. It states that one ought to do something only if one can do it. Much depends on how the two modals are understood. For the purposes of this discussion, 'ought' is to be understood as moral 'ought', in contrast with, say, prudential or epistemic. Moral 'ought' is expressed through moral requirements, prohibitions, and permissions. I shall treat 'ought' as interchangeable with

⁸⁶ The Doctrine of Double Effect states that it is permissible to bring about harm in pursuit of a good if and only if the harm is in some sense proportional to the good, it is not brought as a means to the good but as a side effect of it, and it is not intended but merely foreseen (see, e.g., Boyle 1980, Cavanaugh 2006). The Doctrine of Triple Effect states that, under similar conditions, it is impermissible to act intending to bring harm in pursuit of a good but it is permissible to act on condition of that harm occurring (Kamm 2007: Ch. 4). Both doctrines prohibit intending harm, either as an end or as a means, but not intending harmless but morally objectionable things, such as to get praised for saving a drowning child – the selfish rescuer case is not representative of their implications. Nonetheless, for reasons of simplicity, I shall use it as a paradigm case for any view taking intentions as relevant to permissibility.

morally obligatory or required, ‘ought not’ with morally impermissible or forbidden, and morally permissible as the negation of ‘ought not’.

The ‘ought’ in the Ought Implies Can is sometimes taken to mean the objective ‘ought’, that is, as what in fact one ought to do and not what one (justifiably) believes one ought to do (Vranas 2007: 168, Graham 2011: 339-41). Scanlon (2008: 49-52) holds that moral ‘ought’ is subjective, being what one would believe one ought to do, if one assessed the available evidence in a reasonable manner. Some other authors (Gibbard 2005, Parfit 2011a: Ch. 7) argue that these are but different senses of ‘ought’, fitting different purposes. As nothing in this chapter hinges on which, if any, of these senses is the right one, I shall leave that debate aside.

Regarding ‘can’, it is generally agreed that it denotes ability (Haji 2002: 16, Vranas 2007: 169-71, Graham 2011: 341-2, Wedgwood 2013: 75). The ability is not metaphysical possibility. I cannot be morally required to save a drowning child running like a cheetah to pull him out, even though there is a possible world where I am able to run that fast. The ability is therefore a kind of physical possibility. More precisely, it is specific ability, the one dependent on a person, a time, and a circumstance. So although it might be physically possible for me to run like a competitive sprinter, if I train hard and have the genetic predispositions, I am currently unable to do it. I then cannot be required to save the child running like a competitive sprinter. And even if I were able, I could not be required to do it if a mountain stood between us or if I were unconscious. The latter shows that the specific ability at work here includes the (environmental, psychological, or other) opportunity to act (Mele 2003: 447).

Importantly, the ability is not fixed to these particularities to the extent that one is able to do only that which, at the moment of acting, his physical, psychological, or epistemic conditions

allow. One is able to do more than one would actually do. How much more is a question which, even if it has a precise answer, cannot be dealt with here. Certainly, although abilities come in degrees, some have thresholds. For example, an adult, cognitively and physically functional, and socialized person is able, at the bare minimum, to compute one digit numbers, to run faster than 10 mph, to know basic empirical and normative facts, including facts about morality – e.g., that suffering and death are bad, that, barring overriding reasons, promises ought to be kept, and that the reason to help those in need is not self-promotion. It is also true, I believe, that such a person has abilities to do or know the latter even when he does not do or know them. I shall defend that last claim below.

Another important issue is that of the relation between the ‘can’ and the opportunity not to act or, more broadly, between the Ought Implies Can and the Ought Implies Can Refrain. On the view I shall rely on (Zimmerman 1993, 1996: Ch. 3, Haji 1993, 2002: Chs. 3-4, 2012: Chs. 2-3, Wedgwood 2013), the latter follows from the former. For it to be true that one ought to do something, it must be true not only that one can do it but that one can refrain from doing it; it is a two-way ability. Put counterfactually, one is morally required to act or not to act only if one can act otherwise. Without alternative possibilities, one acts neither permissibly nor impermissibly.

1.2 Alternative Possibilities and Thick Act-Types

We are now in the position to derive P1. It seems to me that conditioning the choice to act for a reason as necessary for the requirement or the permission to act for a reason is best interpreted as saying that alternative possibilities are necessary for deontic morality dealing with act-types of

the sort. The following derivation is laid out with respect to ‘morally required’. It should not be difficult to construct parallel ones for ‘morally forbidden’ and ‘morally permissible’. Here it is:

- (1) One is morally required to act only if one can act. (Ought Implies Can)
- (2) One is morally required to act if and only if one is morally required not to refrain from acting.
(Ought is Equivalent to Ought Not to Refrain)
- (3) (From (1) and (2)) Therefore, one is morally required to act only if one can refrain from acting. (Ought Implies Can Refrain)
- (4) If one is morally required to act, *simpliciter*, only if one can refrain from acting, *simpliciter*, then one is morally required to act for some reason only if one can refrain from acting for that reason (i.e. only if one can act for another reason). (Thin to Thick Act-Types Bridging)
- (5) (From (3) and (4)) Therefore, one is required to act for some reason only if one can refrain from acting for that reason (i.e. only if one can act for another reason). (Ought Implies Can Refrain from Acting for a Reason)
- (6) (From (5)) Therefore, one is required to act for some reason only if one can choose not to act for that reason (i.e. choose to act for another reason). (P1)

I have already explained (1) and (3). (2) expresses an equivalence between the moral obligation to do something and the moral impermissibility to avoid or to refrain from doing it. I shall take the validity of steps (1)-(3) for granted, and focus on the remaining premises.

To understand (4) and (5), let us first distinguish act types from act tokens. The former are repeatable events that are or could be performed by different agents at different times, such as

‘killing’, ‘lying’, ‘helping’, etc. Act tokens are unrepeatable instances of act types, performed by a particular agent at a particular time, such as ‘Jones killing Smith at 5 pm’ (see Goldman 1970: Ch. 1). On the view I shall rely on (Zimmerman 1996: 53-7, Wedgwood 2010), moral obligation concerns act types, not act tokens (thus defined).

Act types can be described in more or less detail. For example, ‘killing’ is a broader act-type than ‘homicide’. One way in which act types differ in descriptions is on the basis of whether intentions or one’s reasons feature in their content. Call act types described without intentions or one’s reasons, thin, (e.g. ‘shooting’) and those described with intentions or one’s reasons, thick, (e.g. ‘shooting in order to kill’) (Wedgwood 2011a: 387-91, 2001b: 465-6). Some thick act types are covered by distinct concepts. Lying is making a statement that one believes to be false with the intention to deceive.⁸⁷ Without that intention, lying is indistinguishable from other untruthful practices, such as writing fiction or making pranks. The same goes for blackmail, theft, bribery, forgery, and many other actions (cf. Husak 2009: 58-9). However, a thick act type is often merely constructed for argumentative purposes and is not covered by a distinct concept (e.g. ‘rescuing the child in order to get praised’, ‘turning the trolley in order to save the five’, etc.).

If the AFC is sound, there is no meaningful sense in which permissibility can be tied to thick act types. Nonetheless, the distinction between thick and thin act types makes sense even if it has no relevance for the issue of permissibility.

Return now to premises (4) and (5). The former states that if Ought Implies Can Refrain applies to thin act types, then it applies to thick act types. (5) is just the conditional’s consequent.

⁸⁷ There is a debate about the appropriateness of that definition. It was argued (Sorensen 2007) that the so-called bald faced lies present an exception to the criterion of intending to deceive. Those cases aside, the definition holds and suffices for our needs here.

Note that (4) does not state that the moral obligation regarding thin act types is defined in terms of the moral obligation regarding thick act types but that there is a correlation between them. The correlation looks rather intuitive; if alternative possibilities are necessary for requirements to act or not to act, they are necessary for requirements to act or not to act for a reason.

Finally, since an alternative way to say that one can refrain from doing something, is that one can choose not to do it, (5) leads to (6). As we shall see, the soundness of the AFC crucially depends on how the verb ‘choose’ is understood. I shall argue that Scanlon equivocates between different senses of ‘choose’.

1.3 Objections to P1

Objections to P1 are objections to Ought Implies Can and Ought Implies Can Refrain. I shall focus on what is frequently thought as the most important kind of objections, those coming from Frankfurt-style cases. I shall suggest that they are not decisive though they raise issues of internal consistency for Scanlon’s overall project.

Not much needs to be said in favor of the two principles. They are often taken as axioms of moral theory, and they are rather intuitive. The main reason for it, it seems me, lies in the idea that a requirement which would ask one to do what one is unable to do is remarkably unfair (cf. Coop 2003: 272, Fischer 2006: 204). That explanation is not without controversy and there are others (see Vranas 2007: 198, n. 4) but, given the principles’s intuitive and theoretical appeal, the burden of proof falls on the critics. Let us then turn to Frankfurt-style cases.

Frankfurt (1969, cf. 1988) has argued that the Principle of Alternate Possibilities, which states that one is responsible for what one has done only if one could have acted otherwise, is false. In Frankfurt-style cases, one cannot act otherwise because one's counterfactual opportunity to act is precluded by the intervener and yet he does exactly what the intervener would make him to do. It might then be thought that the same line of reasoning applies against Ought Implies Can Refrain and, consequently, against Ought Implies Can (Fischer 2006: Ch. 10-1).

In one such case (Frankfurt 1969: 835-6), Black, an evil neurosurgeon, implants a chip in Jones' brain, unbeknownst to him. Black can manipulate Jones in such a way that will ensure that Jones kills Smith. Black would intervene only if Jones were about to decide not to kill Smith. As it happens, Black never has to intervene because Jones, for reasons of his own, kills Smith.

Now if it is indeed true that Jones lacked the opportunity not to kill Smith, including the opportunity to kill him for a reason different from his own—which is disputed by the 'flicker of freedom' strategists—then we should, I believe, dig in our heels and say that Jones neither killed Smith permissibly nor impermissibly. We should judge it just as if Smith died in a mudslide (cf. Zimmerman 1996: 89, Vranas 2007: 195-6).

One objection (Widerker 1991) is that Jones's action is blameworthy and an action must be impermissible if it is blameworthy. But, arguably, some blameworthy actions are permissible. Scanlon (2008: 125) believes that this is true of many, if not all, otherwise permissible actions done for wrong reasons. Since the defenders of the relevance of intentions disagree, saying that, thickly described, those are impermissible, we might appeal to less contentious cases. Although their permissibility was also disputed, a better example of optional but blameworthy actions are,

I believe, suberogatory actions, e.g. moving one's lawn on early Sunday morning (Haji 2002: 39, see Driver 1992).

A more serious challenge is that the reason why Jones's action is intuitively blameworthy is (in part) because he kills Smith for reasons of his own. An action can then be blameworthy because it is done for a wrong reason even if one has no opportunity to act for a different reason. But why then not assume that an action can be (in part) impermissible because it is done for a wrong reason even if one could not but act for that reason? Either such actions can be both blameworthy and impermissible, or they can be neither.

Note that one cannot get out of the dilemma arguing that an action is never impermissible because of one's reasons for it. For the first horn only states that an action *can* be impermissible because of one's reasons for it. If true, that suffices to show that Ought Implies Can Refrain from Acting for a Reason is false and, therefore, that P1 is too. Whether an action that is blameworthy because it is done for a wrong reason is sometimes impermissible because of the same reason is a completely different question.

Another way out is the flicker of freedom strategy, which states that the intuition about the blameworthiness of Jones's action is explained not (only) by the fact that he acted for reasons of his own but by the fact that he thereby had alternative possibilities with respect to reasons for acting. That Jones killed for his own reasons shows that he was able not to intend to kill Smith. Black could not intervene ensuring that Jones kills Smith unless Jones could not, voluntarily, come to intend not to kill, which would trigger such intervention. Of course, Black may have ensured that Jones does not form the intention not to kill Smith, allowing him only to believe that

there are reasons in favor of kill him but then, on this view, Jones's action is not blameworthy (see e.g. Naylor 1984, Wedgwood 2013: 75-6).

Although this strategy faces some powerful criticisms (Fischer 1994: 137-47, 2006: 40-5, Stump 1999), I am inclined to believe that it is still the best reply to Frankfurt-style cases on the market. However, another issue is more pertinent for our purposes. Since Scanlon (2015: 100-1) takes Frankfurt-style cases among the reasons to reject the Principle of Alternative Possibilities and to think that what matters for blameworthiness of an action are one's attitudes, such as one's reasons for acting, actually reflected in the action, he owes us an explanation why actual attitudes are not sufficient for the possibility of an action being impermissible because of them. In other words, why do we need Ought Implies Can Refrain from Acting for a Reason? As the objection in question shows, one cannot have one's cake and eat it too.

Scanlon might reply that we do not need Ought Implies Can Refrain and its derivatives for P1. He (2015: 97) does claim that an action need not be something that a person could have avoided doing and still be blameworthy as long as it was "produced by and therefore reflective of the person's view of the reasons he or she has". But this assumes, as Scanlon (2008: 194) seems to agree, that even if, to be reflective of one's attitudes, an action need not be something one can refrain from, one's attitudes need to be under one's control in the minimal sense that one could have formed others. Scanlon (2015: 98) is right, I believe, that one can be open to criticism for one's current attitudes even when he is not to be criticized for their origin (e.g. repulsive attitudes formed by heavy upbringing), but he is wrong that Frankfurt-style cases, assuming they involve complete curtailment of one's options to form alternative attitudes and, hence, of one's reflective abilities, are even remotely similar.

I therefore believe that, not only is Ought Implies Can Refrain necessary for P1, but the Principle of Alternative Possibilities is necessary for blameworthiness. But even if I am wrong about the latter, it seems right to think that the choice of acting for one reason rather than another should be interpreted in terms of the ability to act for a reason other than that one actually acts for. If so, the dilemma remains for those who, like Scanlon, think that alternative possibilities are not need for blame-ascriptions but, presumably, are needed for deontic ascriptions.⁸⁸

I shall now address two central assumptions of the AFC – that it is impossible to choose to believe that something is a reason for an action and that it is impossible to choose to act for a reason.

2. Choosing Reasons for Action

Scanlon (2008: 193) seems to suggest that the assumption that one cannot choose to believe that something is a reason for action follows from the truth of doxastic involuntarism, the view that it is not only psychologically but also conceptually impossible to believe at will. Scanlon does not say why he believes that that view is true, referring (2008: 220, n. 13.) only to Hieronymi (2006, cf. 2011), so I shall take him to rely on her influential account.

I shall start briefly explaining why it is not possible to believe at will, proceeding to argue that the AFC's main error lies in the equivocation between the concept of 'choice' featuring in P2 and the one featuring in the other premises. My point is that, even if one cannot choose to believe at will, there is a sense of 'choice' such that one can choose one's reason for action. The relevant

⁸⁸ The point is generalizable to other forms of curtailment of alternative possibilities, such as hyper-compulsion cases (see, e.g. Graham 2011).

sense is that of control over beliefs expressed in the specific ability to believe otherwise, hence, in the ability to recognize certain reasons as relevant.

2.1 Why We Cannot Believe at Will

Let us first distinguish between two kinds of control, or ability, exercised with respect to beliefs. The first is the ability to think about what is true – to look for and assess available evidence, and to believe otherwise on the basis of opposing and better evidence. It is the ability to evaluate and reevaluate what is true. Call it the evaluative control (Hieronymi 2006: 53).

Exercising evaluative control is a voluntary process. It is under one's control to look for sufficient reasons to believe something. However, once one finds sufficient reasons to believe it, i.e. once one gets convinced by what they suggest, one, *ipso facto*, believes what they suggest. Suppose I want to know when the Great War started. I search it online, look at history textbooks, etc. If the evidence I find convincing suggests that it started in 1914, I cannot help believing that it did not. I cannot meaningfully say: "I am convinced that the war started in 1914. I shall now choose whether to believe it or not".

There is another kind of reasons with respect to believing. It does not count in favor of a belief but shows it good to have (Hieronymi 2006: 52, 57). Suppose that you hold me at gunpoint threatening to pull the trigger unless I believe that the Great War started in 1918. If I find my life dear I shall believe it would be good to believe that it started in 1918. I shall want to believe it but I shall not be able to believe it unless I see good reasons for believing it (cf. Alston 1988: 263, Parfit 2011a: 420).

However, there are ways to respond to reasons that do not bear on the truth of a belief but show it good to have. One can change the world so that it corresponds to what would be better for one to believe or change one's mind by taking action to affect one's existing beliefs. If I want to believe that my computer is shut down, I can shut it down. If I want to believe that the Great War started in 1918, I can approach evidence selectively, focusing on conspiracy books that point towards that conclusion. I might also take a drug which does the job, or get hypnotized to believe it or engage in some other forms of self-manipulation for that purpose. This is the second kind of control we enjoy with respect to beliefs. Call it the manipulative control (Hieronimy 2006: 54-6, cf. Parfit 2011a: 421-4).

With the help of the concepts introduced, it is easy to see why we cannot believe at will. To believe at will, as understood in this debate (Hieronimy 2006: 58-64), would be to choose to believe differently than the reasons one finds convincing suggest. As we saw, in some cases, one can believe that it is good to believe that something is true. But that is not the same as to believe that that is true. To believe against the best evidence one might manipulate oneself or the external world. But that is to act in order to induce a belief in oneself. Either way, we cannot believe at will.

Does that result undermine the possibility of choosing reasons for action? *Pace* Scanlon, I shall argue that it does not.

2.2 An Important Equivocation

It is worth quoting Scanlon (2008: 60) at full length here:

[W]e do not, it seems to me, *choose* which things to take as ultimate reasons. We have to decide whether something is a reason or not—this is part of our being responsible. But deciding in such a case is not choosing, because it lacks the relevant element of free play. When one sees each of several courses of action as supported by sufficient reasons, one can choose which of them to take. In this sense one can choose which reasons to act on. But [...] when one sees several considerations as counting in favor of the same action, one cannot choose to act on one of them rather than another unless one downgrades some of the reasons by changing one's mind about whether they really do count (or count sufficiently) in favor of the action in question. Changing one's mind in this way is a judgment—a decision—but not a *choice*. Holding constant one's judgment about which considerations count in favor of an action, there is no such thing as choosing, or “singling out,” one of these as *the* reason that one is acting on.

He (2008: 61) then adds:

One can, of course, choose to bring it about, or try to bring it about, that one regards or does not regard certain things as reasons. [...] But what I choose in these cases is not to see or not see certain things as reasons, but to undertake a process of self-manipulation aimed at bringing it about that I have such attitudes.

These paragraphs explain how Scanlon understands the move from P2 to P3. The first step in the argument states that forming a belief that something is a reason for action is involuntary. As we saw, when one finds good reasons for believing something, including that something is a reason for an action, one cannot but believe it. The second step rules out manipulative control (Scanlon 2008: 60-1). One can choose to cause oneself to regard considerations one did not take to count in favor of an action as decisive by manipulating oneself or the external world. But a morality which would require such manipulations in order to avoid acting for a wrong reason is highly

unattractive. What is more, as Scanlon (2008: 61) observes, to engage in manipulative control is to bring oneself to believe something and not to choose to believe it on the basis of evidence.

These two steps add up to justify P2. It is the third step, which aims to justify P3, that is controversial. Scanlon takes the fact that one cannot choose to believe that something is a reason for action to imply that one cannot choose one's reasons for action and, consequently, that one cannot choose to act for a reason. Both inferences are, I believe, flawed. I discuss them in turn.

Scanlon starts from a true assumption that choosing one's reasons for action involves the change in one's mind about their strength. One downgrades the other reasons, as he (2008: 60) puts it. The problem arises because he takes this to show that one cannot make a choice when judging reasons as more appropriate than others or inappropriate at all. This is because, Scanlon (2008: 60) claims, one thereby makes a decision, which is different from choosing.

I have no disagreement with Scanlon about the process of making up one's mind about reasons for action. I disagree that changing one's mind about which reasons count in favor of an action is not choosing, in the sense of it needed to establish the possibility of choosing reasons for action and, consequently, of acting for a reason. The sense of choice relevant here is that of the ability (and the opportunity) to believe otherwise. We enjoy that ability because we exercise evaluative control over beliefs. It is up to us to look for and to assess the evidence for or against a belief. We can try to find additional reasons to make beliefs more justified and we can revise or abandon beliefs in virtue of what we take to be better reasons. The same is true of control over beliefs about which reasons count in favor of an action and about the relative strength of those reasons. Nothing more is needed for the possibility of choice of one's reasons for acting. That

belief formation is involuntary is beside the point. Forming a belief and acquiring it are logically independent phenomena.

Recall the example of the selfish rescuer who does not care about the drowning child and thinks that no other consideration apart from his self-promotion favors saving the child. As said in the first section, assuming he is a cognitively functional and socialized adult, the rescuer is able to believe that the child's life is relevant. Since moral considerations like that one are not esoteric, it is not impossible to demand him to know better. Suppose he responds: "I could not save the child but for the sake of self-promotion. I was *convinced* that this selfish reason is the only relevant one. I thus simply could not *choose* whether to believe that. So I could not choose to believe or to act for a different reason." We could reply: "It is true that you could not choose whether to believe that that reason is the only relevant one once you became convinced of it. But you should not have become convinced of it in the first place. You should have known better. You were able to know better, so you could have acted for another reason."⁸⁹

The rescuer's reasoning resembles a popular argument for the impossibility of doxastic 'oughts'. The argument (Alston 1988: 115-52, Levy 2007: 133-4) states that, because one cannot form a belief voluntarily, one cannot be required to believe other than what one believes nor can one be responsible for it (in other versions of the argument). We cannot blame those who believe that the Earth is flat or that the Holocaust never happened nor can we say that they are required, epistemically or morally, to believe differently.

⁸⁹ To my knowledge, McMahan (2009a) is the only author objecting to the AFC on the same grounds. He (2009: 356) writes: "When an agent fails to see the force of a normative reason she actually has, and thus fails to act as she ought to, we (...) condemn her for having failed to engage sufficiently in a kind of deliberation of which she is capable and that should eventually have enabled her to see the force of the reason and to decide to act on it." McMahan does not develop that thought further.

I find this a *reductio ad absurdum* of that argument. The argument went wrong, I believe, at the point one took the impossibility of belief formation to entail the impossibility of choosing to believe otherwise. The inference is problematic since the truth of the antecedent is irrelevant for the truth of the consequent. When Holocaust deniers get persuaded by historical revisionists that gas chambers are a hoax, they *ipso facto*, believe so. But that does not mean that they ought not to know better.

It is customary to respond that evaluative control is indirect: obligation and responsibility to believe, as suggested, are derived from obligation and responsibility to act, namely to decide to approach a body of evidence and evaluate it. But such an indirect control, some people (e.g. Zimmerman 2008: 175-93, Levy 2007: 145-8) argue, does not suffice for obliging or holding one responsible for a belief. I remain unconvinced by those arguments. However, engaging with them here would lead me too much astray into complex issues about the nature of doxastic obligation and responsibility.

A more interesting objection is that, if one does not recognize some reason at the moment of acting, one is unable to act for it, nor can one be obliged to act for it at that moment (but only on some future occasion). For example, Ross (2002: 5, 1968: 114-6, cf. Prichard 2002: 3) wrote:

It is not the case that I can by choice alone produce a certain motive (whether this be an ordinary desire or the sense of obligation) in myself at a moment's notice, still less that I can at a moment's notice make it effective in stimulating me to act. I can act from a certain motive only if I have the motive; if not, the most I can do is to cultivate it by suitably directing my attention or by acting in certain appropriate ways so that on some future occasion it *will* be present in me, and I shall be able to act from it. My *present* duty, therefore, cannot be to act here and now from it.

The claim that one is unable to believe something if one does not in fact believe it is an instance of a more general view about abilities being exhausted by what one actually does, believes, etc. In the first section, I suggested that one is often able to do, at the time of doing it, more than what one in fact does. This is clear when we consider cases where one could have easily performed better than one did, e.g. a skillful basketball player who misses an easy shot or a mathematician who miscomputes two one-digit numbers. I have also said that, as normally functioning adults, we are able to do some things above certain thresholds, including developing a moral sense about which reasons for actions are wrong, at least for a range of situations. The latter is a crucial part of our moral development.

As long as the time span before one's action was sufficient for one to engage to develop an ability to do or believe what was expected of him as a normally functioning adult, one cannot justifiably object that he was unable to do or believe it just because he did not do or believe it.⁹⁰ For sure, the selfish rescuer cannot appeal to the lack of time to realize that his reason for saving is wrong; he had all his life to arrive at that belief. This is true even if he was thought vainglory all his life; it was enough for him to know that death is bad and ought to be prevented at a trivial cost (cf. Adams 1985: 27). The morality which would demand him to act but not from vainglory might be demanding. But that shows, at best, that some moral requirements to act for a reason are unappealing but not that they are impossible. I therefore do not think that Scanlon can help himself adopting Ross's argument.

⁹⁰ It might be objected that this presupposes a libertarian-style ability to believe otherwise in spite of all the actual conditions of the world at the time of believing something. While it is in line with that view, it is also in line with various compatibilist views, such as the view that, for it to be true that one could have believed otherwise, it is enough that he would have believed it if he had decided to.

Finally, one can object that the AFC holds if one lacks the ability to believe otherwise, as it is presumably true in Frankfurt-style cases. But if one has no ability to believe that the reason one takes as decisive counts less than one thought or not at all, then one has no ability to act for a different reason. And since P1 presupposes Ought Implies Can Refrain from Acting for a Reason, such a person cannot fall under the requirement to act for a different reason. Therefore, the AFC neither holds in such cases.

Interestingly, as the first of his quotes given above shows, Scanlon accepts that we are responsible for what we believe. Why does he then fail to see its relevance for the possibility of choosing one's reasons for action? I think he falls prey to an equivocation due to the ambiguity of the verb 'choose'. When saying that one cannot choose one's reasons for action, we use the verb 'choose' in the sense of being up to one whether to believe otherwise, that is, in the sense of being able to make up one's mind about the relevance and strength of certain reasons and not, as Scanlon thinks, in the sense of singling out a preferable reason from the set of alternatives.

Both senses of choice seem to be present in ordinary speech. We use the 'singling out' sense when, for example, one opts for a chocolate ice-cream among the available flavors. But we also say that we should choose life-paths for ourselves, by which we do not mean that we should single out a preferable life-path from some list of the available paths but that we should be able and have the opportunity to pursue a life-path. This is merely a terminological point and might not be the best error theory of Scanlon's view. Even if one disagrees that the concept 'choice' can be used in the latter way, it remains true that we have the ability expressed in that usage.

2.3 Why We Can Choose to Act for a Reason

The most prominent strand of responses to the AFC was to deny P4 (Walen 2009, Lillehammer 2010, Kolodny 2011a, Nelkin and Rickless 2014). It was argued that, although one cannot choose to believe that something is a reason for an action, one can choose to act for a reason. I agree that P4 is false but for the same reason I believe P3 is: one can choose to act for a reason because one can choose one's reasons for action.

The former choice might occur differently, depending on how one sees the considerations counting in favor of an action. Suppose first, somewhat artificially, that one believes that only two reasons favor an action, one right and the other wrong. Suppose also that, not being a moral exemplar, one takes the latter to be conclusive and the former to be merely sufficient. This would happen if the rescuer recognized the child's death as a serious consideration in favor of saving but took it to have less strength than his self-promotion (cf. Nelkin and Rickless 2014: 143). Can he still be required to save the child for the right reason? On the view in question, the answer is 'Yes', as he might choose to act for what he sees as a weaker reason. That would not be irrational as long as he believed that the weaker reason is merely sufficient (Kolodny 2011a: 105, n. 9).

What about the case where one is as wicked as to believe that nothing but a wrong reason favors an action? This is the original selfish rescuer example. Scanlon (2008: 61) meant the AFC to apply only to such cases.⁹¹ For different reasons, these authors deny that they are of great moral significance.

⁹¹ Considering a similar case with the exception that the rescuer sees neither any self-interested reason to save the child, McMahan (2009a: 355) and Kolodny (2011a: 106) argue that Scanlon must conclude, absurdly, that the rescuer is not obligated to save the child because he cannot choose to save for a reason he does not take to count in favor of saving. But this follows only if Scanlon makes a further assumption that one cannot act unless one acts for a reason, which he does not make (personal communication).

For instance, Kolodny (2011a: 106-7) argues that while it is true that the rescuer cannot be morally required to save for the right reason he can be required to save but not for the wrong reason. Given the rescuer's belief that nothing but his selfish interest favors saving, morality can demand him to save irrationally, for no reason at all.

On the other hand, Walen (2009: 18-21) argues that the selfish rescuer cannot save nor be obliged to save for the right reason when he does not see it. However, he thinks that such cases are rare. To completely disregard a child's death as a consideration in rescuing shows that one is seriously morally impaired, at the border with sociopathy. Walen (2009: 20) claims that the same is true if one disregards any reason but the wrong one. There is no right reason so esoteric that a normally functioning adult cannot be able to recognize its force.

This may sound overly optimistic. It can be said that, while spotting the wrong reasons is not difficult, recognizing the right reasons might often be. But even if this is true, we should note that the central issue of the debate is not whether one can be morally required to act for a right reason but whether one can be required to act but not for a wrong one. The Doctrine of Double Effect and the similar principles prohibit only acting for wrong reasons.⁹² Scanlon might then be attacking a straw man claiming that it is not possible to be required to act for the right reason. However, one can restate his question as whether morality can require acting but not for a wrong reason. If the AFC is sound, the answer is still no.

There is a more important problem with the stated response. Its upshot is that, when one would not take anything but the wrong reason into consideration, one either could not be morally

⁹² Hurka (2014b: 48-9) interprets the doctrine to prohibit acting for a wrong reason rather than to require acting but not for a wrong reason. On his interpretation, the one who would act for a wrong reason should not act at all, even if the consequences of abandoning the action would be detrimental. Many defenders of the doctrine understand it in the other, less restrictive, way (e.g. Tadros 2011b: 156-7, Wedgwood 2011a: 388-91, 2011b: 468-9, Liao 2012: 716).

required to act but not for it unless one could be required to act irrationally, for no reason at all. On the account I defended, in such cases, one can be required to act but not for a wrong reason because one has the ability to look for and assess other reasons.

It seems much more intuitive to reproach those who act for a wrong reason saying that they should have had a different judgment about which reasons count in favor of the action rather than that they should have acted against their better judgment about it. We do not think that, in such situations, one should have acted irrationally but that one should have known better. This is not to say that one can never be morally required to act irrationally, nor even that this could not happen in the selfish rescuer case, but that we need not settle those questions to see whether one can be required to act for a reason. Apart from being more intuitive, my account is simpler.

3. How the AFC Proves Too Much

The AFC entails that permissibility applies to thin act types only. Thick act types of the sort ‘X-ing for a reason’ are neither impermissible nor permissible. I shall now briefly argue that, if that is true, it can never be impermissible to lie or to perform any other *prima facie* wrong thick act-type. That is a striking conclusion and by itself a compelling reason to reject the AFC.

Imagine that Smith’s acquaintance, Jones, enjoys an ideal companionship with his wife. Being envious of Jones, Smith intentionally deceives him by telling him that his wife is unfaithful. Since Smith cannot choose to believe that there is a reason to tell it to Jones other than to deceive him, the AFC implies that he cannot be morally forbidden to tell it for the sake of

deceiving Jones. In other words, Smith cannot be forbidden to lie, which cannot be right as lying is clearly impermissible in this case.

In reply, Scanlon can admit that the AFC revises ordinary language, rendering thick act type obligations meaningless. But he can suggest that morality requires, permits, and forbids thin act types and that that is sufficient to resolve the problem. For example, the AFC is compatible with there being a moral prohibition against telling falsehoods such as Smith's in circumstances such as his. Smith could do away with his duty not to lie without there being a duty of a kind 'do not to tell falsehoods in order to deceive' (cf. Stocker 1970b: 596-600). Although revisionary of ordinary language, the proposed account is not revisionary of moral obligations.

This reply will not work. To understand why, imagine that Jones adores pranks and that Smith knows that. Jones would welcome Smith telling him the same falsehood if he discovered that all was a prank. It is plausible to assume that it is permissible for Smith to tell it in order to make a prank. But if the AFC is sound, Smith cannot be permitted to do *that*. He can only be permitted to tell it or not to tell it, *simpliciter*.

Scanlon cannot reply that making pranks is neither permissible nor impermissible and that what Smith may do is tell the falsehood, *simpliciter*. As we saw, he would have to say that the same thin act type is impermissible. Smith's action would be characterized as impermissible and permissible, which makes sense only if one accepts that permissibility applies to thick act types. It is then possible to say that, under one description, an action is permissible and, under another one, it is not.

The objection stands independently of whether Smith acts impermissibly because his intention is deceitful as well as whether he acts permissibly because his intention is to make the

prank. Suppose that Smith is morally forbidden to lie because he breaches Jones's trust and that no harm of that sort occurs when he makes the prank. Contrived cases aside, the harm of a deceit and the harm of a prank differ in degrees. When Jones realizes that it all was a prank, he feels relieved, which does not occur when he is made to believe the falsehood. The prohibition against lying is just a prohibition of the form 'do not harm to a degree X in circumstances C'.

If this is true, we can substitute the prohibition against lying with any other prohibition which results in the same degree of harm. But that does not make sense. We want prohibitions to be distinctive for each wrong even when the harm involved is identical. To preserve the wrong of lying we have to preserve the prohibition against telling the falsehood in order to deceive.

The problem is generalizable to any example of thick act types of the same thin act type differing in permissibility. For example, I see no way in which Scanlon can consistently hold to the AFC and claim that tactical bombing is sometimes permissible but terror bombing is not. Even if the difference is not in the bombers's reasons for acting but in the expected harm to non-combatants (Scanlon 2008: 28-9), the prohibition against more harmful bombing is applicable only if the prohibition against bombing in order to kill non-combatants exist. The 'in order to kill non-combatants' part is what makes the act of terror bombing a *distinct* wrong, even if it is not what makes it a *wrong*. It therefore seems to me that thick act type requirements and prohibitions are inescapable.

Conclusion

I gave three objections to the AFC. In the first section I argued that P1 follows from the Ought Implies Can Refrain and that, even if some standard objections to that principle, like Frankfurt-style cases, can be resisted, Scanlon needs to explain why that principle is necessary for deontic ascriptions but is not, as he believes, for ascriptions of blameworthiness.

In the second section, I rejected P3 and P4. I argued that, even if direct control over belief formation is impossible, Scanlon is mistaken to infer from it that it is impossible to choose one's reason for action. I argued that one has the ability to look for and to assess evidence for beliefs about reasons for action and that nothing more than that indirect control is needed to establish that one can choose to believe that something is a reason (in that sense of 'choice').

Finally, in the third section, I presented a *reductio ad absurdum* of the AFC, arguing that it implies that *prima facie* wrong thick act-types, such as lying, are never impermissible because thick act type prohibitions are impossible.

These objections establish that morality *can* but not that it *does* require or forbid acting for a reason. They have no bearing on the questions when and why intentions or one's reasons for action matter for permissibility. Scanlon might be right that they never matter but that cannot be because of the AFC.

7. Bad Intentions and Third Party Intervention

Introduction

On one influential argument, bad intentions, such as the intention to harm, are irrelevant to the permissibility of actions—in the sense of rendering otherwise permissible ones impermissible or wrong ones more wrong—because a third party may and, sometimes, should encourage actions done with them. So either these actions are not impermissible or it is sometimes permissible to encourage impermissible actions. But the latter is impossible, the claim goes, so these actions are permissible. Bad intentions do not affect their permissibility. Call this argument *the Argument from Encouragement* (AFE).

A classical statement of the AFE is in Thomson (1999). She imagines the following case (*Bad Doctor*).⁹³ A patient is near death and in anguish. He requests the doctor to end his life by a lethal dose of some standard drug. Suppose that, given the patient's condition and consent, there is a decisive reason to do it. The doctor, however, does not care about the patient and administers the drug intending only to kill him as an end in revenge. Thomson (1999: 516, cf. n. 19) writes:

⁹³ Its earliest mention is probably in Levy 1986: 34.

If we love [the patient], we too want [the doctor] to inject the drug. We can consistently believe it would be morally impermissible for her to act, while nevertheless wanting her to. But morality calls for us to feel ashamed of ourselves if we do. What is morally impermissible is, after all, exactly that: morality requires that the agent not do the thing, the agent must not do it. So if we really do believe it morally impermissible for the doctor to inject the drug, then it is a bad business in us to want her to, wanting this for the sake of the benefit to be got by the patient if she does.

Thomson seems to suggest that those who believe that the doctor's intention to kill the patient in revenge renders his action impermissible, must say that they cannot want him to administer the drug with that intention or, at best, that they can want it but not without feeling shame for that desire, assuming that they take morality seriously. But, certainly, one should want the doctor to proceed without feeling ashamed. Moreover, being in the position and having the authority to do so, one should encourage him, regardless of knowing that he would act with that intention. The objection serves as a *reductio* of the conclusion that administering the drug is impermissible and in general of the view that intentions are relevant to permissibility.

A similar point is sometimes made by stressing the absurdity of third party prevention of acting with bad intentions rather than of not encouraging it. Consider the following variation on *Bystander*, the original trolley case (*Bad Bystander*).⁹⁴ A runaway trolley is headed toward five persons trapped on a railway track. They will be killed unless something is done. The trolley can be redirected onto a side track where one person is trapped. I am a bystander and near me is a switch. If I throw it, the trolley will get redirected onto the side track, thereby hitting the one and killing him. Now suppose that I throw it only in order to kill the one because I hate him, merely

⁹⁴ The example is familiar from McIntyre 2001: 228, Kamm 2007: 132, and Liao 2012: 716.

foreseeing saving the five. I would not redirect the trolley were it not for the one being trapped on the side track.

Regarding a similar case, Tadros (2011b: 160) writes: “Wouldn’t it be odd for [someone] to push [me] out of the way so that [he] can turn the trolley with my better motivations?” Those who find the question rhetorical (and, as we shall see, Tadros is not one of them), can pursue an argument tantamount to the AFE. They can argue, roughly, that intentions cannot be relevant to permissibility since, if they were, third parties who would fail to prevent otherwise permissible actions when they are done with bad intentions would act impermissibly, as such actions should be prevented. Call this *the Argument from Prevention* (AFP).

The AFE and the AFP pose a dilemma for the defenders of the relevance of intentions: either to claim that a third party has a strong reason to intervene in acting with a bad intention, by restraining from encouraging it or by preventing it, or to accept that it is sometimes permissible to encourage, or at least not to fail to prevent, an impermissible action.

In this chapter I assess different ways to resolve the dilemma. Taking the first horn entails that bad intentions are as morally important as to oblige a third party to intervene in acting with them even when there is no reason to do it with identical acting with good intentions. Comparing reasons to intervene in otherwise permissible acting with bad intentions with reasons to intervene in otherwise permissible acting to bring other harmless wrongs, I argue that both are at best very weak reasons and cannot lend obligations. I then turn to ways to take the second horn. The idea is that there are impermissible actions that can be permissibly encouraged. I first argue that some ways to show this are question-begging because they appeal to otherwise permissible actions done with bad intentions. I then argue that there are actions which are impermissible in virtue of

factors other than bad intentions but which may and perhaps should be encouraged. This suffices to refute the AFE/AFP.

1. The First Horn

1.1 Three Views on the Wrongness of Intending Harm

To take the first horn, one has to show that third parties have a conclusive reason to prevent or at least to refrain from encouraging an action when and because it is done with a bad intention. I start by outlining a particular defense of this claim advanced in Tadros (2011b). I then set out two other possibilities.

Tadros (2011b: 161) agrees that it would be odd for you to prevent me, a bystander at the switch, from redirecting the trolley in order to kill the one so that you redirect the trolley in order to save the five. He reasons that I may be more skilled in redirecting trolleys so, if you push me, you might jeopardize the chances to save the five. However, he adds that this does not imply that there is no reason for you not to intervene but just that the reason to do it is not strong enough. By redirecting the trolley intending to kill the one, I disrespect his moral status and, therefore, act impermissibly (Tadros 2011b: 161-2). But morality does not call for the prevention of each and every impermissible action. The weight of reasons that favor a third party intervention in acting with bad intentions is not exclusively a function of showing respect for a victim's moral status by not harming him with a bad intention. It also depends, among other things, on ensuring that the action's overall effects are not thereby worsened. In particular, we have much stronger reasons to

see to it that more people are saved wrongfully than that less are saved respectfully (Tadros 2011b: 165).⁹⁵

To avoid unnecessary complications, imagine, *per impossible*, that the predictable and the actual consequences of your and my redirecting would be exactly the same. We can stipulate that redirecting with intention to kill brings no additional harm or offense to the victim or those close to him (e.g. the one and his family and friends do not know that I intend to kill him) and that you harmlessly push me to take my place. Assume further that we are equally skillful, that we act as complete behavioral doppelgängers, etc. The only difference is in our intentions.⁹⁶

Tadros (2011b: 162) claims that, under those conditions, it would not be odd for you to push me and redirect the trolley yourself intending only to save the five. In fact, he thinks that that is what any well motivated third party is morally required to do because there is no other way to ensure that the one is killed respectfully. To be sure, in terms of the harm the one suffers, it is immaterial who redirects the trolley. But since I redirect it intending to kill him, I wrong him in the way you would not. Apart from the right not to be harmed, the one has a right not to be harmed with a disrespectful intention, which is, on Tadros's view, as stringent as to give rise to a duty for a third party to intervene and stop me.⁹⁷

⁹⁵ McMahan (2009a: 357-8) similarly claims that a military commander deliberating whether to assign a task to carry out a permissible raid to a skillful bomber, whom he knows would act only in order to cause terror and intimidation, or to an incompetent but good-minded bomber, whom he reasonably fears could get shot down by the enemy, should choose the former. His rationale is that, although the commander has reasons not to send the terror bomber—among which is the fact that it may be worse, personally and impersonally, if the victims are killed with the bad intention than with the good one—they are outweighed by the overall lesser harm the terror bomber would bring about (cf. Thomson 1999: 516, n. 19).

⁹⁶ In what follows, I shall always have in mind such contrived versions of *Bad Bystander* and similar cases.

⁹⁷ Some of Tadros' (2011b: 164) examples of disrespectful intentions, such as the intention to redirect the trolley away from the five because they are God's work of art, are dubiously disrespectful. Many religious people see others as God's works of art but respect them. It is better not to raise the bar too high, making rational understanding of what renders persons valuable necessary for forming intentions that are not bad. I shall, however, stick to intending harm out of motives everyone agrees are bad, e.g. intending to kill in revenge.

This view can be generalized to any instance of acting with a bad intention when there is no reason to prevent an identical acting with a good intention. For instance, Tadros has to accept that if, in *Bad Doctor*, another doctor could have administered the drug intending only to relieve the patient's suffering, he would have been morally obliged to do it, thereby preventing the other doctor from acting in the same way, apart from the intention to kill the patient. Thomson invited us to consider *Bad Doctor* on the premise that it is counterintuitive that anyone has such a duty. I share her intuition and so do many people whom I presented the case with. But since philosophy is not a popularity contest, to make headway, we need subtler criticisms of this view.

The task is twofold. First, we need to see whether the one is wronged when and because I redirect the trolley with the intention to kill him. Second, we need to see whether that gives a third party a conclusive reason to prevent me from acting with that intention. The first question is answered by accounts of the wrongness of acting with bad intentions, i.e. accounts that specify why there is a *pro tanto* reason against intending harm (or other bad effects, e.g. involvement in ways that foreseeably cause harm). They do not make a further claim that such intending suffices to render otherwise permissible actions impermissible or to oblige a third party to intervene. I examine three such accounts, victim-centered, agent-centered, and impersonal, and then discuss whether the reason against intending harm, if it exists, outweighs reasons against preventing it.

According to victim-centered accounts, among which is Tadros's view, the constraint on acting with a bad intention is based on a morally relevant feature of victims. What that feature is depends on a particular account. Respect for the victim's moral status is one influential answer. Tadros grounds moral status on autonomy of rational beings. He (2011b: 129-30, 153, 2014: 24) claims that to harm a person as a means intentionally is relevantly similar to coercing that person

to serve an end at the expense of being harmed in that way. But since we are autonomous, we are able to set ends for ourselves and to legitimately refuse to adopt an end. Compelling us to act against our will is what is disrespectful to us as rational beings. A paragon of disrespectful end is the end realized by harming us as a means, even if otherwise optimific. That end, Tadros claims, imposes a cost that we are not required to bear.

There are several problems here. First, in *Bad Bystander* I do not kill the one as a means. I kill him as an end and as a side effect of saving the five but intend to kill him only as an end. The justification just outlined is silent about the legitimate refusal to be intentionally killed as an end. Tadros could reply that, if we can legitimately choose not to adopt an end which harms us as a means and imposes on us the cost we are not required to bear, then we can, for the same reason, legitimately choose not to adopt an end which harms us as an end.

But that gets us to another worry. It seems that, by the same token, we can be granted the right to refuse to be harmed as a merely foreseen side effect. Tadros (2011b: 136-7) disagrees. He writes: “[I]n cases where a person is harmed as a side effect to save a life she is not compelled to adopt the saving of the life as her goal. She does not contribute to the pursuit of the goal and so pursuit of the goal is not imposed on her.” This suggests that what makes it true that the one in *Bystander* lacks the right not to serve the end of saving the five (and what makes it, therefore, true that he is not wronged when killed as a foreseen side effect) is that his death is not a causally necessary means to that end. But that is just another way of claiming that intended harming as a means (or intending harm as a means) is a wrong and foreseen harming as a side effect is not.

Either intending harm is a wrong because, qua autonomous beings, we can choose not to be harmed with intention to harm when and because that is costly for us, in which case we can,

for the same reason, choose not to be foreseeably harmed, or it is a wrong because we can choose not to pursue an end we have no duty to pursue, in which case all turns on whether and why we have such a duty. The second option makes the account trivial while the first one makes the duty to prevent intending harm hard to justify, at least without a further argument, given that there is no corresponding duty to prevent foreseeing harm when they would lead to identical otherwise permissible actions.

Although Tadros's account of the wrongness of intending harm is unsatisfactory, there are others worth considering. I do not believe that similar victim-centered views employing Kantian notions like inviolability or using as a means can succeed (e.g. Kamm 1996: 271-82, Nelkin and Rickless 2014: 149-150), and for similar reasons. Let us then turn to agent-centered views.

According to them, the agent has a reason to attend to himself in acting; deontological constraints, including the constraint on intending harm, are derived, as it were, from the inside of the agent outward. Intending harm is primarily bad for the agent (McMahan 2009a: 357). On this note, Wedgwood (2011a: 393, cf. 2009: 335-6, cf. Chappell 2002: 224-5) writes: "When your act has a bad consequence, the more agentially involved you are in bringing (it) about, the stronger the reason against the act will be", and much stronger it will be, he adds, if you intend than if you merely foresee that consequence.

However, this does not explain why intentional agential involvement in causing some bad effect is especially objectionable. It might be because it impairs our moral characters. If so, we would have a reason to avoid intending harm and third parties would have a reason to prevent us from it because there would be a reason to act out of good character and to prevent those who fail in that respect.

We do not, I believe, have such reasons. We might have a reason to improve ourselves morally, including to cultivate good intentions, but we do not have a reason not to act because our characters fail a certain standard. And while we might have a reason to help others become better people, that does not include thwarting their permissible actions (cf. Kamm 1996: 241). Otherwise we would have a reason to prevent one to act not only with a bad intention but out of a false moral belief (e.g. an employer who holds that certain races are inferior but who is not motivated by it in the hiring process).

Nagel (1986: 181-3) gives a more sophisticated account. He characterizes intending harm as “swimming head-on against the normative current”. When we intend harm, even as a means, Nagel claims, our actions are guided by evil, but evil is something to be repelled by rather than to be embraced. We ought not to be attracted by evil but to work on eliminating it. Nagel believes that the constraint on intending harm is agent-relative. But we can accept his explanation even if the constraint concerns us all in the sense of having reason to minimize its violations, which is what we must believe if we think that any third party has a reason to prevent intending harm.

Nagel (1986: 183) himself raises a powerful objection: this explanation might be limited to intending harm as an end. We might not be guided by evil in intending harm as a means. Much depends here on what we take intending to be. Imagine a doctor who prescribes chemotherapy to a patient with cancer, foreseeing that some of his organs will be damaged in the course of it but that that damage is necessary to destroy the cancer cells (*Chemotherapy*). Must the doctor intend to damage the organs? Some people think that the answer is ‘Yes’ as we must intend whatever we take to be causally necessary to our ends. I disagree. Even if the doctor will use chemotherapy until the organs are damaged, he need not intend to damage them but only to involve them as a

means of killing cancer cells. We can then be guided by evil and yet merely foresee it (cf. Kamm 1996: 254, n. 10). But even if he intends the damage, he does not wrong the patient.

Finally, one might think that it is impersonally bad that our wills are guided by evil (cf. FitzPatrick 2012: 103-4, McMahan 2009a: 357-8). Harm, and suffering in particular, are bad not only for those who suffer but impersonally. Injustice and inequality, some people think, are also doubly bad. Intending harm might add up to impersonal badness of harm because of its volitional commitment to harm, or what Nagel calls guidance by evil.

However, even if this is true, it does not establish that intending harm is a wrong *per se*. If intending harm is impersonally bad, we all have reason to want or hope that there is less of it. But it is a big leap from that claim to the claim that we all have reason not to perform otherwise permissible actions with intentions to harm or to prevent them.

I therefore doubt that intending harm is a wrong *per se*. I shall not dwell deeper into the strengths and weaknesses of these three accounts. I now want to grant that one of them offers the correct explanation for why reasons not to intend harm count *pro tanto* against the permissibility of actions. I shall argue that, even with that assumption, it does not follow that third parties are permitted to prevent otherwise permissible actions when and because they are done with bad intentions.⁹⁸

⁹⁸ Hanser (2005: 466, n. 20) suggests that we have a reason to prevent acting with bad intentions because it makes agents open to blame and we do have reason to prevent others act in ways that will leave them open to blame. This is true but immaterial for refuting the AFE/AFP. The aim of those arguments is to prove that bad intentions are non-derivatively irrelevant to impermissibility and not that they are non-derivatively irrelevant to blameworthiness or derivatively relevant to impermissibility through their relevance to blameworthiness. We therefore cannot take the first horn arguing that a third party has a reason to prevent or not to encourage an action because it is open to blame but because it is done with bad intention, even if the latter sometimes accounts for the former.

1.2 The Strength of Reasons to Prevent Intending Harm

My argument to that point is inductive. I compare cases where one would be harmed as much by the agent as by a third party but also harmlessly wronged by the agent. By analogy, I argue that, since a third party has no duty to intervene in such cases, we have reason to believe that the same is true in *Bad Doctor* or *Bad Bystander*.

But before it, note that the oddness of third party intervention in the latter cases partially comes from violating one's institutional or professional duties (cf. Liao 2012: 717-21). Albeit the doctor in *Bad Doctor* intends to kill the patient in revenge, the reason to stop him and administer the drug ourselves might not override his duty, *qua* doctor, to do it. Similarly, Tadros discusses a trolley variation where I am the trolley driver and would redirect it intending to kill the one while you, a passenger, would push me out of the way and redirect the trolley intending to save the five (*Bad Trolley Driver*). Even if you have a reason to do it, that reason might not override my right, *qua* driver, to decide how to drive the trolley.

To avoid these difficulties, we should focus on cases without confounding institutional factors. *Bystander* and *Bad Bystander* are good examples of such cases. Compare them with the following case first (*Clumsy Bystander*). You are naturally gifted in redirecting trolleys. You have never redirected one towards people trapped on tracks but, if you were to find yourself in a case like *Bystander* and you decided to throw the switch, your trolley would instantly kill whoever is on the side track. I am less gifted in that respect. My trolley would hit the one, cut his toe and only then run him over. As it happens, I find myself in *Bystander*. Aware of my clumsiness, you deliberate whether to push me out of the way and redirect the trolley yourself.

Do you have a sufficient reason to do it? The answer depends, I believe, on the extent and duration of the pain the one would feel. If the one would be almost instantaneously killed after his toe loss, I doubt that you would have a sufficient reason to intervene. The cost of prevention would be greater than the harm itself. Suppose that you know that he would be in intense pain for half a minute before the trolley would kill him. That fact would give you a sufficient reason to intervene. And if he suffered longer, that might give you a decisive reason, or a duty, to do it.⁹⁹

Clumsy Bystander shows that reasons for intervention are given by the victim's harm and that they are sufficient once the harm is greater than the cost of intervening. In *Bad Bystander*, the agent and the third party cause identical harm to the victim. If there is a reason to intervene in that case, it is given by the wrong of intending harm. Assuming that wrong exists, we should ask whether it is as great as to permit and, as Tadros thinks, oblige a third party to intervene. To do so, we should contrast the prevention of other harmless wrongs, holding all else equal.

Consider the wrong of unfairness. In *Unfair Bystander* everything is as in *Bystander* but there are two side tracks, left and right of the main track, each with one person trapped on it. If I decide to redirect the trolley, I have a strong reason to choose the side fairly, perhaps by tossing a coin or using some other randomizing procedure (cf. Uniacke 2014: 407-8). Suppose I decide to redirect to the left but for an irrelevant reason, e.g. because the switch for redirecting to the right is a few centimeters away from me and I am too lazy to stretch my hand. Although the person on

⁹⁹ These calculations are of course very imprecise. I do not think there is an answer to the question when exactly your reason to intervene becomes sufficient or decisive.

the left would be as dead as if the trolley were to hit him after being redirected fairly, he would be wronged over and above it because he would get no chance not to get killed.¹⁰⁰

It might be replied that he would not be wronged because he already had an equal chance of being saved as he was equally likely to be the one whom I could save by throwing the nearer switch (Parfit 2015b: 57). That reply works when the costs of saving the further person are not insignificant but this is not such a case: the right switch is just a few centimeters away. The reply does show that the reason to use a fair procedure is likely not decisive and that I would redirect permissibly even if I did it unfairly. What should interest us here is whether you have a sufficient reason to prevent me and redirect the trolley yourself, using some fair procedure.

It might be said that what gives you that reason is the fact that the one on the left would be worse off if I redirected unfairly; what matters is not just the actual harm but the risk of harm. I agree with this, although I do not believe that reducing the risk of harm matters as much as to give you a decisive reason to intervene. This seems to be because, as just said, the one on the left has been already given enough and as much chance to survive.

¹⁰⁰ Note that I do not wrong him because I act for a wrong reason but because I violate the requirement of fairness in saving. It is not clear that the reason I act for is wrong. Certainly, it is less wrong than if I were to redirect the trolley towards the one out of a racial animus. At best, my reason is wrong because it is not right, i.e. because it is not the reason of fairness. But then, every reason which is not that of fairness is wrong in this case. This shows that I wrong the one on the left not because of the reason I act for but because of my failure to use a fair decision procedure.

Kamm (2014: 485-6) gives a different argument. She imagines a variation of *Unfair Bystander* in which the trolley is redirected to the right but I can re-decide where it will go giving equal chance to both sides. I flip a coin but only in order to increase the chances to redirect to the left as I have a racist prejudice of the one trapped on that side. The coin selects the left track and I happily redirect to it. Kamm claims that I still act permissibly since the bad intention that guides me to use the fair procedure does not affect the permissibility of my action. But the defenders of the relevance of intentions can reply that, since it is impermissible to redirect being led by a bad intention in *Bad Bystander*, it is impermissible to use a fair procedure to redirect being led by a bad intention. The intention to kill the one because I hate him is not much morally different from the intention to flip a coin to increase the chances to kill the one because I hate him. If the former renders redirecting impermissible, so does the latter. Since this antecedent is what the debate is all about, Kamm cannot assume its falsity without begging the question.

Unfair Bystander shows that third parties need not be obliged to prevent any wrongdoing. However, notice that unfairness is here a different kind of harmless wrong than intending harm could be. Unfairness is a wrong because it thwarts the victim's chances not to be harmed. There is a description of an identical action, done fairly, where one would not be harmed. Since there is no description of an identical action done but not with intention to harm where one would not be harmed (in *Bad Bystander* and similar cases), we should compare the wrong of intending harm with wrongs which are the same in this respect.

One such wrong is that of harmless violation of private property. In *Phasing Bystander*, all is the same as in *Bystander* except that, before I redirect the trolley, I trespass the property of the one trapped on the side track. There are other roads to reach the switch but I do it via his land because it is a convenient way and I do not care about violating his property rights. It is plausible to assume that no one has a right to intentional and non-consensual trespass even when doing so causes no additional harm to one's property (Ripstein 2006). To make the case that of harmless wrongdoing, imagine that I can become intangible at will. I phase through this man's land, leaving absolutely no harmful effects, and then I redirect the trolley.

Taken independently of the trespass, redirecting is permissible. But it is not far-fetched to say that I wrong the one by harmlessly trespassing his land. Despite that, it is plausible that the wrong is not as important as to oblige or even permit you to intervene and redirect the trolley yourself, without violating the one's property rights.

Compare *Phasing Bystander* with *Clumsy Bystander* and *Unfair Bystander*. Intuitively, in *Clumsy Bystander*, you have more reason to intervene than in *Unfair Bystander*, and you have much less or almost no reason to do it in *Phasing Bystander*. Tadros would, I hope, concede this.

He should then explain why we should treat *Bad Bystander* differently from *Phasing Bystander*. He can say that they differ because the wrong of intending harm impairs the agent's agency to a greater extent than harmless trespassing, or because it is more disrespectful towards the victim, or because it is a greater impersonal wrong. But even if these wrong differ in kind, I do not think that they differ in stringency, or at least to an extent that merits serious discussion.

1.3 The Strength of Reasons not to Encourage Intending Harm

I shall now consider two alternative takes on the first horn of the dilemma, as ways to support the AFE rather than AFP. The first states that, even if we may not prevent acting with a bad intention, we should refrain from encouraging it.

There are numerous ways to refrain from encouraging an action, ranging from different types of discouraging to ignoring it. However, if there is no sufficient reason to prevent me from redirecting in *Bad Bystander*, there is no sufficient reason, I believe, to divert my attention so as to take my place, or to try to convince me not to act, or to condemn me to the same effect. If it is impermissible to manipulate me in these ways to get me to abstain from redirecting in *Phasing Bystander*, it is impermissible to do it to get me abstain from redirecting in *Bad Bystander*. The reasons not to encourage intending harm as such are similarly weak.

The second alternative is the view that, even if it is impermissible to prevent or to fail to encourage acting with a bad intention, it is obligatory to ensure, using morally acceptable means, that one acts but not with that intention. Note that there are two ways a third party can prevent an otherwise permissible acting with a bad intention: either by taking one's place and acting with a

good intention or by making one act but not with a bad intention. It is only the latter that, on this view, a third party is morally required to do (Walen 2004: 346-8, Hanser 2005: 465).

For example, you should persuade me to abandon intending to kill the one or give me a special pill for it. But if there is no time for conversation or no such pill exists or if I refuse to forgo the intention despite your efforts, you are permitted to let me act with it. The responsibility for intending to kill falls on my shoulders.

However, we can still ask whether we are required to intervene in cases where we know that a convincing argument or a special pill would make one abandon a bad intention before one acts. I doubt that we are. You would not have a decisive reason to secretly put a pill in my drink so that, in *Clumsy Bystander* or *Unfair Bystander*, I do no additional wrong. And you would not have even a sufficient reason to do it in *Phasing Bystander*. Morality is not as paternalistic as to permit us to manipulate others so as to ensure that they do not cause minor wrongs. We are not permitted to stuff others with intention-changing pills or to lecture them on the moral importance of abandoning bad intentions.

Imagine that we have a pill that, if put in a water supply, would erase all racist intentions in a population. Many would agree that we would be obliged to get rid of racism in such a way. Similarly, if we could motivate racists to abandon their intentions by a persuasive argument, we would have a duty to advocate that argument everywhere. How to account for these intuitions?

The answer lies, I believe, in the fact that racist and other bad intentions, as a rule, result in harmful and other wrongful actions (see Scanlon 2008: Ch. 2). With exception of the contrived cases we are dealing with, a malevolent person rarely acts identically to a benevolent one. While we are obliged to prevent others to act with bad intentions when and because they lead to great

wrongdoing, we are not permitted to prevent others to act with bad intentions when they lead to otherwise permissible actions. We have the above intuition not because we believe that we ought to prevent racist intentions as such but because we believe that we ought to prevent great harms and wrongs those intentions usually lead to.¹⁰¹

2. The Second Horn

2.1 The Question-Begging Views

To take the second horn, we have to show that it is sometimes permissible to encourage or to fail to prevent impermissible actions. On Walen's (2004) view, we can do that if we consider cases in which it is permissible to encourage actions which are impermissible because they are done with bad intentions.

Walen (2004: 342-4) claims that because, in *Bad Doctor*, the doctor could administer the drug intending only to alleviate the patient's suffering, he ought to do it with that intention and not with intention to kill the patient. However, that the doctor can administer the drug with the good intention entails that he ought not to do it with the bad intention only if the reason not to do it with the bad one outweighs reasons to administer the drug, *simpliciter*. But since that is exactly

¹⁰¹ On a different view (FitzPatrick 2012: 112-8), it is permissible to tolerate acting with a bad intention but it is not permissible to act with it if one could let a better motivated third party act identically but with a good intention. For example, if you were not present in *Bad Bystander*, I would be permitted to redirect the trolley. But if you were, I would be obliged to let you do it. However, you would not be permitted to force me to do it (FitzPatrick 2012: 117).

This intermediate position is unstable. If I am permitted to redirect with intention to kill the one, then the reason not to do it is at best very weak. Your presence could not make that reason stronger as I would then have a strong reason not to act. But on this view, I do not have that reason; I only have a strong reason to let you act instead of me. Your presence would have to provide me with another reason, namely to let you act. But that reason could not be given by the fact that I would be acting badly if I did not let you, since the badness of my action is at best minor. It needs to be the other way around: I would be acting badly because I did not respond to that reason. Nor could it be your presence that provides that reason, as I have no reason to let others act instead so that they commit no wrong unless intending to kill the one is a major wrong, which this view denies. Either way, it is untenable.

the point of disagreement between the defenders and the critics of the relevance of intentions, Walen begs the question against the latter.

Suppose the doctor is the only person capable of administering the drug. Others ask him to do it and he agrees under the condition that he does it intending to kill the patient. Although it is permissible to encourage him to act, it is debatable whether he would act impermissibly. Walen (2004: 352-3) claims that the patient has a right against the doctor not to act with that intention and against those encouraging him to act with it but that the latter is a weaker right. The rationale is that the right not to be harmed with a bad intention does not apply transitively with the same strength against a third party encouraging harming with that intention. It is the agent who ought to care not to act with a bad intention and not a third party, apart from caring for expectable harm of acting with a bad intention (Walen 2004: 352).

The problem with this view is that it can be accepted by the critics of the relevance of intentions. Even if the right not to be harmed with a bad intention is stronger than the right not to encourage such harming, it is an open question how much stronger it is. It might as well be that the two are very weak and that, therefore, the agent acts permissibly.

On another view (McMahan 2009a: 357, FitzPatrick 2012: 117-8), it is sole responsibility of the agent to avoid intending harm because reasons against intending harm are agent-relative; a third party lacks them by definition. Those reasons make it true that an agent who intends harm *himself* ought not to intend harm and act with it. But a third party is permitted to encourage him to intend harm if he would otherwise act permissibly.

That reasons against intending harm are exclusively agent-relative is a contentious claim. If it were true, the agent would himself have the aim not to intend harm but other people would

have no reason to wish or hope that there is less such intentions or to minimize their occurrences. There would then have to be something morally special about those who intend harm or about their relation to their intentions apart from the fact that harm is intended. I fail to see what that could be. It seems more likely that, if intending harm is especially objectionable, it is so agent-relatively *and* agent-neutrally. Notice that this is compatible with the correct account of reasons against intending harm being agent- or victim-centered. Even if intending harm is a wrong *per se* because, say, an agent who intends harm is guided by evil, it does not follow, at least not without an additional argument, that its wrongness is exhausted by the agent-relative prohibition of it.

A more serious problem is that, even if reasons against intending harm are solely agent-relative, they need not be weighty reasons. If intending harm makes an action a bit worse but not impermissible, then a third party permissibly encourages a permissible action. The defenders of the view in question can of course debate the weight of reasons against intending harm but they cannot take the second horn of the dilemma to avoid the AFE/AFP without begging the question.

We therefore need cases of encouragement of actions which are impermissible though not in virtue of badness of intentions they are done with but in virtue of some other facts.

2.2 Encouraging the Impermissible

Two such cases feature in the literature. In the first, *Preemptive Murderer* (Walen 2004: 344-9, Hanser 2005: 465), A is about to murder B, intending to kill him, not realizing that B is about to murder C, intending to kill him. D is a bystander. Aware of B's plan, he can encourage A to preemptively kill B. If A were to kill B intending to defend C, he would permissibly kill in other-

defense. But since A does not know B's plan, he cannot do so. He is just a victim of good luck, about to prevent a murder accidentally, by committing a murder himself. The thing is that D is permitted and perhaps obliged to encourage A to shoot knowing that he intends to kill B. Since B culpably attacks C and killing him is the only way of saving C, D is justified in encouraging A to preemptively kill B, despite the fact that A murders B. Therefore, it is sometimes permissible to encourage impermissible acting.

The second case, *Claw Hammer*, is particularly gruesome (Tadros 2011b: 161-2). A child is drowning out of reach. There is a boat nearby which you cannot operate. I am an experienced sailor and I could save the child with little efforts. However, I would have to dive into the water once I reach him, thereby ruining my new suit as there is no time to take it off. I care a lot for the suit, as much as that I would rather let the child die. The gory twist is that you happen to carry a claw hammer which you can give me and which I would use to save the child, jamming it in his eye socket to pull him out. Are you permitted to do it?

If yes, you would encourage an unnecessary butchery. The cost of saving harmlessly is just a suit, so it is very plausible that I would act impermissibly in saving with the claw hammer. However, as Tadros (2011b: 162) observes, it is better to save the child at the cost of an eye than to let him drown. You are then allowed, and, perhaps, even obliged, to pass me the claw hammer. Hence, an impermissible action can be permissibly encouraged (cf. Tadros 2011b: 131).¹⁰²

Preemptive Murderer and *Claw Hammer* illustrate actions which are impermissible not in virtue of bad intentions they are done with. C acts impermissibly because he knows that B is an

¹⁰² Since Tadros believes that it is sometimes permissible to encourage an impermissible action but not an action done with a bad intention, he owes us an explanation why the latter is different. It cannot be because intending to kill is, in itself, a stronger reason for a third party to intervene than butchering an innocent person's eye is. This casts doubt on the internal consistency of his view.

innocent person who does not consent to be killed and has no other reason to die. I save the child impermissibly because I do it in a harmful way when the costs of saving harmlessly are trivial. It is immaterial to the issue of permissibility whether my intention in saving is bad or the best.

The cases share some morally important features: third parties cannot do what is best, the agents can but do not want to do it and would rather do what is suboptimal under the condition that third parties encourage them to do it, and the suboptimal is a striking wrongdoing. The main, or maybe the only, reason why it is permissible to encourage such actions is that doing nothing is much worse.

We can draw two conclusions from here. First, since neither case stipulates that intentions are relevant to permissibility, the assumption that it is impossible to encourage an impermissible action permissibly, lying behind the AFE/AFP, is refuted on the grounds that both sides of the debate can reasonably accept. Second, the cases show a way out of the dilemma but they do not prove the irrelevance of intentions. There are morally relevant differences between encouraging impermissible actions and otherwise permissible actions done with bad intentions.

Compare, for example, *Claw Hammer* with *Bad Bystander* (cf. Tadros 2012: 106-7). One relevant difference is that, in the former, you, as a third party, cannot do what is best. You cannot save the child at the cost of a suit (as opposed to saving her with the claw hammer or not saving her), since by assumption you cannot operate the boat. But in *Bad Bystander* you can do what is best. You can redirect the trolley in order to save the five and not in order to kill the one.

Suppose that in *Claw Hammer II* you can operate the boat and save the child harmlessly. You would then be required to save her yourself and even to prevent me from saving her with the claw hammer. Saving the child's eye suffices to oblige you to intervene, much more than saving

the one's toe does in *Clumsy Bystander*. However, the wrong of harming with intention to harm, in itself, matters insufficiently, if it matters at all, to permit such an intervention.

Conclusion

The AFE/AFP aim to establish the irrelevance of bad intentions to impermissibility through the irrelevance of third party intervention in acting with bad intentions. They bring up a dilemma for the defenders of the relevance of intentions: either to accept that it is obligatory to intervene in acting with bad intentions or that it is sometimes permissible to tolerate impermissible acting. I argued that, although the former raises theoretically important strategies, none of them stands to critical scrutiny and that the latter is an easy way out of the dilemma, as there are cases where we have good reason to support impermissible behavior.

8. Considering Intentions in Decision Making:

What is so Odd about it?

Introduction

If intentions are relevant to the moral permissibility of actions, we arguably ought to take what we intend into account in moral decision making. For instance, if we intend harm but would act otherwise permissibly, we ought to drop that intention and then act (Wedgwood 2009b, 2011a: 388-91, 2011b: 468-69, Tadros 2011b: 156-7, Liao 2012: 716). Similarly, a moral adviser should tell to an agent to reflect on his intentions before acting even if he knows that the agent would act otherwise permissibly.

To many, such deliberation and advice seem rather odd. In her well-known discussion of the ethics of self-defense, Thomson (1991: 293) writes:

Suppose a pilot comes to us with a request for advice. “See, we’re at war with a villainous country called Bad, and superiors have ordered me to drop some bombs at Placetown in Bad. Now there’s a children’s hospital there too. Is it permissible for me to drop the bombs?” And suppose we make the following reply:

“Well, it all depends on what your intentions would be in dropping the bombs. If you would be intending to destroy the munitions factory and thereby win the war, merely foreseeing, though not intending, the deaths of the children, then yes, you may drop the bombs. On the other hand, if you would be intending to destroy the children and thereby terrorize the Bads and thereby win the war, merely foreseeing, though not intending, the destruction of the munitions factory, then no, you may not drop the bombs.” What a queer performance this would be! Can anyone really think that the pilot should decide whether he may drop the bombs by looking inward for the intention with which he would be dropping them if he dropped them?

Thomson (1999: 514-5) makes a similar point discussing the role of intentions in the ethics of terminal sedation:

Consider drug-injecting and imagine the following scenario. You are a doctor, and your patient is near death, in terrible pain that cannot be alleviated by any less than a lethal dose of morphine, and for that reason, he asks for the injection. Suppose that, pursuant to the hospital’s rules, you ask me—I am the appropriate hospital official—whether it is legally permissible for you to inject the lethal dose. I reply: “Well, I don’t know. I can’t tell unless you tell me what your intention would be in injecting the drug. If you would be injecting to cause death, either as means or end, then no. But if you would be injecting only to cause relief from pain, then yes.” This is an absurdity. How on earth could it be thought proper for the legal permissibility of acceding to turn on what the doctor intends to bring about by acceding? Surely it should turn on the patient’s condition and wishes.¹⁰³

Some other critics of the relevance of intentions to permissibility claim, explicitly or in passing, that such advice would be “odd”, or something “you would not” and “should not say” (Bennett 1981: 96-8, Scanlon 2000: 305, 2008: 19-20, 30). The tone of Thomson’s remarks suggests that

¹⁰³ Thomson (1999: 515) thinks the same applies to moral permissibility. We can put legal permissibility aside.

she intended them as an objection to the view that intentions matter to permissibility, indeed as a *reductio* of it. Call this objection *the Objection from Oddness* (OFO).

However, the intuition about the oddness need not be shared. It might seem that there is nothing odd going on. And even if one is inclined to think that there is something odd, one might not understand what exactly that is. Therefore, the OFO needs to be developed and justified. We need to show why it is odd to consider intentions in decision making and why that counts against their relevance to permissibility, of course without appealing to the fact that they are irrelevant to permissibility, as that would beg the question (Lippert-Rasmussen 2010: 554, 2014: 147-8).

In this chapter, I examine three groups of views (some of them are explicitly endorsed and others are proposed so as to exhaust most, if not all, possibilities): that considering intentions in decision making is uncommon, that it is somehow conceptually confused, and that it focuses on morally unimportant considerations. I argue that the third view is probably what most of those who appeal to the OFO have in mind.

The minimum that the OFO must state is that, if it is odd (in the best understanding of it) to consider intentions in decision making, we have good reason to assume that they are irrelevant to permissibility (Lippert-Rasmussen 2014: 146-7). One might say that the inference is false as the oddness remains even if intentions are relevant. If I am right that the intuitive unimportance of intentions is what explains the oddness, that cannot be the reason why the inference is false. Rather, it is because it is circular; the OFO is not an objection proper.

A brief clarification is needed before we proceed. The OFO can take deliberative (or first-person) form and advisory (or third-person) form. Examples that motivate it are often presented as dialogues between the one seeking an advice and an informed and morally sensible third party,

often depicted as the one's superior (although nothing important hinges on this fact). But if it is odd to advise others to consider intentions in decision making, it is no less odd to consider them in personal decision making (cf. Lippert-Rasmussen 2010: 544, Tadros 2011b: 146, 149). There seems to be no difference in oddness of considering intentions between deliberative and advisory perspectives. Some disagree and I shall examine their reasons below. But otherwise I shall switch between the two perspectives assuming that nothing is lost in the argument.

1. Doing the Unordinary

One reason why it might be odd to consider intentions is that no one does it. One might reason:

- (1) If it is unordinary to consider or advise others to consider something in decision making, then it is odd to do it.
- (2) It is unordinary to consider or advise others to consider intentions in decision making.
- (3) Therefore, it is odd to consider or advise others to consider intentions in decision making.

1.1 The Fallacy

On the face of it, the first premise might look true. Because it is unordinary to consider morally irrelevant factors in decision making, such as one's body fat or Zodiac sign, it might be thought that the entailment from the unordinariness to the oddness holds. But these are morally irrelevant not because they are not usually taken into consideration but because there is nothing in them or that to which they relate that counts in favor or against an action.

An instance of this fallacy is the claim that, if intentions were a proper object of practical deliberation, they would be like other factors we consider, in particular the facts about the effects of actions or the way they are done. These facts are, so to speak, outward phenomena. They are about the world and not about our mental lives. So if we were to consider intentions in decision making, we would have, as it were, to look inward, turning practical deliberation into a form of self-investigation, which might be seen as odd (Thomson 1991: 293, Scanlon 2011b: 491, n. 1, cf. Tadros 2011b: 146).

This might be indeed odd but not because of its putative unordinariness. For even if no one has ever decided whether and how to act by looking inward to his intentions, it could still be cogent that he should do it.

1.2 The Facts

The second premise is also false. As McMahan (2009a: 348) observes: “Most people throughout recorded history have not found it queer to advise others that the permissibility of their action can depend on their intention in acting”. McMahan is, however, less optimistic about the history of considering intentions in personal deliberations (he has certain conceptual reservations about it, which I turn to later). I think there is ample evidence for both.

One example: when, in early 1942, Churchill decided that the RAF raids the populated areas of Nazi Germany in order to break the civilian morale, many British officers objected appealing to badness of the intention behind such raids. They had thought that it is a matter of

principle that civilian casualties are brought “only insofar as [they] remained a by-product of the primary intention to hit a military target” (cited in Walzer 1977: 257).

The officers were of course expected to obey the legal conventions on aerial warfare of the time, all of which proclaimed that deliberately aiming to kill civilians is impermissible (see *Hague Convention* 1907, IV, art. 25 and 1922-1923 *Hague Rules of Air Warfare*, arts. 22 and 24, cf. Grayling 2006: Chs. 4 and 5). However, their objection was essentially about moral, not legal, permissibility. They knew that strategic bombing might cause as many civilian casualties as area bombing but they objected to the latter because of what it intended: to demoralize, terrorize, and revenge. Whether these objections were legitimate is another question. But that they were raised is sufficient to demonstrate that it need not be unordinary to make military decisions on the basis of intentions behind them.

Some may argue that medical decisions are different. A doctor who would abstain from administering a lethal drug to a terminally ill patient—who might have even requested for it—on the basis of his bad intention toward the patient would make a precedent in clinical practice. But this is far from truth.

A recent study by Douglas, Kerridge, and Ankeny (2013) brings in-depth interviews with eighteen palliative care specialists regarding their experiences of considering intentions when performing terminal sedations. It shows almost a unanimous agreement about the relevance of cultivating the intention to ease patients’ anguish, as opposed to the intention to cause death or harm. Although majority of the doctors said that they were not religious, they insisted that, even when the patients’ deaths were a foreseeable effect of injecting the drugs, they acted intending to ease suffering and not intending to kill. The doctors were explicit that, apart from not wishing to

be the kind of people who intend to kill, they took their intentions into account when deciding whether what they were doing had been permissible (cf. Kendall 2000).¹⁰⁴

To repeat, I am not saying that the fact that doctors assess the permissibility of harmful medical actions on the basis of their intentions entails anything about the soundness of a moral requirement which would demand them to do it. It may as well be true that the current practice is preposterous. But the fact that they do it suffices to show that the practice is not unordinary.

Everyday situations follow the same logic. We sometimes ask ourselves: “Should I help him because I promised or because I expect to benefit from him later on?”, “Am I kind in order to be seen as compassionate?”, “Do I love her or am I with her because I get used to it?”, etc. We also give advices: “If you give to the needy, do not do it to cherish your vanity”, “If you honor your parents only in order to keep the Commandment, you are doing it wrong”, “You may refuse to shake his hands because you are afraid of Ebola but not because he is black”, etc.

2. Doing the Impossible

The oddness may be due to a conceptual confusion or impossibility of a kind. One might argue:

- (1) If it is impossible or conceptually confused to consider or advise others to consider something in decision making, then it is odd to do it.
- (2) It is impossible or conceptually confused to consider or advise others to consider intentions in decision making.

¹⁰⁴ Needless to say, the doctors were schooled to comply with the Doctrine of Double Effect and they were expected to explain to the death audit committee that their intentions were the best (see Trankle 2014). Some bias towards the relevance of intentions is therefore inevitable. Nevertheless, we can safely assume that some reports are genuine.

(3) Therefore, it is odd to consider or advise others to consider intentions in decision making.

Consider the impossibility part first. The first premise rests on what might be called the Non-Odd Implies Can, a principle analogous to the Ought Implies Can (and a contrapositive of P1). Three remarks are necessary about the latter. First, for our purposes, the ‘ought’ in it is moral, as opposed to prudential or epistemic ‘ought’. It concerns moral permissibility and impermissibility. The oddness must then also be a morally loaded term. Second, the ‘can’ is that of the ability and the opportunity to act (Vranas 2007: 169-71, Graham 2011: 341-2). If we are unable and have no opportunity to act, we neither ought nor ought not to act and it is why it is odd to act. Third, the Ought Implies Can is famously attacked by the Frankfurt-type cases and cases of compulsion and addiction (see Frankfurt 1969, Graham 2011). I do not think that these objections succeed but I shall not try to answer them here. I shall therefore assume that the first premise is true.

2.1 Cartesian Psychology

Two arguments in favor of the second premise are relevant. First, considering intentions in decision making involves, as it were, looking inward for them. Anscombe (1981: 58) claims that the talk about looking inward and giving a little speech *in foro interno* before acting is a remnant of a naïve Cartesian psychology. She focuses on the moral implications of that idea, briefly remarking that, on the Cartesian view, intention is “an interior act of the mind which could be produced at will” (1981: 58, cf. Chappell 2013). I do not think that either of these points renders looking inward impossible or nonsensical. I examine them in turn.

Suppose first that acts of mind are to be explained, in some specified sense, in terms of acts of body, that is, that logical behaviorism is true.¹⁰⁵ Would the talk about looking inward still make sense? Of course it would. It was never meant to be but a metaphor for self-reflexivity (cf. Hanser 2005: 458-9). Self-reflexivity might take the form of an internal monologue, even a long one—as when one ponders his own deepest motives for acting—but it usually does not. In most cases, acting with bad intentions leads to strikingly different consequences than acting with good ones (Scanlon 2008: Ch. 2). So it is not difficult to realize the practical difference of acting with one rather than another intention without engaging in an internal monologue (Walen 2011: 23).

In other, more contrived cases, the consequences of an action and the way it would be done would not change if one were to act with another intention. The defenders of the relevance of intentions to permissibility usually have in mind such cases when postulating the requirement to forgo a bad intention and carry on with acting with a different one. We can ask two questions: first, can one consider an intention which one would act with and contrast it with one's possible intentions, holding fixed the predicted consequences of the action and all other facts about it; and second, would one have to engage in an internal monologue while doing it?

McMahan (2009a: 348) and Lippert-Rasmussen (2010: 554, n. 24, 559-60, 2014: 149-51) answer the first question negatively. They claim that intentions in general remain unsettled until one has resolved whether it is permissible to act. What one determines about the permissibility of an action, McMahan (2009a: 348) writes, “will usually affect what she finds if she looks inward for her intentions”. Similarly, Lippert-Rasmussen (2014: 149) claims that considering intentions

¹⁰⁵ As a student of Wittgenstein and a critique of the Cartesian view, Anscombe is often mistakenly assumed to be a behaviorist despite her explicit rejection of it. See, e.g., Anscombe 2005: Ch. 1. However, we can assume, *arguendo*, that behaviorism is true.

independently of permissibility judgments would not treat them “as an object of deliberation, but as a constraint in deliberations”. It would present future intentions as if they “settled, morally speaking, what (one) should do rather than something she could and should shape in accordance with her beliefs about moral permissibility” (Lippert-Rasmussen 2014: 150).

They can be understood as saying either that the judgment about the permissibility of an action often makes one unjustifiably believe that his intention is better or worse than it actually is or that one would (or should) not, at least not under normal conditions, make the judgment about the goodness or badness of his intention in isolation from the judgment about the permissibility of his action. The former is sometimes true but irrelevant to the first question while the latter is relevant but, I think, false.

I often deliberate whether what I am about to do—which I know would be permissible and even supererogatory—I would do only with selfish intent. If I answer positively but proceed to act, I do not conclude that my intention in acting ceases to be bad or that my action ceases to be permissible but that I may proceed despite that intention. Whether such an inference is sound is another issue but the phenomenology is not unique to me. It is common to any rational agent. With minimal cognitive efforts, we can reflect on our intentions even after we resolved the issue of permissibility. We need not, nor do I see good reason why we should, shape intentions by prior permissibility judgments (although, I agree, we sometimes do it). If I think it is supererogatory to give half of my salary to Oxfam, why would (or should) that make me believe, self-deceit cases aside, that *the* intent with which I do it, namely, to boast about it in front of my charity-minded friends, is good when it obviously is not?

It is also unclear why Lippert-Rasmussen thinks that considering intentions in decision making implies thinking about them as constraints on our actions imposed by an external source rather than something we control. To consider intentions is to deliberate about them and possibly to reconsider them in light of new beliefs, including beliefs about permissibility (assuming that we think intentions matter for permissibility). If I am aware that behind my intention to dispense largesse to the poor is the intention to boast about it, I might, if I find a need for it, reconsider the latter without thereby reconsidering what I am about to do (for instance, by committing myself to act without boasting about it). I can do it because I see that intention as bad, or because I fear disapprobation of others if they discover it, or for some other reason. In all of these cases, I am in control of deciding to ponder on my intention, to scrutinize it, and to forgo it. That self-reflective activity might include an internal monologue but it might as well be momentous.¹⁰⁶

2.2 Intending at Will

Anscombe's other worry about intending at will is the crucial premise in the second argument in favor of the second premise. On that note, Scanlon (2008: 58-62) argues that, because we cannot choose to believe that something is a reason for acting, we cannot choose our reasons for acting and, correspondingly, our intentions (on his view, what one believes is as a reason for acting is a necessary, and perhaps a sufficient, condition of intending, Scanlon 2008: 11, cf. Bennett 1995: 193-4). This implies, Scanlon believes, that we cannot act for a reason and, therefore, be required

¹⁰⁶ Interestingly, Lippert-Rasmussen (2014: 148-9) sees no oddness in considering intentions in retrospective terms and in cases where a third party deliberates about what an agent ought to do, given the intentions he holds and those available to him, without giving an advice. If we find considering intentions odd even in such cases, that is because, Lippert-Rasmussen (2014: 148) claims, we find intentions irrelevant to permissibility. I shall argue later that this is true of both kinds of cases.

or forbidden to act for a reason (or to act but not for a reason). It is only Scanlon's first claim that is of interest here.

I agree with this claim but I do not think that it shows that looking inward for intentions is nonsensical or impossible.¹⁰⁷ Note that the same is true of beliefs; we cannot believe at will. But we nevertheless have the ability to self-reflectively engage with what we believe. While I cannot believe at will that *the* reason to give to the poor is to boast about it or that Holocaust is a hoax, I am responsible for believing so. Being self-reflective, I am able to ponder on those beliefs and to scrutinize the evidence against them. If I find the evidence convincing, I can drop those beliefs and form others. The same is true of the ability to consider intentions. Since we are able to intend otherwise, we are in full control not only over what we intend but over what we do not but can intend.

As an illustration, suppose that one sees no other reason for acting apart from a clearly wrong reason. For example, I believe that nothing speaks in favor of giving to the poor except for boasting about it. Scanlon (2008: 61-2) takes such cases as a background for his argument. Can I act for a different reason even then? If I care about morality and I think that I have a reason to reflect on my reason for acting, I can take a step back to ponder on it even if, at that moment, I do not find it wrong. I am typically able and have the opportunity to see a reason as wrong even if I do not find it wrong. Such reasons are not esoteric (cf. Scanlon 2008: 58, 61). Therefore, I am able to realize that vainglory is not the purpose of charity even if I will in fact act out of it.

¹⁰⁷ Scanlon (2008) does not support his view of the oddness of considering intentions with this argument. I examine it as a possibility. The same is true of some other arguments discussed later.

2.3 Act-Types and Possible Intentions

I now turn to the charge that there is a conceptual confusion in considering intentions in decision making. We can again grant the truth of the first premise since it is hard to think of a requirement which would be conceptually confused and yet would not be odd exactly because of it.

If intentions are relevant to permissibility, one is morally required to ask oneself or those he advises: “Am I/you to X with a bad intention?” proceeding with a conditional: “If yes, then I/you must not X”, which many find odd. The inference is ambiguous between two descriptions. It means either that I/you must not X, *simpliciter*, or that I/you must not X with that intention. On one view, roughly, while the former is odd, the latter is not.

Wedgwood (2009b, 2011a: 390-1, 2011b: 468-9) reads Thomson (and Scanlon) as saying that it is odd to advise the prospective bomber that he should not drop the bombs if his intention turns out to be to kill the children but that he should drop the bombs if his intention is to destroy the munitions factory. This objection, Wedgwood claims, is a travesty of the view that intentions matter to permissibility because it confuses different act descriptions: thin act-types, or acting, *simpliciter*, with thick act-types, such as acting with some intention. Thomson assumes that her opponents are committed to prohibiting the one who would in fact act with a bad intention to act, *simpliciter*. However, they are committed only to prohibiting him to act with a bad intention.

So, while they would say that it is impermissible to drop the bombs intending to kill the children, they would not say that it is impermissible to drop them if, intention-independently, there is a sufficient reason for doing it. After all, there could be a bomber who would drop the bombs intending to destroy the munitions factory and, even more, that could be the one who will

in fact drop them intending to kill the children. As said above, normally functioning adults have the ability to ponder on the content of their intentions, so the bomber could realize that killing the children is not a morally proper goal.

Therefore, according to Wedgwood, the permissibility of an action turns on the intention one could and must have if he were to act permissibly but not on the intention one would have if one were to act (cf. FitzPatrick 2012). Wedgwood sees nothing odd in the following advice: “You may drop the bombs but not intending to kill the children. Whether you will in fact intend to kill them is, however, irrelevant to what you may do.”

I disagree that possible intentions determine the permissibility of actions (if any does). I also doubt that the critics of the relevance of intentions fail to distinguish thin from thick act types or that, even if they do, that accounts for the oddness. I defend these claims in turn.

Suppose that the bomber who will drop the bombs cannot but not drop them intending to kill the children and that there is no other bomber around. It is not that it is impossible for him to act with a different intention or that he is unable to do it but that the possible world where he acts with it is so remote that we would not normally say that realizing the ability to intend otherwise was an option at hand. Imagine, for instance, that throughout his life he has been thought that it is a worthwhile goal to kill the enemy’s children and that he had rather limited opportunities to question that belief. While he still could intend otherwise, the morality which would require him to do it is rather demanding.

This gives us a hint about what makes considering intentions odd (more about it below) as well as shows how actual intentions matter. If the bomber were to take seriously Wedgwood’s advice, he could legitimately reply: “I do care about morality. I know I am able not to intend to

kill the enemy's children but how on Earth could that not be the good intention? I need time and effort to realize what I get wrong." At this point, we cannot repeat the advice. Moral advice are not just utterances of true statements but action-guiding rules. Advisers should make sure that one could act upon them, given a number of factors, both internal and external to the agent. At the very least, advisers must take into account: the psychology of the agent and the history of its formation (the agent's intentions, belief set, character traits, and so on), the likelihood, given that psychology, that he will employ his ability to intend otherwise, as well as the constraints of time. The last factor is very important as it shows that, even in more ordinary cases, where the agent realizes the badness of his intention without great effort but cannot change it because he needs to act immediately, Wedgwood's advice makes little sense.

Therefore, permissibility cannot turn on possible intentions unless one would expectably act or refuse to act with them after a careful and rational deliberation, given the above constraints (and perhaps some others). But then the oddness remains, at least to my ear.

The oddness cannot also be due to the confusion of thick and thin act types. Suppose that the critics of the relevance of intentions believe that a bad intention the agent in fact acts with must render what he does, in the thin act type sense, permissible. While this would be false (the thin act type could be done with a good intention), it is irrelevant for the OFO. The charge runs best if one could not have helped oneself but acted with a bad intention and there was no other, better motivated, person to step in. Wedgwood must say that, in such a case, one should not have acted, *simpliciter*. For instance, the wickedly educated bomber is morally forbidden to drop the bombs, which amounts to saying that the raid should be aborted. It will seem no less odd to many that the bomber should deliberate towards that conclusion.

2.4 Actualism and Possibilism

Despite these disagreements, I agree with Wedgwood (2011a: 390) that the defenders of the relevance of intentions who embrace actualism about moral obligations—according to which, roughly, one ought to select between options according to what one *would* rather *could* do after selecting the option (see Jackson and Pargetter 1986)—must claim, rather implausibly, that it is impermissible to act, *simpliciter*, for those who would in fact act with a bad intention. Lippert-Rasmussen (2010: 557, 2014: 152-3) argues that the oddness is in part a result of many people’s actualist leanings and that the advice of the form “If you would X with a good intention, then it is permissible to X, but if you would X with a bad intention, then it is not” is incompatible with actualism, which entails a straightforward “Yes” or “No” to the question of permissibility.

To this I have three replies. First, it is debatable how many people accept actualism. I, for one, find it deeply implausible because it implies that we are morally required to do horrendous things as long as we would do them (and doing them would be better, in the relevant sense, than not acting in those ways, Wedgwood 2009c, Ross 2012). Second, the conditional advice is not incompatible with actualism. An actualist defender of the relevance of intentions who is unaware of the advisee’s intention has to frame the advice in the conditional form. His view is that one should act as one would in fact act, so he cannot comment on what one should do unless he is aware with which intention one will actually act. Third, if the actualist adviser is aware that one will act with a bad intention, his answer will be that it is impermissible for one to act, *simpliciter*. This is where he departs from a possibilist friend of the relevance of intentions—possibilism is the view, roughly, that one should act as one *could* but will not in fact act as long as the former

leads to a better outcome than the latter—who will say that it is permissible to act, *simpliciter*, since the advisee could act with a good intention. But I criticized this view for failing to take into account the fact that moral requirements which purport to be action-guiding rely, to an extent, on contingent facts about agents' psychology and the external world.

So I do not see how accepting either actualism or possibilism—in conjunction with the thesis that intentions matter for permissibility—gets us off the OFO's hook. On the one hand, the problem cannot be the conditional form of the advice as both views imply it depending on the adviser's knowledge of the advisee's actual intentions. On the other hand, both views suggest, oddly, that one should abandon the action, no matter how important it may be, if one could not help oneself but act with a bad intention (except for the most stringent type of possibilism which disregards any actual state of affairs, thus failing to be action-guiding).

2.5 The Question-Begging Charges

I now turn to three question-begging charges. On the first one, to consider intentions in decisions making is to make a category mistake. Intentions do not non-derivatively matter to permissibility but to blameworthiness or character evaluation (see Bennett 1981: 97-98, Rachels 1986: 93-95, Thomson 1991: 295-96, 1999: 517, Norcross 1999: 121-22, Scanlon 2008: 52-56).

However, the defenders of the relevance of intentions do not deny that intentions matter to these moral categories; they only claim that intentions, in addition, non-derivatively matter to permissibility. Absent an argument to prove that they cannot or do not matter in such a way, the charge begs the question. Moreover, even if intentions mattered only to blameworthiness, the one

who would find a conclusive reason to avoid acting blameworthy, even if he believed that it is permissible to act in that way, would have to reflect on his intentions before acting. Thomson's bomber might believe that bombing the enemy is permissible despite his intention being only to kill the children, but reject that intention wanting to avoid the risk of being blamed for it.

It might be replied that the OFO states that the oddness appears only when one considers intentions (or more precisely, facts about what one intends) as non-derivative reasons to forbid performing otherwise permissible actions. But if there is nothing odd in considering intentions as derivative reasons that affect permissibility, the oddness in considering them as non-derivative reasons has to do with them being such reasons rather than with the category mistake. The latter is at best an epiphenomenon of the former. I say at best because the intuitions about the oddness might differ and, as we shall see, there are cases where those who find considering intentions odd might intuit differently.

On the second charge, considering intentions makes no sense since, when one acts with a bad intention, one does not act for the sake of properties that render an action permissible but the action still has those properties even if one would change his intention (see Kamm 2004: 666-7, 2008: 168, 2014: 380, Rachels 1986: 92, cf. Hanser 2005: 461-4).

However, the defenders of the relevance of intentions hold that the intention-independent right-making properties of a permissible action are not the only ones that have an impact on its permissibility. Acting with a bad intention is itself a wrong-making property (cf. Chappell 2002: 226, Walen 2006: 42-3). Moreover, these people believe that such properties override the right-making ones, rendering the actions impermissible. So it makes sense, and it is indeed of utmost importance, to avoid acting with bad intentions if their view is true.

The last charge of the sort states that, when considering whether we may or may not act, we consider the properties of an action and only then decide about the goodness or badness of our intentions. It is not only that, as McMahan and Lippert-Rasmussen claim, the intentions with which we act are often affected by our judgments of permissibility. It is that we first determine the factors that affect the permissibility of actions and then, in virtue of them, decide about the goodness of our intentions, and not the other way around. To consider intentions before acting is to put the cart before the horse (see Kamm 2004: 667, cf. Scanlon 2008: 28).

In some cases, this objection works. We might be aware that having a certain intention will dispose us to perform an action which we can assess as impermissible irrespective of that intention and then conclude that, if and when, we would act in that way, the intention would be bad because it would lead to an impermissible action. However, since we can assess an action as permissible and yet think of the intention disposing us to do it as bad (e.g. giving to the needy in vainglory) as well as assess an action as impermissible and think of the intention disposing us to do it as bad but for an unrelated reason (e.g. killing non-combatants in war only for the sake of personal financial gains), the objection fails. This is not to say that, in such cases, we are morally required to consider intentions in order to determine the permissibility of our actions but that, if such a requirement existed, it would not be conceptually confounded.

3. Doing the Unnecessary

I now turn to views according to which the oddness is due to intentions being relatively morally unimportant considerations. The argument is:

- (1) If a consideration matters little or nothing morally, it is unnecessary and, for that reason, odd to consider or to advise others to consider it in decision making.
- (2) Intentions matter little or nothing morally.
- (3) Therefore, it is odd to consider or to advise others to consider intentions in decision making.

We saw above that the first premise is true on the examples of morally irrelevant considerations which are not odd because they are not usually taken into consideration but because they matter little or nothing in comparison with others. The second premise is where the rubber hits the road. However, since the OFO's aim is to justify the irrelevance of intentions to permissibility, the task of arguing for the second premise makes the OFO futile. If we explain the oddness appealing to the irrelevance of intentions, the OFO repeats what it has to prove. To many people's ears and in most cases, the oddness is, I believe, due to this reason.

3.1 Moral Self-Indulgence

Before I turn to my preferred explanation, let me rule out a view in its neighborhood. It has been argued that, faced with conclusive or sufficient reasons to act, it would be overly self-indulgent to pay attention to one's intentions before acting and to refuse to act if they are bad. Such a move would show an excessive concern for the purity of one's moral character or record – it would be, as one author said, “monstrously narcissistic” (Nye 2013: 271, cf. Kamm 1996: 249-51, Lippert-Rasmussen 2010: 560, n. 40, but see McMahan 2009a: 357).

Recall the stories from the beginning. If the bomber is hesitant whether to drop the bombs because he hates the enemy's children, he might decide to abort the mission because he cannot get rid of the intention to kill the children. But that would mean that the war prolongs. Similarly, following the advice to consider his intentions, the doctor aware of his hatred to the patient might leave the patient in anguish. According to the charge, such decisions are odd because the bomber and the doctor put the purity of their hearts above and before a pressing moral demand, namely the prevention of suffering.

But even if this is an odd prioritization, its oddness need not be due to self-indulgence. The justification of the constraint on intending harm might not be in the agent but in the victim (Quinn 1989: 348-51, Cavanaugh 2006: 147-58, Tadros 2011b: 130-2, *pace* Nagel 1986: 181-2, Wedgwood 2011a: 393). The constraint might arise because intending harm wrongs the victims by disrespecting them or by violating their status as autonomous beings. The bomber and the doctor might then claim that they were not self-indulgent but rather deliberated in those ways to avoid treating their victims disrespectfully.

There is another issue here. If I am concerned with personal moral purity and believe that intending harm or some other bad effect taints it, I will find acting in some otherwise permissible way more discomforting than the person who would not care less for it (cf. Williams 1981: 45-8). Placing too much emphasis on that feeling, to the extent of deciding to refrain from acting, might partially account for the oddness.

However, behind this lies a more fundamental explanation: revising bad intentions before acting is needed only if avoiding acting with them is more important than other considerations speaking in favor of acting. With respect to the mentioned cases, many people find that hard to

believe (Thomson 1991: 297-8, 1999: 514-6, Bennett 1995: 218, McIntyre 2001: 225-6, Enoch 2007: 81, Scanlon 2008: 20, 31, FitzPatrick 2012: 114, 117). This, it seems to me, is the key for the oddness or, more precisely, for the gut feeling of oddness that those people have.

3.2 Tracing the Oddness

On one view, it is odd to consider intentions in warfare but not to avoid direct discrimination (Kolodny 2011a: 119, Dworkin 2011: 470, n. 11). It does not seem odd to say to oneself: “Am I refusing to rent him the room because he has a criminal record or because he is black?” or to advise a colleague: “If you fire her because she is an incompetent worker, that’s OK. But if you do it because she refused to sleep with you, it’s not.”

However, this might be because the ordinary cases of discrimination are structurally quite different from Thomson’s cases. To make them more similar, we should make three assumptions. First, assume there is a conclusive reason to refuse to rent or to fire; say, the tenant is predictably dangerous while the worker is incompetent. Second, assume that the actions do not contribute to societal exclusions or deny the victims important social goods and opportunities nor they express the agents’s proneness to do it in future (this blocks the predictive significance of discriminatory intentions, cf. Scanlon 2008: 73, as well as societal role-based reasons not to take discriminatory reasons into account in public decision making such as hiring). Third, assume that the victims are not aware of the intentions with which they are treated (as that might be offensive to them and, thus, speak against acting, cf. Scanlon 2008: 72, Kamm 2014: 482-3). I do not claim that, under these conditions, direct discrimination is possible or that, if it is, it is permissible to do it but that

the oddness would remain for the same reason as above. The absence of the oddness in everyday instances of direct discrimination is due to the presence of those conditions, which are signified by the intention to discriminate.

For similar reasons, when the agents are members of just institutions in virtue of which they have professional duties to act or not to act in certain ways, considering intentions appears more odd than in the absence of the institutional role factor (cf. Liao 2012: 706). In examples typically used to illustrate the OFO, agents are members of just institutions. The pilot who drops the bombs intending only to kill the children is a member of armed forces waging a just war; the doctor who intends to kill the patient while performing terminal sedation is a qualified specialist working in a hospital, and so on.

There are least four reasons that might explain why people find it odd for professionals to consider intentions in decision making. First, since professional duties are by and large stringent, it might be thought that, if the duty to avoid acting with bad intentions exists, it cannot override them. Second, the former duties are not easy to comply with; they require attention, skills, and devotion. To require of professionals to pay attention to their intentions before carrying on with acting might be thought to require one thing too many. Third, professionals follow the orders and the protocols of their institutions, even when they act with bad intentions (cf. Hanser 2005: 457, 460). Since they rarely act without authorization, it might be said that the bulk of responsibility for their actions lies in their superiors and, consequently, that it makes no sense for the latter to deliberate about which intentions the former would act with (but it might make sense to inquire about their intentions before authorizing them to proceed if they could foresee that the difference in intentions would lead to the difference in consequences of acting). Fourth, since it is often

difficult to know one's real intention, the moral advice to act but not with a bad intention would not be helpful unless one was already motivated to avoid acting impermissibly. This is a general difficulty with moral advice but it becomes especially vivid if the policies specifying the content of professional duties were to depend on it (cf. Lippert-Rasmussen 2010: 560, n. 40.).

However, consider identical cases to those from above but devoid of the institutional role factor. In the first case, the person who drops the bombs is not a pilot but a well-skilled thug who has no connection to any regular army or military institution (Liao 2012: 719-20). In the second case, the person who administers the drug is the patient's enemy who finds himself at the right place at the right time.

No doubt that many will still find considering intentions odd. However, this is clearly not because intentions are unimportant but because none of the agents is authorized to act in these ways. Despite the opportunity to do much good by killing terminally ill patients at their requests, an unauthorized person sneaking into hospitals and doing it clearly acts impermissibly even if his intentions are the best. That is why it is odd to deliberate about it at all. Nonetheless, if we vary the examples such as to eliminate the context of war or hospital rules, we get the situations where the oddness still remains. And the only explanation of it, it seems to me, is that it is unnecessary to consider intentions due to their moral unimportance.

This conclusion is unfavorable to both sides of the debate. The defenders of the relevance of intentions might agree with it but then they must undertake the burden of proof. They have to show, on theoretical grounds, why the intuitive unimportance of intentions is not a good enough reason to think that they are non-derivatively irrelevant to permissibility. Or they might disagree with the intuition, saying that there is nothing odd in considering intentions, but then they owe us

an error theory of the opposing intuition. On the other hand, the critics owe us a different account of the oddness, avoiding those I criticized. Or they might embrace my explanation at the expense of discarding the OFO. It seems to me that the defenders are better off by diving into theory and the critics by discarding the OFO.¹⁰⁸

Conclusion

One of the main objections to the relevance of intentions to permissibility states that it would be odd to consider one's own intentions (or to advise others to do it) in order to decide whether it is permissible to act. I examined three kinds of explanation of that oddness, granting the intuition. I argued that the oddness is not due to the unordinariness of considering intentions, as that is not unordinary, and, even if it were, it could not explain it, nor that it is due to an impossibility or a conceptual confusion, as none of the suggested ways how that could be true works. I then argued that the oddness is best explained by the intuition that intentions are intuitively relatively morally unimportant, at least in Thomson's classic cases. As the defenders of the relevance of intentions share that intuition, it is likely that they are taken by it in positing the OFO. If so, the OFO is not an objection proper but just another way of restating their view.

¹⁰⁸ The oddness might not appear in considering intentions when about to perform merely permissible actions. It might be said that, if there are *pro tanto* reasons against acting with bad intentions, they tip the balance, rendering merely permissible actions impermissible (Arneson 2006: 782-3). It might be similarly suggested that, while it is odd to consider intentions if the stakes are high, it is not odd to do it when there is, intention-independently, as much reason to act as not to act. The plausibility of that suggestion depends on whether acting with a bad intention is a *pro tanto* reason against an action's permissibility. I cannot discuss that issue here.

PART THREE

Towards the Irrelevance Thesis

9. Bad Intentions as Wrongs *Per Se*

Introduction

Are bad intentions wrongs *per se*? In other words, are there reasons against intending harm and other bad effects that are not derived from reasons against harming or bringing those effects?¹⁰⁹

There are three ways to answer these questions. We might give an argument to the point that reasons against intending harm are impossible. One influential argument reaches that verdict from the impossibility of choosing intentions with which we act (Ross 2002: 5, Bennett 1995: 194-6, Scanlon 2008: 58-62). Alternatively, we might accept the possibility of reasons against intending harm and go for the case-based dialectic, assessing cases where those reasons are said to explain why an action is impermissible. This is the most popular strategy in the debate about the relevance of intentions to permissibility (see e.g. Bennett 1995: 214-21, Scanlon 2008: 12-36, Tadros 2011b: 149-60, Nelkin and Rickless 2014: 134-40, 144-7). Finally, we might engage with justifications of reasons against intending harm as reasons affecting the permissibility of actions. This is what I shall do in this chapter, arguing against nine prominent justifications. Their failure

¹⁰⁹ I am concerned with intending harm which results in harm, that is, with act-types of harming with intent to harm. I thus use intending harm and intended harming, and reasons against them, as synonyms. I also use bad intention to mean any intention whose content is a bad effect, like harm.

is not a decisive evidence to conclude that there are no reasons against intending harm—for that we need the convergence of the other strategies—but it is a sufficient evidence for serious doubt.

Reasons against intending harm matter because they ground the deontological constraint on intending harm. Since that constraint is an essential part of the Doctrine of Double Effect and other moral principles stating the non-derivative relevance of bad intentions to permissibility, the principles make sense only if there are reasons against intending harm and they are of overriding strength (Kagan 1989: 165-6). I leave the latter condition aside here. I am interested only in what could justify them as *pro tanto* reasons.

Any plausible justification must satisfy the following desideratum. The facts that would give rise to reasons against intending harm might give rise to reasons against merely foreseeing it. But since these principles state that it is worse to intend than to merely foresee harm (or that it is harder to justify it), to the extent that some of them invariably prohibit the former, the facts that would give rise to reasons against intending harm must be morally special. Intending harm must be objectionable over and above merely foreseeing it (Kagan 1989: 166, Quinn 1989: 348).

I evaluate instances of three general justifications of reasons against acting in some way: agent-centered, victim-centered, and impersonal. With regard to the issue in question, they state, respectively, that bad intentions are wrongs because they are bad for the agent, or for the victim, or because they are bad, period (McMahan 2009a: 357-8). In what follows, I reject in turn the views of intending harm as corrupting agency and moral character, betraying integrity, showing

an attraction to evil, disrespecting and using the victim, violating his inviolability and autonomy, and being impersonally bad.¹¹⁰

1. Agent-Centered Views

On the agent-centered views, deontological constraints are grounded in some morally relevant feature of the agent or the way *he* relates to his action or to the victim. Therefore, it is sometimes said that agent-centered views derive deontological constraints “from inside the agent outward”, as opposed to victim-centered views, which derive them “from outside the agent inward” (Kamm 2007: 26, cf. Darwall 1986, Anderson 1993: 73-9).

1.1 Agency

When some people defend reasons against intending harm, they appeal to the fact that intending harm involves greater or more intimate agency. Wedgwood (2011a: 393, 2009a: 335-6) claims that, when an act has a bad consequence, “the more agentially involved you are in bringing about that consequence, the stronger the reason against the act will be”, and it will be much stronger, he adds, if you intend than if you merely foresee that consequence. On the same note, it is argued that acting with intention to harm has a greater degree of “actionhood” than acting with a mere foresight to harm (Chappell 2002: 224-5), that we are more closely “engaged with the world as

¹¹⁰ In doing so, I repeat many objections raised almost three decades ago by Kagan (1989: 165-82), modifying and improving them in certain ways. Attractive but flawed ideas do not go out fashion easily, as Kagan was well aware of. The epigraph of his book is a quote from Gide: “Everything that needs to be said has already been said. But since no one was listening, everything must be said again.”

agents” when we intend harm (Duff 1990: 113), and that “we are the authors (...) in the way we are not when we merely anticipate [harm]” (Moore 1997: 409).¹¹¹

However, when we foreseeably cause or allow harm, we are agentially involved as well. We act no less voluntarily and intentionally. And we could not foreseeably cause harm unless we directed our actions towards it. It then remains unclear why agential involvement or actionhood is greater, in the sense of being more morally objectionable, when we intend harm. Two kinds of response present themselves to those attracted to the agent-centered justification: the one that traces the offense to some property of the agent, such as the corruption of his moral character or the betrayal of his integrity, and another locating it in his relation to some morally salient feature of intending harm. I discuss them in turn.

1.2 Moral Character and Integrity

On some views (e.g. Boyle 1980: 536-7, Masek 2010: 581-5), intending harm impairs moral character. When we intend harm, we become worse people and our actions are *pro tanto* wrong. We then have reason to attend to ourselves in acting, being vigilant not to stain our characters with bad intentions.

An obvious problem with this explanation is that it passes the buck. We are left with the question why intending harm corrupts the character more than merely foreseeing it. Defenders of this view can help themselves with Nagel’s explanation of wrongness of intending harm (his own view is different; I discuss it below). For Nagel (1986: 181-3), to intend harm is to have a will

¹¹¹ These authors also claim that the degree of agency is greater in actively causing harm than in failing to prevent it (see Wedgwood 2009a: 334-5, 2011: 392-3, Chappell 2002: 215, Moore 1997: 409).

guided by evil. Since evil is something to be repelled by and not to be embraced, when we intend harm, we “swim head-on against the normative current” (1986: 182). We are guided by what we have reason to avoid. A similar rationale is that, when we intend some effect, we thereby commit to achieve it. That fact distinguishes intending from merely foreseeing, not just conceptually but morally. Volitional commitment to harm makes it especially bad to intend it (Cavanaugh 2006: 134-47).

However, even if those explanations are plausible, they render the character-impairment view obsolete. If intending harm is wrong over and above foreseeing it because one is guided by or attracted to evil, the fact that that makes one a worse person adds nothing to the wrongness of intending harm.

There are further problems. This view goes against the common picture of the division of labour between reasons not to act in certain ways and reasons not to be certain kind of people. Permissibility and character are usually thought of as separate moral categories, the former never being grounded on the latter. One can act permissibly and yet be wholly vicious. One can reply that the implications need not be as revisionary. The view might state only that there is a reason not to act in ways that corrupt the character but not that that reason is decisive. But this would not avoid the objection, for the following reason.

Many states of mind can be said to corrupt the character: some false moral beliefs, such as racist or sexist beliefs, some inadequate desires or emotions, such as hatred or revengefulness, etc. Even if we have reasons to work on changing or eliminating them for the sake of moral self-improvement, it is too far of a stretch to say that those reasons count against the permissibility of acting, even if just *pro tanto*. Consider an employer who disciplined himself not to be motivated

by sexist thoughts but who cannot get rid of them for good. On this view, his hiring decisions, even when perfectly just, are always *pro tanto* wrong. Moreover, since no one is immaculate in all thoughts, desires, and emotions, this view implies that we always act *pro tanto* wrongly. This cannot be right. It seems more likely that reasons not to cultivate certain motives or intentions, like reasons not to form certain beliefs or respond with certain emotions, do not ground reasons against acting, although they have a big influence on a person's character.

Exceptions might be cases where having a bad intention would make one undermine his character as a whole, or betray his integrity, as in Williams's (1973: 97-9) case of a chemist and committed pacifist, whose only option to provide for his family is to accept a job in a factory that produces chemical weapons.¹¹² Assuming that he will intend to be part of the team that produces the weapons, were he to accept the job, and that there is a strong reason for him to accept it, one might argue that his commitment to pacifism defines who he is and gives him a reason, perhaps even a decisive one, not to accept the job. In fact, on some views (Fried 1978: 2, Darwall 1986: 306-17), deontological constraints in general stem from the prohibition on betraying integrity, i.e. from the reason not to violate the principles one reflectively accepts.

Three things can be said in reply. First, upholding integrity is consistent with intending harm. A judge is no less righteous if he intends a criminal to suffer when sentencing him to years in prison. He need not think that the criminal deserves to suffer but that he will not be taught a lesson unless he does.¹¹³ Second, if reasons against intending harm are based on reasons against betraying integrity, the former are in charge much less frequently than one might hope. Losing

¹¹² Integrity is understood here not as a character trait but as a self-conscious commitment to goals and projects that make up one's practical identity (Williams 1981: 49, cf. Finnis 2011a: 194-6).

¹¹³ Compare him with a judge who intends a defendant to suffer, period, due to personal animosity (Sidgwick 1981: 202). He would compromise his integrity as a judge.

integrity is not an everyday episode; intending harm is. Third, even if betraying integrity counts against acting with intent to harm, there is nothing about it that counts exclusively against acting in such a way. If Williams's chemist has a reason to refuse the job offer in order to preserve his integrity, he has it whether he would intend or foresee the harm by taking the job. It is the nature of the job that would compromise his ideals, were he to accept it, not his attitude towards it.¹¹⁴

Those attracted to the agent-centered justification might then turn from intending harm as corrupting the agent to corrupting his relation to the relevant normative features of that intention.

1.3 Guidance by Evil

As we saw, Nagel claims that to intend harm, even as a means, is to be guided by evil. What does it mean to be *guided by evil*? And why is it morally objectionable?

We should first note that Nagel (1986: 164-5, 176) thinks that deontological constraints, and reasons that ground them, are not just agent-centered but agent-relative. They have essential reference to the agent they constrain. Agent-relative reasons are reasons that *I* do not do or intend something. This is not to say that no one but me would have an agent-relative reason in the same situation. Neither is it to say that everyone else cannot have reasons to hope that my actions or intentions do not come about or reasons to prevent them; a reason can be both agent-relative and agent-neutral. It only means that it is not possible to reduce deontological reasons to impersonal ones (Nagel 1986: 177).

¹¹⁴ Kagan (1989: 31) argues that the appeal to integrity cannot justify the constraint against doing (and, by extension, intending) harm but only an option to refrain from it because, if one is *willing* to do (or intend) harm, nothing can stop him, morally speaking. This is a misunderstanding. If reasons not to betray integrity do exist, one cannot release oneself from their demands at will, just like one cannot release oneself from other-regarding reasons at will.

This helps Nagel to unravel the phenomenon of two viewpoints that each person can take: from within and from outside himself. We can see ourselves as doing certain things and see those things as happening. When we intend harm, it is the former, personal, standpoint we occupy that helps us acknowledge the reasons to abandon that intention. If we only occupied the impersonal standpoint, our sole concern would be to make things go best, regardless of means and attitudes.

Reasons not to cause harm intentionally, Nagel (1986: 180) writes, are “magnified and lit up from [our own] point of view”. Nothing of the sort, at least not to that extent, occurs when we foresee causing or allowing harm. It is unclear what the metaphor of magnification stands for. Perhaps it is a feature of common phenomenology, although I doubt that we are sensitive to such refined deontological distinctions. My own view from within is bland in that respect.¹¹⁵

Nagel (1986: 180) also talks about the special responsibility we have not to intend harm. Whatever he means by responsibility—aptness for other’s reactive attitudes, greater debit on our moral ledger, special answerability, etc.—it is irrelevant to our issue, for the same reason the fact that intending harm affects our moral character is irrelevant. These categories are best kept apart as we have no good reason to think that one grounds the other (Scanlon 2008: 52-6).

We are then left with guidance by evil as a purportedly morally relevant fact. Let us try to unpack the notion of evil first. Evil cannot be a wrong since, if intending harm is a wrong *per se* because one is guided by that wrong, the account is circular (McCarthy 2002: 627, Nelkin and Rickless 2014: 132: n. 14). And if it is a wrong because one is guided by some other, more basic, wrong, the account is self-refuting. It gives up on bad intentions as wrongs *per se*.

¹¹⁵ Wedgwood (2009a: 336) uses the magnification metaphor to point out the greater agential contribution of intending harm to the badness of the course of that action. He compares a person dying prematurely from natural causes from him being murdered. But clearly the difference is not due to the intention to kill; we can murder foreseeably.

Therefore, evil must be something (very) bad. The badness cannot be deontic, for similar reasons to those just given. If the evil in intending harm is deontically bad, it is bad because the wrongness of acting with intention to harm adds up to the badness of what happens. But the issue is whether acting with that intention is a wrong *per se* in the first place. Evil is then non-deontic badness.

Non-deontic badness can be intrinsic or extrinsic. Something is intrinsically bad if it is bad solely in virtue of its intrinsic properties, and extrinsically if it is bad derivatively, in virtue of its relation to other extrinsically or intrinsically bad things. On this view, suffering, death, and other misfortunes, are intrinsically bad only because of some additional attendant circumstances, e.g. that someone dislikes to suffer, that he wishes to go on living, that death would deprive him of valuable experiences, and so on (Wedgwood 2011a: 395). As Wedgwood (2011a: 396) notes, we rarely intend intrinsic badness, so defined. Even murderers typically do not intend to deprive their victims of valuable life experiences, in addition to intending to kill them. Therefore, in most cases, the evil in intending harm is that of extrinsic badness.

We can now turn to the notion of guidance. Nagel (1986: 181-2) claims that, to be guided by evil, means to “be prepared to adjust [one’s action] to ensure the production of evil: a falling-off in the level of the desired evil becomes a reason for altering what one does so that the evil is restored or maintained”. Some people (e.g. Nelkin and Rickless 2014: 132, n. 14) read this as a way of saying that one intends harm. If so, the account is empty of useful content and it repeats what it has to explain in fanciful terms. This is a very uncharitable reading.

On a better, but still mistaken reading (Wedgwood 2009a: 335), to be guided by evil in intending harm is to invest more thought and effort in bringing it about than when foreseeing it.

But surely, I need not spend more time thinking when I am to harm someone intentionally than foreseeably, nor need I find it harder to do it. And even if I did, an extra thought or drop of sweat would not make a moral difference.

What Nagel means by guidance, I believe, is a kind of commitment. To be guided by evil is to commit oneself to bring it about and to do what is needed for that goal to ensue, such as to take necessary means to it, to avoid steps that would obstruct or prevent it, perhaps even to take recurrent action if the initial one fails. This is a very influential account of the difference between intending and merely foreseeing harm (see Bratman 1987: Chs. 3, 10, Cavanaugh 2006: 134-47) and Nagel, as I understand him, is that account's forerunner (cf. Kagan 1989: 166-7).¹¹⁶

To commit oneself to bring evil, even only as a means and with a heavy heart, entails that one is attracted by it, Nagel (1986: 182) thinks. However, since the nature of evil is to be repelled by it, to have that attitude is to go against the moral grain. This is where the inappropriateness of intending harm lies.

Are we really guided by evil only when we intend harm? And is that so objectionable? I believe that the answer is "No" to both. Being guided by evil is not characteristic of intending harm. We are often guided by it when we merely foresee harm, whether as an unavoidable side effect of an end we pursue or as a means to it. And there is often no moral objection to it.

Consider first morally justified foreseen but unintended harming as a means of benefiting ourselves. In *Self-Defense*, a culpable attacker tries to kill me but I get him first. I intend to save my life, merely foreseeing his death. I want to be sure that the attacker is dead as I know that he

¹¹⁶ The idea of committing to evil as a sign of moral perversion is of course much older. Milton's Satan cries out "Evil be thou my Good" and Goethe's Mephistopheles reveals himself as "a part of that power which eternally wills evil".

has a destructive arsenal and is willing to use every bit of it to kill me. So I go on shooting until I am sure I killed him. I am guided by his death, committed to bring it about. But I do not intend it, nor am I attracted by it. I have no hard feelings towards the attacker and I would not object if he survived and got arrested before I kill him. I just know that, in all probability, that is not going to happen. Clearly, I have all the right to shoot my attacker dead.

Consider next justified merely foreseen harming of one person as a means of benefiting that very person. In *Chemotherapy*, a doctor prescribes chemotherapy to a patient with cancer, knowing that that will damage his organs but that the damage is causally necessary to get rid of the cancer cells. The doctor need not intend the damage. Apart from intending to cure the patient, he might intend to affect his organs, foreseeing the damage as a regrettable but necessary means of curing him. However, it is very likely that he will be guided by the damage as a sign that the treatment is going well. It would be bad news for the doctor if, after several sessions, no damage occurred. He might then go on with chemotherapy until enough harm is done. Suppose that, after many tries, he realizes that the patient is getting cured though his organs remain unscathed. That would be good news, indeed a miracle. Only then would the doctor not welcome the damage or be guided by it. Still, it does not follow that he would wrong the patient before the discovery (cf. Kagan 1989: 168, Parfit 2015b: 10-1).

Consider finally justified merely foreseen harming of some as a side effect of doing good to others. In *Cloudy Bombing*, a bomber waging war on the just side intends to raid the enemy's munitions factory in order to bring the war to an end and save millions but foreseeing that he will kill some non-combatants as a collateral. Since the factory is covered with thick clouds, his only sign that the mission was successful is the sight of dead non-combatants. He keeps bombing until

he sees them dead (Kamm 1996: 254, n. 10, 2007: 177, n. 10). He is committed to kill them and guided by their deaths but he does not intend to kill them. Nor does he wrong them, at least not over and above the wrong of foreseeably killing them.

There are, I believe, instances of justified foreseen harming of some as a means of doing good to others, but since they are controversial we need not discuss them here. One might claim that harm must be intended in all the cases above. This is hardest to believe in *Cloudy Bombing*. The most we can say about the bomber, I believe, is that he aims at non-combatant deaths. This is why some people think that he intends it. But not every aiming is intending (cf. Kamm 2007: 124, n. 6). The sense in which the bomber aims at non-combatant deaths is minimal; it means no more than that he pursues it. In that sense, merely foreseen causing an effect is aiming at it too.

I do not know the necessary and sufficient conditions for intending. Intending might even be a family resemblance concept. However, it seems to me that the ordinary meaning of the term is consistent with the judgment that in none of the above cases harm is intended. If so, volitional commitment or guidance does not imply intending, as that term is ordinarily used.

Nonetheless, intending might be a philosopher's term of art. Nagel, among others, might take aiming and intending as synonyms (cf. 1986: 182-3). In *Self-Defense* and *Chemotherapy*, harm would then be intended as a means to good ends. However, since there is nothing bad about aiming at evil as a means to justified self-defense or to curing sick people, it follows that there is nothing bad about intending evil as a means to such good ends. Nagel (1986: 182-3) himself raises this objection but fails to reply to it.

He could have replied in two ways. First, he could have said that his account is limited to intending harm to an innocent person, who does not consent to it, as a means of benefiting other

innocent persons. Evil might be only a variant of extrinsic badness, that of instrumental badness for one to benefit others. This is suggested by Nagel's (1986: 176) example in which, to drive my injured friends to the hospital, I must twist some child's arm, since only his cries would persuade his terrified grandmother to give me the car keys.

The situation is not more horrible because I intend the child's pain as a means to my end. Being an experienced torturer, I might twist his arm habitually, with the sole intent of getting the information about the keys. I might also do it as a side effect of getting the grandmother to give me the keys, say, by pressing a switch that will turn on an audio message with child screams so as to deceive her, but foreseeing that that will also activate a lever twisting the child's arm a bit later. I do not think that the degree of wronging the child in the original example is different from these other versions. The prohibition on twisting his arm is located elsewhere.

Moreover, as Kagan (1989: 169-70) observes, if we limit the guidance by evil rationale to cases of harming as a means, we give up on that rationale. We instead rely on the assumption that what is so offensive about intending harm is the nature of the relation between the agent and his victim and not the agent and his intention.

A better reply is that intention to harm, even as a means, exhibits an attraction to evil that mere foresight of harm, and even aiming at it, do not. The essence of objection to intending harm is attraction by what should repel us, and not the fact that we aim at evil or that we are guided by it. If the doctor in *Chemotherapy* intends to harm the patient's organs as a means of curing him, rather than just to affect or use them, he will likely be dissatisfied to hear that chemotherapy had no negative effect on the organs. He would not be much different from a sadist doctor who would

not care about curing the patient and who would prescribe chemotherapy only in order to damage his organs as an end. The same is true of other cases.

It is important to bear in mind here the difference between the question whether attraction to an effect, in our case harm, distinguishes intending from merely foreseeing it, and the question whether that attraction is the marker of moral difference between them. The former, conceptual, difference might as well lie in the attitude. This is what a version of the Counterfactual Test tells us: we intend an effect if we would not be glad were we to act and the effect did not come about.

However, we have no reason to think that attraction to harm makes intending it singularly objectionable (cf. Kagan 1989: 172). It seems more natural to take that attraction as a character defect rather than as a non-derivative reason not to harm. On some views, however, bad motives, like revenge or hatred, are intrinsically bad. Intending harm might be similarly intrinsically bad and the reason of its badness might be attraction to evil. I do not believe that this can differentiate morally intending from merely foreseeing harm. I shall discuss such views in the last section.

There is another objection to the agent-centered views, namely that they are focused on the agent whereas the source of the complaint naturally comes from the victim. Since the victim-centered justification has tradition on its side, the agent-centered justification is often said to be revisionary, let alone self-indulgent (Kamm 1996: 249-51, Lippert-Rasmussen 2010: 560, Nelkin and Rickless 2014: 132, n. 14) and narcissistic (Nye 2013: 271).

While I see the intuitive force of this complaint, I do not think that it is decisive. Its best target is the agent-centered view of reasons against intending harm combined with the claim that those reasons are as strong as to outweigh competing reasons that strongly favor acting, such as reasons to save innocent people from death. It might be rightly objected that that view has an odd

self-indulgent implication that the agent is to be concerned with his bad intention when the lives of others are in danger and he could help at no cost. It would be a mistake to think that he has a duty to try to change his intention at the expense of risking to fail to save those people. The same objection is less persuasive against the view that intending harm is *pro tanto* wrong because of some agent-centered fact. In other words, the rescuer might have a very weak reason to consider his intention apart from the claims of those in need, a reason that soon gets lost under the weight of those claims. The objections I gave are better suited to tackle that other view.¹¹⁷

2. Victim-Centered Views

Another strain of justification is victim-centered. It derives the objection to intending harm from the way it wrongs the victim. Victim-centered views, almost as a rule, appeal to some Kantian insight, such as disrespect, violation of autonomy, using as a means, etc.

2.1 Respect

Quinn (1989: 348) claims that those who harm innocent, non-threatening, persons with intentions to harm them or to involve them in foreseen harming in order to benefit other such people “show a shocking failure of respect for the persons who are harmed; they treat their victims as they would treat laboratory animals”. Following Quinn, many other people write that intending harm “does not adequately respect one as a human being” (20: 163), that it “shows a lack of respect”

¹¹⁷ McMahan (2009a: 357) claims that deontological morality is essentially agent-centered but what he means but it is that it is essentially agent-relative (“We are (...) primarily responsible for our own action, and our own intentions”).

and even “disrespect” (Liao 2012: 723-4, Nelkin and Rickless 2014: 149, 150). What they mean by respect is recognition respect, that is, respect for a person on the basis of morally salient facts that ground his moral status, e.g. his rational nature or dignity (Darwall 1977).

It is perfectly natural to characterize some instances of intended harming as lacking that form of respect. A sadistic doctor who prescribes chemotherapy to a patient only in order to see him suffering is disrespecting the patient. A bomber who targets non-combatants as a means of demoralizing and forcing the enemy to surrender shows little respect for non-combatant victims. It is, however, less obvious whether the harm inflicted intentionally is, other things being equal, more disrespectful than the harm inflicted merely foreseeably and, if it is, whether that accounts for the supposed difference in wrongness between them. This touches on the issue of justificatory role of respect in general: whether disrespect is a fundamental reason why an action is *pro tanto* wrong.

As Quinn (1989: 348, cf. Tadros 2011b: 125) admits, the appeal to respect alone will not suffice. Some instances of merely foreseen harming are disrespectful. He mentions a case where a strategic bomber raids an unimportant enemy facility, foreseeing non-combatant casualties as an unavoidable side effect. But if the bomber disrespects non-combatants, it is because he does not have sufficient justification for the raid. The same is true, I believe, of the bomber who raids non-combatants as a means of demoralizing the enemy. This suggests that in many, perhaps most cases, lack of respect and disrespect are epiphenomena of more basic facts rendering an action wrong (Kagan 1986: 175-6, Scanlon 2008: 223, n. 18). Therefore, even if intended harming is more disrespectful than merely foreseen harming and even if wrongness tracks disrespectfulness,

the former's greater wrongness requires explanation in terms of some morally relevant fact other than disrespect.

Moreover, the assumption that intended harming is more disrespectful rests on thin ice. In *Chemotherapy*, the doctor does not seem to disrespect the patient whose organs he intends to harmfully affect. Conversely, in Nagel's child-torturing case, I do disrespect the child by twisting his arm against his will, regardless of whether I do it with intention to coerce his grandmother to give me the car keys or habitually, as an experienced torturer, concerned only for the well-being of my injured friends.

It can be replied that intending harm as an end is always disrespectful. That might be true but, as I have said, it does not follow that harming with that intention is wrong because it is disrespectful, not even that it is *pro tanto* wrong. Cases where one has sufficient reason to harm the innocent, does it disrespectfully by intending to harm him as an end, and yet his action is not *pro tanto* wrong (because, say, the harm is a side effect of a greater good) are a possibility worth discussing. For instance, if I redirect a runaway trolley away from the main track where it would otherwise kill five people onto a side track where my foe lies, just in order to kill my foe, I might not act *pro tanto* wrongly, but I act disrespectfully, towards my foe and the five (cf. Kamm 2007: 132-3). What is more, if intending harm as an end is always disrespectful but intending harm as means is not, and might even be respectful, the respect view is at best incomplete.

Therefore, the appeal to respect is unhelpful either because more basic factors are in play or because some actual or seemingly disrespectful harmful actions are not *pro tanto* wrong.¹¹⁸

¹¹⁸ A disrespectful action, even if otherwise permissible, often impairs one's relation with others and gives them reasons to blame him (Scanlon 2008: 115-7). This might partly explain why some people are drawn to disrespect as a ground of reasons against intending harm but I doubt that it is what drives the authors I cited.

2.2 Inviolability

Another idea is that intended harming, especially as a means to a greater good, is wrong because it violates our moral status as rational beings. On one popular version of that idea, moral status is inviolability. Each of us has a right that constrains others's behavior, limiting what they can do to him in pursuit of the impersonally best outcomes. Since we are rational beings of high cognitive capacities, we are not violable for the sake of maximizing the good (Kamm 1992, 1995, 1996: 271-82, 2007: 26-30, 253-6, 268-75, Nagel 2008: 111-3, Tadros 2011b: 124-7, Nozick 1974: 31).

Inviolability is an agent-neutral value, but it is a value to be protected, not promoted. As Kamm (2007: 29) puts it: "Each agent must protect this value and does so in being constrained by the rights of the first person he encounters. (...) The value resides *in* persons and we act in the light of it." Inviolability is not absolute. It might be overridden by a much greater good. It is also a matter of degree. Human persons have a more elevated moral status than non-human animals (some of which might be persons).

One reason to suppose that we are inviolable is instrumental. It might be better for us to live in a world in which we are inviolable (Kamm 2007: 270-1). Kamm thinks that inviolability is as valuable as to make it the case that a world in which it is respected is better for us than a world were it is not so respected even if we are more harmed or at greater risk of being harmed in the former. "We may all lead harder lives", she (1996: 294) writes, "but our dignity is greater". But that we would be better off as inviolable does not imply that we are inviolable. Most of us would be better off if we were millionaires.

Moreover, as Kamm (2007: 29-30) notes, from the *ex ante* perspective (e.g. the veil of ignorance), it might be in our interest to opt for a society in which inviolability is occasionally violated for the greater good. However, we cannot bargain away our inviolability, Kamm (2007: 30, 254) thinks, so she concludes that its justification must lie in our status as rational beings. It is simply in virtue of who we, namely persons, that we are inviolable. This might sound as a cop out but I shall grant it for the sake of the argument.

The problem with the appeal to inviolability is, in one way, the same as the problem with the appeal to respect. Even if we grant that rationality endows us with inviolability and that that feature is the ground of deontological constraints, we are left with the question why inviolability is sensitive exactly to deontological constraints we think it is sensitive to. Our inviolability might be violated to a greater extent when we are harmed intentionally than foreseeably, or as a means than as a side effect, but it might as well be the other way around, or it might be that inviolability is violated to the same extent in each of those cases. There is no reason to think that inviolability constrains intended harming stricter than merely foreseen one, nor that it constrains harming as a means stricter than harming as a side effect. Like the notion of respect, inviolability, by itself, is of no use in justifying deontological boundaries.

It might be replied that the greatest violation of inviolability is harming intentionally as a means to a greater good because that compels one to serve others's ends against his will (Tadros 2011b: 127, 136). This is a legitimate reply but it turns the justificatory focus from inviolability to the prohibition on harmfully forcing a person to serve optimific ends. It is not inviolability but autonomy that does the work.

2.3 Autonomy

Tadros (2011b: 136) writes: “When a person is harmed as a means to an end that person is also required to bear a cost for the sake of that end. His life, or his autonomy, is sacrificed for the sake of that end. It is wrong to require a person to sacrifice his life or his autonomy for the sake of an end that he need not adopt as his end at that cost.” Tadros (2011b: 14) believes that harmful using as a means is necessarily intended. The objection is then not just to the act-type of harming as a means, but to the act-type of harming as a means with intent to harm in that way. Presumably, the rationale extends to intended harming as an end and harming as an end.

Autonomy matters for the same reason inviolability does. Since we are rational creatures, we can choose how to live our lives according to our best judgment. We can choose which ends to adopt and which to refuse; we are “independent end-setters” (Tadros 2011b: 127, cf. 130, 2015: 65). Autonomy is sometimes overridden by more important factors. We may be compelled, even harmfully, to serve an end. In Tadros’s terminology, we sometimes have enforceable duties to serve ends and “there will often be nothing wrong with compelling the person to bear that cost for the sake of that end” (Tadros 2011b: 136). But *prima facie*, we do not have a duty to serve the greater good, apart from serving it voluntarily (cf. Quinn 1989: 349).

Again, the main question is why autonomy does not provide us with the right to refuse to be harmed as a merely foreseen side effect and, if it does, why that right is not as stringent as the right to refuse to be harmed intentionally as a means to an end. Tadros’s (2011b: 136-7, 2015a: 67) reply is that harming as a side effect of a greater good does not compel one to adopt the good

as his end. On the ordinary meaning of ‘compel’, this is plainly false. To compel is to force one to do something, or as Tadros puts it, to adopt an end, against that person’s will. It is not limited to forcing to adopt it as a means.

It can be replied that we cannot be compelled to *serve* an end as a side effect and that this is what we do when we are harmed as a means (cf. Nelkin and Rickless 2014: 133). But there might be a sense in which we serve an end when we are harmed as a foreseen side-effect to it; we serve it as its side-effects. What is of greater importance, however, is that it is question-begging to defend the special wrongness of intended harming as a means appealing to verbal substitutes for the causal relation of ‘being a necessary means’ (like ‘to serve’). It is to restate, in an ornate style, what has to be proven: that intended harming as a causally necessary means is a wrong and merely foreseen harming as a side effect is not, or that the former is a greater wrong than the latter. This is what Tadros does when he says (2011: 137) that we are not compelled to adopt an end when we are harmed as a side effect because we do not *contribute* to the pursuit of that end. But to contribute to an end means, roughly, to be a causally necessary means to it.^{119, 120}

It might be thought that Tadros has another reason to count intended harming as a means to an end as a wrong since he might sometimes be interpreted as saying that a person harmed as a means is wronged not only when he has no enforceable duty to serve an end as a means but also because of that duty (Tadros 2011b: 129). If we do not have that duty because, *qua* autonomous beings, we enjoy, as Quinn (1989: 351) puts it, a veto power over certain ends, we are back to the

¹¹⁹ Contributing might also mean being causally sufficient but not necessary. I take it that most people who hold that intended harming as a causally necessary means is a wrong also hold that that other causal means are wrong as well, and to more or less the same extent (cf. Kagan 1989: 140-1, on ‘weak’ and ‘strong’ means).

¹²⁰ Tadros (2011b: 137) similarly begs the question when he explains the wrongness of harming an innocent person by pushing him off a bridge as a means of saving five others saying that “[w]e could rescue the five *only by* requiring the one to participate in pursuit of a goal that he is not required to pursue” (my emphasis).

same problems. And if there are other reasons not to have an enforceable duty to be harmed as a means, everything turns on working out those reasons.

2.4 Using as a Means

The remaining option is to posit using as a means for one's purposes as the ground of reasons against intending harm. The idea seems to lie behind Quinn's claim that intentionally harming as a means uses victims as if they were guinea pigs. Many other authors make similar remarks. For instance, Tadros (2011b: 124) claims: "If the person uses me as a means to pursue their goal, I am treated like an object. Objects, but not humans, may be used to pursue goals." And Nelkin and Rickless (2014: 149, 150) write: "Harmful use of victims infringes their right not to be so used. (...) The fact that the civilians have not consented [to be killed as a side effect of strategic bombing] is morally relevant. But it is also important that they are not used."

This view is incomplete in at least three ways. First, it cannot account for intending harm as an end (Kagan 1989: 179). True, intentions to harm as an end are sometimes covert intentions to harm as a means to another end. A sadist might intend not only that his victims are in pain but that they are humiliated as a result of being in pain. There is a sense in which he uses them as a means. But many wrongdoers do not intend to harm their victims in order to make them suffer or to deprive them of valuable experiences, apart from intending to harm them in particular ways.

It might be replied that we nevertheless always intend to act so as to achieve an objective. To intend harm as an end is actually to intend harm as a means to achieve an objective, such as to revenge against someone or to make him suffer (Nelkin and Rickless 2014: 148). However, if by

‘objective’ we mean an end, we can intend an end without intending it as a means, as I have just said. A possible reply is that the core objective of every action is self-interest. But even if this is true, it only shows that self-interest always motivates us and not that we need to intend to act as a means of satisfying it. When I intend to exercise, I do not need to intend to exercise in order to stay fit, even if staying fit is my self-interested motive for exercising.

Second, the view cannot account for wrongful intentional causing of harm to a person, or harmfully affecting him, in ways that do not harmfully use him. On the ordinary sense of ‘use’, we use a person if we want his presence to help us achieve our goals. However, other people can be obstacles to our actions. Our goals would be best achieved had they not been where they are (Parfit 2011a: 221-2, Persson 2013: 141, cf. Kagan 1989: 140-1). In some such cases, as in self-defense from culpable attackers, harming is permissible. But in many others, it is not. If I blow up one person blocking the mouth of a cave as a means of rescuing several other trapped persons, I do not use him but I do unjustifiably sacrifice him for that goal (Foot 2002a: 21).¹²¹

Third, in omitting to act, by definition, we do not act, so we cannot use anyone. But to omit to act is often harmful and wrongful. Suppose a child is drowning in a shallow pond and I am the only person who can rescue him, at no cost to myself. If I walk on, intentionally leaving him to drown, I wrong him and act impermissibly but I do not use him.¹²²

These problems are secondary to the problem of justification. The charge, “You were just using me!” is common in everyday speech but it is questionable whether it has moral substance apart from being a trope for expressing more fundamental moral concerns. It is often not wrong,

¹²¹ This leads Tadros (2015a) to supplement the prohibition on intended harmful using as a means with the prohibition on intended harmful affecting as a means. My later objections target the affecting view too. (Tadros justifies both views appealing primarily to considerations of autonomy.)

¹²² In Kant’s sense of ‘use’ (*brauchst*), I do. But that is a rather technical sense of ‘use’ or ‘treat’.

not even *pro tanto*, to use people as means or tools. We may use others as watches, asking for the time, or as dictionaries, asking for the meaning of a word, or as parasols, staying in their shadow to protect ourselves from the sun. As Scanlon (2008: 107) notes, if using others as a means were morally objectionable as such, any cooperative activity would be *pro tanto* wrong.

It might be said that what is objectionable about intended harming as a means is not using but harmful using without consent. To avoid the difficulties of the autonomy view, this proposal has to insist on the wrong of harmful rather than non-consensual using. But there need not be any difference in harmfulness of being affected as a side effect and used as a means.¹²³ Moreover, as *Chemotherapy* shows, harmful using need not be wrong.

It will not work to respond that in *Chemotherapy* the beneficiary and the victim are one and the same person. The substance of that response seems to be that intended harmful using of one person for the greater good of others is objectionable because it “does not sufficiently respect and take account of the fact that he is a separate person” (Nozick 1974: 33). However, merely foreseen harming of one as a side effect of a greater good for others is no less aggregative.

Another proposal is that intended harming as a means profits from, or takes advantage of, the victim’s misfortune in the way merely foreseen harming does not (Cavanaugh 2006: 97, 153, Quinn 1989: 344). To profit from one’s misfortune might mean that one’s misfortune is causally necessary to an end or causally sufficient but not necessary. However, there is often no wrongdoing in using others in such ways. Doctors, lawyers, vendors, and many others profit from misfortunes of their customers. As long as the customers consent to the services (and they are not otherwise exploitative), no wrong is done. Moreover, we might impermissibly profit from one’s misfortune

¹²³ Unless we stretch the notion of harm beyond its ordinary usage (see Duff 1990: 112, 1996: 338, cf. Westen 2011).

by harming him as a merely foreseen side effect. I can run over and kill a person as a side effect of rushing to the hospital to save five others (Kamm 2007: 22, cf. Kagan 1989: 174).

An influential justification is that intended harming as a means uses a victim as a *mere* means, or treats him *merely* as a means (Cavanaugh 2006: 147-58, Tadros 2011b: 136, Liao 2012: 723, cf. Nozick 1974: 30-1, Kamm 1992: 361). There are many ways to understand that phrase (see Parfit 2011a: Ch. 9). Consider the following three. On what seems to be Kant's (1997a: Ak, 4: 421-30) view, we treat others merely as a means when we treat them in ways to which they could not possibly consent. There is a debate about what Kant meant by possible consent. He probably meant rational consent in the sense of having good reasons to give or refuse it. But he did not mean that we cannot consent to being harmed as an intended means but can to being harmed as a merely foreseen side effect. Kant would also think that harming people without their consent in either of those ways is failing to treat them as ends in themselves.¹²⁴

On another view (Parfit 2011a: 214), we come close to using others as a mere means when we both use them as a means and give too little weight to their well-being or moral claims. We do not use others merely as a means, nor are we close to doing it, if either we are guided by some relevant moral belief or we (would) voluntarily bear a great burden for those people's sake.¹²⁵

Suppose that a runaway trolley is heading towards five trapped people threatening to kill them. I am on a bridge above with another man whom I topple off in front of the trolley, causing

¹²⁴ This is why those who appeal to using merely as a means cannot mean by it just failing to treat as an end in itself (cf. Bennett 1995: 18, Cavanaugh 2006: 148-58).

¹²⁵ The plausibility of the former disjunct depends on which moral beliefs are taken as relevant. Suppose that deeply held but false beliefs are relevant. Then, if by following a code of honor, I allow you to receive the last rites before I execute you in vendetta, I do not treat you merely as a means. I doubt that this is what we would normally say. On the other hand, if only true moral beliefs are relevant, the first disjunct practically reduces to the second.

the trolley to stop before it hits the five but killing that person. Suppose also that, if I were closer to the bridge's edge, I would sacrifice myself to save the five. On this view, I do not use the man merely as a means of saving the five (Parfit 2011a: 230). On a different view (Kamm 2007: 12, 40-1, n. 5), I do use him merely as a means because my behavior could be used as evidence that I regard him as a mere means. Since I would have constrained myself for that person's sake if I had a chance, the evidence of my attitude towards him as a mere means is misleading. But on this view, that does not suggest that I did not use him as a mere means. The view makes it harder to distinguish between harmful using a person as a means of a greater good for others from such harmful using as a mere means. More often than not, the evidence about one's attitude towards those one harms as a means misleadingly suggests that he regards them as mere tools.

I am skeptical that any account can fully capture what we mean by 'merely as a means'. The phrase is heavily moralized and it is difficult to distill its meaning without being tempted to equate wrongful using as a means with using as a mere means, as such. For our purposes, it is more important to see whether, given some understanding of the phrase, the appeal to it justifies the distinction between intending and merely foreseeing harm as morally relevant. I do not think it does, for the following reasons.

First, since we can intend harm as a means without intending it merely as a means, the proposal is at best limited. It is limited in other ways since it inherits most problems of the using view. Second, using others as mere means is often not wrong. Imagine that a doctor can either use chemotherapy or radiation therapy on a patient with cancer. He knows that chemotherapy is more effective but also more painful. Since he does not care about the patient's pain or wishes but wants to cure him in the most effective way, he prescribes chemotherapy. In both non-

Kantian senses of ‘merely as a means’ I brought above, the doctor is treating the patient merely as a means of curing him. But there is nothing wrong with it. I also do not believe that being used as a mere means adds to the wrongness of an otherwise wrong action. It is not less wrong to push the man off the bridge to save the five if I use him as a means rather than as a mere means (cf. Parfit 2011: 230-1). Third, it is hard to see what it is about using as a mere means that gives rise to reasons against intending harm. One might think that such using involves the attitude to the victim as a mere tool. But then we can repeat the question about the attitude itself. As the last example shows, to have that attitude need not even mean that one is attracted to evil.

Finally, I should emphasize that for the sake of the argument I granted that harmful using as a means entails intending to harmfully use as a means. But I do not think that that entailment holds. It is possible to harmfully use others as a means merely foreseeably. A driving instructor I know grabs the hands of his newcomer candidates while they are shifting the gears to show them how to do it properly. It is his habit, he does it intentionally, but he does not intend to do it each time he does it. The using as a means justification is therefore doomed from the start.¹²⁶

3. Impersonal Views

The remaining possibility is that bad intentions are bad impersonally (cf. McMahan 2009a: 357-8, FitzPatrick 2012: 103-4). If that were true, it would be an impersonally bad state of affairs when some person intends to kill or injure another. Reasons against intending harm would be

¹²⁶ Scanlon (2008: 115-7) suggests that using as a means affects the meaning of an action and makes the agent open to blame. Authors who appeal to the using view acknowledge this but hold that using is a distinct wrong (Nelkin and Rickless 2014: 152, n. 35).

agent-neutral in the sense that everyone would have reason not only not to intend harm but to wish or hope that less people intend it and to discourage and prevent those who do.

There are three ways how that could be true. First, intending harm can be impersonally bad for the same reasons it is bad for an agent or a victim. Agent-centered and victim-centered considerations, such as commitment to evil or violation of one's autonomy, could give rise both to agent-relative and agent-neutral reasons against intending harm. That would get us back to the problems with agent-centered and victim-centered views.

Second, intending harm could be intrinsically bad. Although intrinsic badness does not necessarily imply impersonal badness, many things thought to be intrinsically bad, such as pain or inequality, are also thought to be bad, period. If intending harm is intrinsically bad, it is bad solely in virtue of its properties. But there is nothing about intending that would make it bad (in contrast to pain or inequality). Intending, like believing or desiring, is value-neutral. If intending harm is impersonally bad, it is because of its special relation to the badness of harm.¹²⁷

Therefore, the third option is that intending harm is extrinsically impersonally bad. When Nagel talks about guidance by and attraction to evil as the properties of intending harm that make it distinctly objectionable, he might be read as saying that intending harm is not only personally but impersonally bad. As Kamm (1996: 246) puts it, Nagel's view might also be that each of us is constrained by the first bad intention he comes up against, whether or not it is his own. On similar views (Slote 2001, Hurka 2001), other mental states, character traits, emotions, and

¹²⁷ It might be said that this does not preclude calling intending harm intrinsically bad since a thing can be good or bad intrinsically even if it is good or bad solely in virtue of its relation to intrinsic goodness or badness (Hurka 2001: 49-50). I agree but, as I said above, if we take the view that bearers of intrinsic (dis)value are only states of affairs that include whatever is necessary to determine their intrinsic (dis)value, then, strictly speaking, pain and pleasure are intrinsically bad because of additional attendant circumstances (e.g. that one dislikes to be in pain). And since we rarely intend such intrinsic badness, it is more appropriate to say that intending harm is bad extrinsically. Nothing of importance for our question hinges on accepting this or some other view of intrinsic value.

motives in general are bad because they are (or comprise) positive attitudes to intrinsically bad things, like pain, or negative attitude to intrinsically good things, like pleasure. Pain is bad but so is desiring or being glad that someone is in pain.¹²⁸

Suppose that some such view is true. It does not immediately follow that, because a bad motive or intent to harm are impersonally bad, they have a bearing on an action's deontic status. We need an additional argument to establish that there are reasons not to act with intent to harm because that intent is impersonally bad. As Hurka (2010: 61-2) argues, it is possible (and in his view more likely) that intrinsic badness that gives rise to the badness of various mental states and motives independently gives rise to the wrongness of wrong actions. Their wrongness need not come from the badness of motives or intentions nor vice versa. If I murder an innocent person, the fact of his death suffices to make my act impermissible, though it could also be what makes my intention to kill him impersonally bad (together with the intention's relation to that fact). My claim is not that this is true but that it can be. If so, even if intending harm is bad, period, it need not be *pro tanto* wrong to act with it. It might as well be that everyone has reason to hope or wish that less people intend harm because of its likely consequences or the way that affects their moral characters.

Moreover, if intending harm is impersonally bad because it involves a positive attitude to harm as an intrinsically bad thing, then merely foreseeing harm is impersonally bad for the same reason. So even if intending harm is impersonally bad and even if that makes an action *pro tanto* wrong, we would still need an account of why intending harm is worse than merely foreseeing it.

¹²⁸ On those views, not only are intending, desiring, loving, etc. harm impersonally bad, but those attitudes to good are impersonally good. Some authors welcome this implication (Hurka 2001: 13-5, Wedgwood 2009a: 336). But the symmetry better not be perfect. For if reasons against intending harm are as weighty as reasons to intending good, the constraint on intending harm cannot be established (Kagan 1989: 177-81).

What is impersonally bad about intending might rather be the attraction to evil. But as I have just said, that would not grant the conclusion that there is a *pro tanto* reason against intending harm.

Finally, a problem with any impersonal view is that it justifies agent-neutral reasons but, as standardly assumed, deontological constraints are based on agent-relative reasons (cf. Kagan 1989: 181-2). One might argue that reasons against intending harm have their place inside the consequentialist framework and that among the things to be promoted is the minimization of the total number of violations of the constraint on intending harm. Alternatively, one might reject the assumption that deontological constraints must be agent-relative, arguing that there is a way to reconcile their agent-neutrality with the intuitive prohibition on minimizing the total number of their violations by violating them ourselves (i.e. to avoid the so-called ‘paradox of deontology’ while preserving agent-neutrality of deontological reasons). These are interesting possibilities but discussing them goes beyond the scope of my topic. If my objections to impersonal views hold, we need not discuss them with respect to the constraint on intending harm.

Conclusion

I have examined nine influential justifications of reasons against intending harm and showed that they are all untenable. As said before, this is not decisive evidence to think that there are no such reasons, as there might be better justifications, and theoretical or case-based reasons to think the contrary. However, it is a legitimate inference that the failure of these justifications is a good, if not the best, explanation that bad intentions are not wrongs *per se*.

Conclusion

1. The Argument in Brief

The thrust of my argument is that the Irrelevance Thesis is true. More precisely, it is that we have good reasons to believe that intentions are non-derivatively irrelevant to the moral permissibility of actions. I came to that conclusion in three steps. I argued that the best principles stating the contrary—the Doctrines of Double and Triple Effect, the Overdetermination View, the Intentions View—are false. I also argued that the best justifications for the relevance of intentions—agent-based, victim-based, and impersonal—are untenable but that so are the best justifications of the Irrelevance Thesis – the Objections from Confusion and Oddness, the Arguments from Choice and Encouragement. If my argument is sound, the strongest evidence for the Irrelevance Thesis is negative. Let me elaborate.

Call the view that intentions are sometimes non-derivatively relevant to permissibility the Relevance Thesis. The Relevance Thesis makes an existence claim: it states that there are cases where intentions matter to the permissibility of actions. The Irrelevance Thesis denies the truth of the existence claim stated by the Relevance Thesis. There are two ways to show that the denial is credible: either by showing that the thing claimed to exist cannot possibly exist or by refuting the arguments brought up in the defense of the existence claim. The former, if successful, turns

the page over the debate: it shows that no possible argument in the defense of the existence claim can succeed. The latter, on the other hand, shows less: it shows only that the arguments that have been brought up so far in support of the claim fail. However, if true, that conclusion challenges the defenders of the claim to bring up new arguments and it also provides us with what I call the negative evidence to support the Irrelevance Thesis. If the overall argument of my dissertation is sound, it is only that latter strategy which stands scrutiny.

There are two reasons why the onus is now on the defenders of the Relevance Thesis. First, it is because various *kinds* of arguments in favor of the relevance of intentions are shown to be unsound. I examined cases where intentions are thought to be relevant or where the principles appealing to intentions, such as the DDE, are said to have great explanatory power. I showed that these cases can be explained in ways that do not appeal to intentions and that these alternative explanations are as convincing. I also showed that the best principles that rely on intentions have many problems of their own and that they do not give intuitive results in relevant cases. Finally, I showed that the best theoretical justifications of the relevance of intentions are all untenable. The fact that these three kinds of different justifications fail poses a significant burden of proof on the defenders of the Relevance Thesis. I am of course aware that some of my arguments depend on controversial intuitions and assumptions (about permissible harming in overdetermination cases, fine-grained nature of intentions, etc.) and that some of my alternative explanations are not fully developed but I believe that the burden of proof befalls the other side much more heavily.

Second, since the defenders of the Relevance Thesis make an existence claim, they have to sustain it by arguments (unless the claim were about the existence of things that no one would dispute but this is far from true). In the absence of a sound argument, their critics need not bring

up some special argument (except if they want to show not only that the disputed claims are false but that they are necessarily false).

Now the sheer amount of problems that the defenders face does not entitle their critics to the conclusion that intentions are never relevant, even less that they cannot be. As I argued in the second part, the best positive arguments in favor of such conclusions are unsound. For sure, this does not mean that no such (future) argument can succeed. But it is also not impossible for the defenders of the Relevance Thesis to offer new cases where intentions matter, or principles that fare better against the relevant cases, or theoretical justifications which are tenable. As I have just suggested, this should not make us agnostic about the issue. The conclusion of this dissertation is a verdict about the current state of the debate and not the final word about it. Currently, the onus is on the defenders of the Relevance Thesis. Moreover, the conclusion is a verdict about what we now have most reason to believe. For what we know, there are deep problems in assuming the relevance of intentions and little, if any, problems in assuming the opposite. That we currently lack a sound argument proving their irrelevance for good should not discourage us, epistemically or morally, from being skeptical about their relevance and from assigning greater credence to the Irrelevance Thesis. But it remains to be seen which side of the debate will have the last laugh.

I shall now provide a partial error theory about why many people are led to believe that intentions are relevant to permissibility and discuss possible implications of rejecting that belief.

2. An Error Theory

I see twelve reasons why people think that intentions are relevant to permissibility.

First, some people do not distinguish between the derivative and non-derivative relevance of intentions. When they claim that intentions are relevant, they think of cases where intentions are predictively significant, as when intending to kill signifies the risk of murder. They also think of cases where reasons against intending are grounded in more basic reasons, as when intending to discriminate on the basis of one's race signifies that the victim is denied equal treatment or that he is an object of an unjustified exercise of power. Finally, they think of cases where one has a reason not to act with a bad intention because one would be open to blame, or have reasons for regret, or taint his moral character because of acting with it, which would cause him or his dear ones inconvenience or harm greater than the good that would be brought by acting in that way. Derivative relevances of intentions are not superficial but they are uncontroversial.

Second, some people mistakenly reason that if an intention is a necessary condition of an action being that of a certain type, then it must be non-derivatively relevant to the permissibility of that act type. But although one cannot lie unless one intends to deceive (apart perhaps from some rare cases), intending to deceive is not the, or even a, reason why lying is impermissible, when it is. The same is true of other act types for which particular intentions are necessary, such as fraud, robbery, or treason, as well as of generally permissible actions like marrying or making a will.

Third, some people confuse the relevance of evidence to permissibility for the relevance of intentions to permissibility. If I do what I justifiably believe will harm a person whom I intend to harm but, due to a good luck that no one could predict, that person survives only as a result of my action, there is a sense in which I acted impermissibly. However, that is because I justifiably expected to cause harm and not because I intended it.

Fourth, some people confuse the relevance of beliefs to permissibility for the relevance of intentions to permissibility. There might be a sense in which I act impermissibly by sticking pins in a voodoo doll of my hated enemy, thereby intending to kill him, but that is because I believe, mistakenly and unjustifiably, that I am about to kill a person and not because I intend to kill him. The same is true if I believe and intend, against the evidence, that I am to benefit a person but I kill him. There might be a sense in which I act permissibly but that is again due to what I believe and not due to what I intend.¹²⁹

Fifth, some people mistake the relevance of intentions to blameworthiness or viciousness of actions for their relevance to their impermissibility. The confusion is partly generated because the question of the relevance of intentions is sometimes framed ambiguously. It is asked whether intentions are relevant to the wrongness of actions, which might mean several things: whether they are relevant to impermissibility, to blameworthiness, to viciousness, or to some other moral category denoted by the term ‘wrongness’. The confusion also appears if one reasons that, since intending harm is in some sense more wrong than foreseeing it, it must be in the sense that there is a stronger reason against the permissibility of acting with that intention rather than doing so is more blameworthy or that it shows one as a worse person. However, philosophers defending the relevance of intentions should not, I believe, be accused of making that category mistake.

Sixth, some people believe that an action is impermissible because it is blameworthy and it is blameworthy because it is done with some bad intention. Others think that impermissibility is similarly grounded in one’s corrupt character. I have argued that those views are implausible.

¹²⁹ I believe that the belief-relative sense of wrongness exists and is tied to blameworthiness but I am skeptical about the belief-relative sense of impermissibility, as I cannot attach any purpose to it. I leave it as a possibility.

Again, I think, it is uncharitable to describe the vast majority of the defenders of the relevance of intentions as making this mistake.

Seventh, some people embrace the DDE or similar moral principles just because they are standard tools for moral reasoning of their religion. This is true of some uncritical Catholics and it is a clear groupthink phenomenon.

Eighth, some people equate sinning with acting impermissibly. They hold that one ought not to intend harm because one ought not to sin, and one sins just when and because one offends God's law. Since this view takes the relevance of any consideration to be derivative from the will of God, it must state the same about reasons not to intend harm. Although I do not argue for it, I also believe that no divine command theory is plausible.

Ninth, some people do not realize that reasons not to intend harm, if they exist and count against an action's permissibility in a non-derivative way, are relatively weak in comparison to countervailing reasons counting in favor of the action. This is in part because they do not focus on cases where the comparative weakness of those reasons is intuitively compelling. But since some of them do, what leads them to that conclusion are theoretical pre-commitments to moral absolutism or to various implausible views about weighing reasons against intending harm.

Tenth, some people are influenced by legal prohibitions on intending harm and other bad effects and think that there must be moral prohibitions of it. Perhaps the clearest influences are of humanitarian law and the law of attempts. But that intentions are relevant to criminal culpability entails nothing about their role in morality, let alone about their relevance to moral permissibility. The inference might go in the other direction, from morality to law, though that depends on what

we think the law is. I do not think that even that inference is often granted. I shall say more about it below.

Eleventh, some people adhere to the DDE and similar moral principles because they lack a better explanation of hard moral cases, such as the Trolley Problem. Others are convinced that those principles are the best thing on the market when it comes to explaining such cases. I argued that their explanations are unsatisfactory and that there are more convincing ones.

Twelfth, some people believe that, if we give up on the relevance of intentions, we must become consequentialists. That was the hope of early critics of the DDE as well as the fear of its defenders, especially Anscombe. However, as Thomson (1999: 517) remarks, “this merely shows a lack of imagination”. The current stage of moral theorizing is still in its infancy and we have no reason to assume that the simplest theories, like consequentialism, are the best theories.

3. Implications

If intentions are irrelevant to permissibility, what of importance follows for our moral beliefs and practices? For one thing, we get a better understanding of the problem. We clear up conceptual confusions about intentions, permissibility, blameworthiness, wrongness, and other concepts that feature in the debate and the relations between them. This is the ‘therapeutic’ contribution.

Obviously, there is a contribution to our moral knowledge: we get rid of false beliefs and acquire true ones regarding when and why an action is permissible. It becomes easier to realize the real reasons why an action is impermissible as purported reasons not to intend harm or other

bad effects are ruled out from the start. This might help in the search for the correct moral theory, though to a limited extent, since morality covers much more than the questions of permissibility.

Some implications are practical. Those who believe that reasons not to intend harm count decisively against acting and who would not do what they otherwise have sufficient reason to do when and because they intend harm, might realize that they are making a mistake and go on with acting despite intending harm. However, those who intend harm usually do not have sufficient reason to act because they cause unjustified harm. Therefore, changing the belief that intentions are relevant to permissibility would likely not prompt a major reduction of what is bad. But it might make some people do more good on some occasions.

Consider now non-moral beliefs and practices. It might be thought that the Irrelevance Thesis entails the falsity of religious convictions about permissible conduct. The best example is the appeal of Catholic moralists to the DDE with respect to issues as diverse as homicide and the use of contraception. However, we should be careful not to derive too quick conclusions. These views rely on metaphysical and conceptual presuppositions—about the normativity of morality, the nature of good and evil, and the meaning of moral concepts such as intention, action, moral wrongness, etc.—that might be strikingly different from those I outlined in the introduction and which I take as relatively uncontroversial among secular moral philosophers. It goes beyond the scope of this dissertation to engage with those presuppositions but I would not be surprised if my arguments miss those views because of the differences in our starting positions and conceptual apparatus. Nonetheless, since some dialogue is possible, moral theologians are obliged to clarify the differences and, if they think they are minimal, to engage with my arguments directly.

Lastly, there is a question about implications for criminal law. Husak (2009) worries that the price of incorporating the Irrelevance Thesis in criminal law is enormous, requiring sizable changes in its content. Others, in particular Thomson (1999), believe that easing the relevance of intentions in some spheres of criminal law is for the better. For instance, she finds it ludicrous that doctors should face legal action for physician-assisted suicide of the terminally ill if they intend to kill them but not if they merely foresee their deaths.

While in such and similar cases the law might not rest on a sound moral principle, more needs to be said in favor of changing it. One reason not to change the law comes from derivative relevances of intentions. Some acts are greater crimes when done with intention to harm than when done without it partly because the former are more likely to cause greater harm or do more wrong in other ways. A doctor who intends to kill a terminally ill patient, being aware that he is legally permitted to do it despite intending so, might kill the patient in a more harmful way than he would knowing that he faces legal action for acting with that intention. The lack of sanction against intending harm might also impose deterrence costs, incentivizing some doctors to kill their patients when they would not have sufficient reason to do it or to kill them in more harmful ways than they would in the presence of that sanction.

However, this is a two-edged sword. Prohibiting doctors from assisting their terminally ill patients in suicide when and because they would intend to kill them likely prolongs the patients's suffering by incentivizing some doctors who would intend to kill them and who care about not breaking the law to wait for their suitably motivated colleagues to take their place. It might also deter some doctors from taking up the responsibility to assist in suicide of the terminally ill, thus

allowing more suffering in the long run. The costs of deterrence are even greater when refraining from acting with intention to harm leads to deaths of innocent people (Tadros 2011b: 165-6).

It is a task of criminal law theory to decide which of these factors have priority and when. The theory must take into account the predictable significance of intending harm across different cases, the costs for deterrence of liberating the law from prohibiting acting with it, the extent to which the state is morally allowed to investigate defendants's intentions and motivation, and the ways in which the culpability of intending harm, in itself, makes a difference to counting an act as a crime or as a greater crime.

The last point, I believe, is important since, even if the Irrelevance Thesis is incorporated in the law, the law need not change, or not substantially, since intentions would be relevant for criminal culpability in the same way they are relevant for moral culpability. For instance, even if there is no difference in permissibility between actions of a doctor who administers a lethal drug to a terminally ill patient intending to kill him and a doctor who does it without that intention, the former's intention might call for treating his action as a greater crime because it might make the action more blameworthy. Whether intending harm, in itself, suffices for blame or only when it stems from a bad motive, such as hatred or revenge, is another issue that the criminal law theory must tackle. But primarily, its task is to explain whether acting with such intentions, in itself (i.e. independently of its expectable effects for the victim or for potential future criminals), provides enough reason for treating an action as a greater crime.

I cannot engage with any of these questions here. There is likely a sea of others. What matters is that the Irrelevance Thesis opens space for research beyond moral theory. Along with its theoretical and practical impact to morality, this warrants taking the Thesis seriously.

Bibliography

- Adams, Robert M. 1976. Motive Utilitarianism. *Journal of Philosophy*, 73 (14): 467–481.
- Adams, Robert M. 1985. Involuntary Sins. *Philosophical Review*, 94 (1): 3–31.
- Alexander, Larry and Ferzan, Kimberly Kessler. 2009. *Crime and Culpability: A Theory of Criminal Law*. Cambridge: Cambridge University Press.
- Alston, William P. 1988. The Deontological Conception of Epistemic Justification. *Philosophical Perspectives*, 2 (1): 257–299.
- Anderson, Elizabeth. 1993. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.
- Anscombe, Elizabeth. 1981. *Ethics, Religion, and Politics: The Collected Philosophical Papers by G. E. M. Anscombe*, Vol. 3. Minneapolis: University of Minnesota Press.
- Anscombe, Elizabeth. 2000 (1957). *Intention*. Cambridge, MA: Harvard University Press.
- Anscombe, Elizabeth. 2005. *Human Life, Action, and Ethics: Essays by G. E. M. Anscombe*, ed. by Mary Geach and Luke Gormally. Exeter: Imprint Academic.
- Aquinas, Thomas. 1981. *Summa Theologica*. Notre Dame: Christian Classics (reprint of 1948 Benzinger's Bros ed.).
- Armstrong, D. M. 1973. Acting and Trying. *Philosophical Papers*, 2 (1): 1–15.
- Arneson, Richard J. 2006. What is Wrongful Discrimination? *San Diego Law Review*, 43 (4): 775–807.
- Audi, Robert. 1973. Intending. *Journal of Philosophy*, 70 (13): 387–403.
- Audi, Robert. 1993. *Action, Intention, and Reason*. Ithaca: Cornell University Press.
- Banerjee, Konika, Huebner, Bryce, Hauser, Marc. 2010. Intuitive Moral Judgments are Robust across Variation in Gender, Education, Politics, and Religion: A Large-Scale Web-Based Study. *Journal of Cognition and Culture*, 10 (3): 151–187.
- Bennett, Jonathan. 1966. Whatever the Consequences. *Analysis*, 26 (3): 83–102.
- Bennett, Jonathan. 1974. The Conscience of Huckleberry Finn. *Philosophy*, 49 (188): 123–134.
- Bennett, Jonathan. 1981. Morality and Consequences. In Sterling McMurrin (ed.), *The Tanner Lectures on Human Values*, Vol. 2. Salt Lake City: University of Utah Press.
- Bennett, Jonathan. 1995. *The Act Itself*. Oxford: Oxford University Press.
- Bentham, Jeremy. 1907 (1781). *An Introduction to the Principles of Morals and Legislation*. Oxford: Oxford University Press.
- Berker, Selim. 2015. The Unity of Grounding. MS (on file with author), pp. 55.
- Boyle, Joseph M. 1980. Toward Understanding of the Principle of Double Effect. *Ethics*, 90 (4): 527–538.
- Boyle, Joseph M. 1991a. Who is Entitled to Double Effect? *Journal of Medicine and Philosophy*, 16 (5): 475–494.
- Boyle, Joseph M. 1991b. Further Thoughts on Double Effect: Some Preliminary Responses. *Journal of Medicine and Philosophy*, 16 (5): 565–570.

- Boyle, Joseph M. 2004. Medical Ethics and Double Effect: The Case of Terminal Sedation. *Theoretical Medicine and Bioethics*, 25 (1): 51–60.
- Bratman, Michael. 1987. *Intentions, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
- Bratman, Michael. 1997. Two Faces of Intention. In Alfred R. Mele (ed.), *The Philosophy of Action*. Oxford: Oxford University Press.
- Broome, John. 2002. Practical Reasoning. In José Bermúdez and Alan Millar (eds.), *Reason and Nature: Essays in the Theory of Rationality*. Oxford: Oxford University Press.
- Brunero, John. 2014. Cognitivism about Practical Reason. In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Vol. 9. Oxford: Oxford University Press.
- Catechism of the Catholic Church*. 1993. Vatican City. <<http://www.vatican.va/archive/ENG0015/INDEX.HTM>>.
- Cavanaugh, Thomas A. 1999. Double Effect Reasoning: A Response to Bennett. *Journal of Applied Philosophy*, 16 (2): 181–184.
- Cavanaugh, Thomas A. 2006. *Double Effect Reasoning: Doing Good and Avoiding Evil*. New York: Oxford University Press.
- Chan, David K. 2000. Intention and Responsibility in Double Effect Cases. *Ethical Theory and Moral Practice*, 3 (4): 405–434.
- Chappell, Timothy. 2002. Two Distinctions That Do Make a Difference: The Act/Omission Distinction and the Principle of Double Effect. *Philosophy*, 77 (300): 211–233.
- Chappell, Timothy. 2013. What Have I Done? *Diametros*, 38 (1): 86–112.
- Chiao, Vincent. 2010. Intention and Attempt. *Criminal Law and Philosophy*, 4 (1): 37–55.
- Chiao, Vincent. 2013. Punishment and Permissibility in Criminal Law. *Law and Philosophy*, 32 (6): 729–765.
- Chisholm, Roderick M. 1970. The Structure of Intention. *Journal of Philosophy*, 67 (19): 633–647.
- Connell, F. J. 2002. Double Effect, Principle of. In Thomas Carson (ed.), *New Catholic Encyclopedia*, Vol. 4, 2nd ed. Detroit: Gale.
- Convention (IV) Respecting the Laws and Customs of War on Land and its Annex: Regulations Concerning the Laws and Customs of War on Land*. 1907. The Hague. <<https://www.icrc.org/ihl/INTRO/195>>.
- Copp, David. 2003. ‘Ought’ Implies ‘Can’: Blameworthiness, and the Principle of Alternate Possibilities. In Michael McKenna and David Widerker (eds.), *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Burlington: Ashgate.
- Daniels, Norman. 1996. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. New York: Cambridge University Press.
- Darwall, Stephen. 1977. Two Kinds of Respect. *Ethics*, 88 (1): 36–49.
- Darwall, Stephen. 1986. Agent-Centered Restrictions from the Inside Out. *Philosophical Studies*, 50 (3): 291–319.
- Davis, Nancy. 1984. The Doctrine of Double Effect: Problems of Interpretation. *Pacific Philosophical Quarterly*, 65 (2): 107–123.
- Delaney, Neil F. 2007. A Note on Intention and the Doctrine of Double Effect. *Philosophical Studies*, 134 (2): 103–110.
- Delaney, Neil F. 2008. Two Cheers for ‘Closeness’: Terror, Targeting and Double Effect. *Philosophical Studies*, 137 (3): 335–367.
- Di Nucci, Ezio. 2014. *Ethics Without Intention*. London: Bloomsbury.

- Donagan, Alan. 1977. *The Theory of Morality*. Chicago: University of Chicago Press.
- Donagan, Alan. 1991. Moral Absolutism and the Double-Effect Exception: Reflections on Joseph Boyle's 'Who is Entitled to Double Effect?'. *Journal of Medicine and Philosophy*, 16 (5): 495–509.
- Douglas, Charles, Kerridge, Ian, and Ankeny, Rachel. 2013. Narratives of 'Terminal Sedation', and the Importance of the Intention-Foresight Distinction in Palliative Care Practice. *Bioethics* 27 (1): 1–11.
- Driver, Julia. 1992. The Suberogatory. *Australasian Journal of Philosophy*, 70 (3): 286–295.
- Duff, Antony R. 1982. Intention, Responsibility and Double Effect. *Philosophical Quarterly*, 32 (126): 1–16.
- Duff, Antony R. 1989. Intention Legal and Philosophical. *Oxford Journal of Legal Studies*, 9 (1): 76–94.
- Duff, Antony R. 1990. *Intention, Agency & Criminal Liability: Philosophy of Action and the Criminal Law*. Oxford: Basil Blackwell.
- Duff, Antony R. 1996. *Criminal Attempts*. Oxford: Oxford University Press.
- Duff, Antony R. 2007. *Answering for Crime*. Oxford: Hart Publishing.
- Duff, Antony R. 2012. Guiding Commitments and Criminal Liability for Attempts. *Criminal Law and Philosophy*, 6 (3): 411–427.
- Dworkin, Ronald. 2011. *Justice for Hedgehogs*. Cambridge, MA: Harvard University Press.
- Edmonds, David. 2014. *Would You Kill the Fat Man: The Trolley Problem and What Your Answer Tells Us about Right and Wrong*. Princeton: Princeton University Press.
- Elster, Jakob. 2012. Scanlon on Permissibility and Double Effect. *Journal of Moral Philosophy*, 9 (1): 75–102.
- Enoch, David. 2007. Intending, Foreseeing, and the State. *Legal Theory*, 13 (2): 69–99.
- Fine, Kit. 2012. Guide to Ground. In Fabrice Correia and Benjamin Schnieder (eds.), *Metaphysical Grounding*. Cambridge: Cambridge University Press.
- Finnis, John. 2011a. Intention and Side Effects. In his *Intention and Identity: Collected Essays*, Vol. 2. Oxford: Oxford University Press.
- Finnis, John. 2011b. Intention in Tort Law. In his *Intention and Identity: Collected Essays*, Vol. 2. Oxford: Oxford University Press.
- Fischer, John M., Ravizza, Mark and Copp, David. 1993. Quinn on Double Effect: The Problem of Closeness. *Ethics*, 103 (4): 707–725.
- Fischer, John Martin. 1994. *The Metaphysics of Free Will*. Oxford: Blackwell.
- Fischer, John Martin. 2006. *My Way: Essays on Moral Responsibility*. New York: Oxford University Press.
- FitzPatrick, William. 2003. Acts, Intentions, and Moral Permissibility: In Defense of the Doctrine of Double Effect. *Analysis*, 63 (4): 317–321.
- FitzPatrick, William. 2006. The Intend/Foresee Distinction and the Problem of 'Closeness'. *Philosophical Studies*, 128 (3): 585–617.
- FitzPatrick, William. 2009. Thomson's Turnabout on the Trolley. *Analysis*, 69 (4): 636–643.
- FitzPatrick, William. 2012. Intention, Permissibility, and Double Effect. In Mark Timmons (ed.), *Oxford Studies in Normative Ethics*, Vol. 2. Oxford: Oxford University Press.
- Fohr, Susan A. 1998. The Double Effect of Pain Medication: Separating Myth from Reality. *Journal of Palliative Care*, 1 (4): 315–328.
- Foot, Philippa. 2002a. *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Oxford University Press.

- Foot, Philippa. 2002b. *Moral Dilemmas and Other Topics in Moral Philosophy*. Oxford: Oxford University Press.
- Frankfurt, Harry G. 1969. Alternate Possibilities and Moral Responsibility. *Journal of Philosophy*, 66 (23): 829–839.
- Frankfurt, Harry G. 1988. What We are Morally Responsible For. In his *The Importance of What We Care About*. Cambridge: Cambridge University Press.
- Fried, Charles. 1978. *Right and Wrong*. Cambridge, MA: Harvard University Press.
- Frowe, Helen. 2014. *Defensive Killing*. New York: Oxford University Press.
- García, Jorge L. A. 1993. The New Critique of Anti-Consequentialist Moral Theory. *Philosophical Perspectives*, 71 (1): 1–32.
- García, Jorge, L. A. 1995. Intention-Sensitive Ethics. *Public Affairs Quarterly*, 9 (3): 201–213.
- Gibbard, Allan. 2005. Truth and Correct Belief. *Philosophical Issues*, 15 (1): 338–350.
- Ginet, Carl. 1990. *On Action*. Cambridge: Cambridge University Press.
- Glover, Jonathan. 1977. *Causing Death and Saving Lives*. London: Penguin.
- Goldman, Alvin I. 1970. *A Theory of Human Action*. Englewood Cliffs: Prentice-Hall.
- Graham, Peter A. 2010. In Defense of Objectivism about Moral Obligation. *Ethics*, 121 (1): 88–115.
- Graham, Peter A. 2011. ‘Ought’ and Ability. *Philosophical Review*, 120 (3): 337–382.
- Graham, Peter A. 2014. A Sketch of a Theory of Moral Blameworthiness. *Philosophy and Phenomenological Research*, 88 (2): 388–409.
- Grayling, Antony C. 2006. *Among the Dead Cities: Is Targeting Civilians in War Ever Justified?* London: Bloomsbury Publishing.
- Grice, Paul. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Hague Rules of Air Warfare. 1922–1923. The Hague. <http://lawofwar.org/hague_rules_of_air_warfare.htm>.
- Haji, Ishtiyaque. 1993. Alternative Possibilities, Moral Obligation, and Moral Responsibility. *Philosophical Papers*, 22 (1): 41–50.
- Haji, Ishtiyaque. 2002. *Deontic Morality and Control*. Cambridge: Cambridge University Press.
- Haji, Ishtiyaque. 2012. *Reason’s Debt to Freedom: Normative Appraisals, Reasons, and Free Will*. New York: Oxford University Press.
- Hanser, Matthew. 2005. Permissibility and Practical Inference. *Ethics*, 115 (3): 443–470.
- Hanser, Matthew. 2014. Acting Wrongly by Trying. In Mark Timmons (ed.), *Oxford Studies in Normative Ethics*, Vol. 4. Oxford: Oxford University Press.
- Harman, Gilbert. 1976. Practical Reasoning. *Review of Metaphysics*. 29 (3): 431–463.
- Hart, Herbert L. A. 2008. Intention and Punishment. In his *Punishment and Responsibility: Essays in the Philosophy of Law*, 2nd edition. Oxford: Oxford University Press.
- Hauser, Marc D. 2006. *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York: Ecco/Harper Collins.
- Hayenhjelm, Madeleine and Wolf, Jonathan. 2011. The Moral Problem of Risk Impositions: A Survey of the Literature. *European Journal of Philosophy*, 20 (S1): 26–51.
- Hellman, Deborah. 2008. *When is Discrimination Wrong?* Cambridge, MA: Harvard University Press.
- Herman, Barbara. 2011. A Mismatch of Methods. In Derek Parfit, *On What Matters*, Vol. 2. Oxford: Oxford University Press.
- Heuer, Ulrike. 2014. Intentions, Permissibility, and the Reasons for Which We Act. In Gregory Pavlakos and Veronica Rodriguez-Blanco (eds.), *Practical Normativity: Essays on Reasons and Intentions in Law and Practical Reason*. Cambridge: Cambridge University Press.

- Hieronymi, Pamela. 2006. Controlling Attitudes. *Pacific Philosophical Quarterly*, 87 (1): 45–74.
- Hieronymi, Pamela. 2011. Believing at Will. *Canadian Journal of Philosophy*, Suppl. Vol., 35 (1): 149–187.
- Hills, Allison. 2003. Defending Double Effect. *Philosophical Studies*, 116 (2): 133–152.
- Hills, Allison. 2007. Intentions, Foreseen Consequences, and the Doctrine of Double Effect. *Philosophical Studies*, 133 (2): 257–283.
- Hills, Allison. 2009. Review of Scanlon's *Moral Dimensions*. *Ethics*, 119 (4): 792–796.
- Horsnby, Jennifer. 1980. *Actions*. London: Routledge and Kagan Paul.
- Hornsby, Jennifer. 2010. Trying to Act. In Timothy O'Connor and Constantine Sandis (eds.), *A Companion to Philosophy of Action*. Hoboken: Wiley-Blackwell.
- Hurka, Thomas. 2001. *Virtue, Vice, and Value*. New York: Oxford University Press.
- Hurka, Thomas. 2010. Right Act, Virtuous Motive. In Heather Battaly (ed.), *Virtue and Vice: Moral and Epistemic*, Hoboken: Wiley-Blackwell.
- Hurka, Thomas. 2014a. Kamm on Intention and Proportionality in War. *Journal of Moral Philosophy*, 11 (4): 411–427.
- Hurka, Thomas. 2014b. *British Ethical Theorists from Sidgwick to Ewing*. New York: Oxford University Press.
- Hurka, Thomas. 2015. Trolleys and Permissible Harm. In Eric Rakowski (ed.), *The Trolley Problem Mysteries*. New York: Oxford University Press.
- Husak, Douglas. 2009. The Cost to Criminal Theory of Supposing that Intentions are Irrelevant to Permissibility. *Criminal Law and Philosophy*, 3 (1): 51–70.
- Husak, Douglas. 2010. *The Philosophy of Criminal Law: Selected Essays*. New York: Oxford University Press.
- Jackson, Frank and Pargetter, Robert. 1986. Oughts, Options, and Actualism. *Philosophical Review*, 95 (2): 233–255.
- Kaczor, Christopher R. 1998. Double Effect Reasoning: From Jean Pierre Gury to Peter Knauer. *Theological Studies*, 59 (1): 297–316.
- Kagan, Shelly. 1988. The Additive Fallacy. *Ethics*, 99 (1): 5–31.
- Kagan, Shelly. 1989. *The Limits of Morality*. Oxford: Oxford University Press.
- Kagan, Shelly. 1998. *Normative Ethics*. Boulder: Westview Press.
- Kagan, Shelly. 2015. Solving the Trolley Problem. In Eric Rakowski (ed.), *The Trolley Problem Mysteries*. New York: Oxford University Press.
- Kamm, Frances M. 1991. The Doctrine of Double Effect: Reflections on Theoretical and Practical Issues. *Journal of Medicine and Philosophy*, 16 (5): 571–585.
- Kamm, Frances M. 1992. Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status. *Philosophy and Public Affairs*, 21 (4): 354–389.
- Kamm, Frances M. 1995. Inviolability. *Midwest Studies in Philosophy*, 20 (1): 165–175.
- Kamm, Frances M. 1996. *Morality, Mortality: Rights, Duties, and Status*, Vol. 2. New York: Oxford University Press.
- Kamm, Frances M. 1999. Physician-Assisted Suicide, the Doctrine of Double Effect, and the Ground of Value. *Ethics*, 109 (3): 586–605.
- Kamm, Frances M. 2004. Failures of Just War Theory: Terror, Harm and Justice. *Ethics*, 114 (4): 650–692.
- Kamm, Frances M. 2007. *Intricate Ethics*. Oxford: Oxford University Press.
- Kamm, Frances M. 2008. Terrorism and Intending Evil. *Philosophy and Public Affairs*, 36 (2): 157–186.

- Kamm, Frances M. 2012. *The Moral Target: Aiming at Right Conduct in War and Other Conflicts*. New York: Oxford University Press.
- Kamm, Frances M. 2014. Responses. *Journal of Moral Philosophy*, 11 (4): 476–517.
- Kamm, Frances M. 2015. *The Trolley Problem Mysteries*, ed. by Eric Rakowski. New York: Oxford University Press.
- Kant, Immanuel. 1996. *Metaphysics of Morals*, ed. by Marry Gregor. Cambridge: Cambridge University Press.
- Kant, Immanuel. 1997a. *Groundwork of the Metaphysics of Morals*, ed. by Marry Gregor. Cambridge: Cambridge University Press.
- Kant, Immanuel. 1997b. *Critique of Practical Reason*, ed. by Marry Gregor. Cambridge: Cambridge University Press.
- Kendall, C. E. 2000. A Double Dose of Double Effect. *Journal of Medical Ethics*, 26 (3): 204–206.
- Kenny, Anthony J. P. 1968. Intention and Purpose in Law. In Robert Summers (ed.), *Essays in Legal Philosophy*. Berkeley: University of California Press.
- Kenny, Anthony J. P. 1973. The History of Intention in Ethics. In his *The Anatomy of the Soul*, Oxford: Basil Blackwell.
- Kenny, Anthony J. P. 1977. Intentions and *Mens Rea* in Murder. In P. M. S. Hacker and Joseph Raz (eds.), *Law, Morality, and Society*. Oxford: Oxford University Press.
- Knobe, Joshua. 2003. Intentional Action and Side Effects in Ordinary Language. *Analysis*, 63 (279): 190–194.
- Kolodny, Niko. 2011a. Scanlon's Investigation: The Relevance of Intent to Permissibility. *Analytic Philosophy*, 52 (2): 100–123.
- Kolodny, Niko. 2011b. Aims as Reasons. In R. Jay Wallace, Rahul Kumar, and Samuel Freeman (eds.), *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. New York: Oxford University Press.
- Levy, Neil. 2007. Doxastic Responsibility. *Synthese*, 155 (1): 125–155.
- Levy, Sanford S. 1986. The Principle of Double Effect. *Journal of Value Inquiry*, 20 (1): 29–40.
- Lewis, David K. 1986. *Philosophical Papers*, Vol. 2. New York: Oxford University Press.
- Liao, Matthew S., Wiegman, Alex, Alexander, Joshua, Vong, Gerard. 2011. Putting the Trolley in Order: Experimental Philosophy and the Loop Case. *Philosophical Psychology*, 25 (5): 1–11.
- Liao, Matthew S. 2009. The Loop Case and Kamm's Doctrine of Triple Effect. *Philosophical Studies*, 146 (2): 223–231.
- Liao, Matthew S. 2012. Intentions and Moral Permissibility: The Case of Acting Permissibly with Bad Intentions. *Law and Philosophy*, 31 (6): 703–724.
- Liao, Matthew S. 2014. The Closeness Problem and the Doctrine of Double Effect: A Way Forward. *Criminal Law and Philosophy*, 1–15.
- Lillehammer, Hallvard. 2010. Scanlon on Intention and Permissibility. *Analysis*, 70 (3): 578–585.
- Lippert-Rasmussen, Kasper. 2010. Scanlon on the Doctrine of Double Effect. *Social Theory and Practice*, 36 (4): 541–564.
- Lippert-Rasmussen, Kasper. 2013. *Born Free and Equal? A Philosophical Inquiry into the Nature of Discrimination*. New York: Oxford University Press.
- Lippert-Rasmussen, Kasper. 2014. Just War Theory, Intentions, and the Deliberative Perspective Objection. In Helen Frowe and Gerald Lang (eds.), *How We Fight: Ethics in War*. Oxford: Oxford University Press.

- Mangan, Joseph. 1949. An Historical Analysis of the Principle of Double Effect. *Theological Studies*, 10 (1): 41–61.
- Mapel, David. 2001. Revising the Doctrine of Double Effect. *Journal of Applied Philosophy*, 18 (3): 257–272.
- Markovits, Julia. 2010. Acting for the Right Reasons. *Philosophical Review*, 119 (2): 201–242.
- Marquis, Donald B. 1991. Four Versions of Double Effect. *Journal of Medicine and Philosophy*, 16 (5): 515–544.
- Masek, Lawrence. 2010. Intentions, Motives, and the Doctrine of Double Effect. *Philosophical Quarterly*, 60 (240): 567–585.
- McIntyre, Alison. 2001. Doing Away with Double Effect. *Ethics*, 111 (2): 219–255.
- McIntyre, Alison. 2004. The Double Life of Double Effect. *Theoretical Medicine and Bioethics*, 25 (1): 61–74.
- McIntyre, Alison. 2014. Doctrine of Double Effect. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, <<http://plato.stanford.edu/archives/win2014/entries/double-effect/>>.
- McCarthy, David. 2002. Intending Harm, Foreseeing Harm, and Failures of the Will. *Noûs*, 36 (4): 622–642.
- McMahan, Jeff. 1994. Revising the Doctrine of Double Effect. *Journal of Applied Philosophy*, 11 (2): 201–212.
- McMahan, Jeff. 2000. Moral Intuition. In Hugh LaFollette (ed.), *The Blackwell Guide to Ethical Theory*. Oxford: Blackwell.
- McMahan, Jeff. 2005. The Basis of Moral Liability to Defensive Killing. *Philosophical Issues*, 15 (1): 386–405.
- McMahan, Jeff. 2009a. Intention, Permissibility, Terrorism, and War. *Philosophical Perspectives*, 23 (1): 345–372.
- McMahan, Jeff. 2009b. *Killing in War*. Oxford: Oxford University Press.
- Mele, Alfred R. 1989. She Intends to Try. *Philosophical Studies*, 55 (1): 101–106.
- Mele, Alfred R. 1992. Recent Work on Intentional Action. *American Philosophical Quarterly*, 29 (3): 199–217.
- Mele, Alfred R. 2003. Agents' Abilities. *Noûs*, 37 (3): 447–470.
- Mikhail, John. 2011. *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge: Cambridge University Press.
- Mikhail, John. 2013. New Perspective on Moral Cognition: Reply to Zimmerman, Enoch, and Chemla, Engré, and Schlenker. *Jerusalem Review of Legal Studies*, 8 (1): 66–114.
- Mill, John Stuart. 1998 (1861). *Utilitarianism*. Oxford: Oxford University Press.
- Moore, G. E. 2005 (1912). *Ethics*. Oxford: Oxford University Press.
- Moore, Michael S. 1997. *Placing Blame: A Theory of the Criminal Law*. New York: Oxford University Press.
- Moore, Michael S. 2010. *Act and Crime: The Philosophy of Action and Its Implication for Criminal Law*. New York: Oxford University.
- Moore, Michael S. 2011. Intention as a Marker of Moral Responsibility and Legal Punishability. In R. A. Duff and Stuart Green (eds.), *The Philosophical Foundations of Criminal Law*. Oxford: Oxford University Press.
- Nagel, Thomas. 1970. *The Possibility of Altruism*. Oxford: Oxford University Press.
- Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

- Nagel, Thomas. 1979. War and Massacre. In his *Mortal Questions*, New York: Cambridge University Press.
- Nagel, Thomas. 2008. The Value of Inviolability. In Paul Bloomfield (ed.), *Morality and Self-Interest*. New York: Oxford University Press.
- Naylor, Margery Bedford. 1984. Frankfurt on the Principle of Alternate Possibilities. *Philosophical Studies*, 46 (2): 249–258.
- Nelkin, Dana K. 2011. Review of Thomas Scanlon, *Moral Dimensions*. *Philosophical Review*, 120 (4): 603–607.
- Nelkin, Dana K. and Rickless, Samuel C. 2014. Three Cheers for Double Effect. *Philosophy and Phenomenological Research*, 89 (1): 125–158.
- Nelkin, Dana K. and Rickless, Samuel C. 2015. So Close, Yet So Far: Why Solutions to the Closeness Problem for the Doctrine of Double Effect Fall Short. *Noûs*, 49 (2): 376–409.
- Nelkin, Dana K. and Rickless, Samuel C. 2016. The Relevance of Intention to Criminal Wrongdoing. *Criminal Law and Philosophy*, 1–18.
- Norcross, Alastair. 1999. Intending and Foreseeing Death: Potholes on the Road to Hell. *Southwest Philosophy Review*, 15 (1): 115–123.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Oxford: Basil Blackwell.
- Nye, Howard. 2013. Objective Double Effect and the Avoidance of Narcissism. In Mark Timmons (ed.), *Oxford Studies in Normative Ethics*, Vol. 3. Oxford: Oxford University Press.
- Nye, Howard. 2014a. On the Equivalence of Trolleys and Transplants: The Lack of Intrinsic Difference between ‘Collateral Damage’ and Intended Harm. *Utilitas*, 26 (4): 432–479.
- Nye, Howard. 2014b. Quinn’s Interpretation of Double Effect: Problems and Prospects. MS (on file with author), pp. 14.
- Otsuka, Michael. 1997. Kamm on the Morality of Killing. *Ethics*, 108 (1): 197–207.
- Otsuka, Michael. 2008. Double Effect, Triple Effect and the Trolley Problem: Squaring the Circle in Looping Cases. *Utilitas*, 20 (1): 92–110.
- Øverland, Gerhard. 2014. Moral Obstacles: An Alternative to the Doctrine of Double Effect. *Ethics*, 124 (3): 481–506.
- Øverland, Gerhard. 2015. Why Kamm’s Principle of Secondary Permissibility Cannot Save the Doctrine of Double Effect. *Journal of Applied Philosophy*, 1–11.
- Quinn, Warren S. 1989. Actions, Intentions, and Consequences: The Doctrine of Double Effect. *Philosophy and Public Affairs*, 18 (4): 334–351.
- Quinn, Warren S. 1991. Reply to Boyle’s ‘Who is Entitled to Double Effect?’. *Journal of Medicine and Philosophy*, 16 (5): 511–514.
- Quong, Jonathan. 2013. Intentions and Permissibility: A Puzzling Asymmetry. MS (on file with author), pp. 50.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Parfit, Derek. 2011a. *On What Matters*, Vol. 1. Oxford: Oxford University Press.
- Parfit, Derek. 2011b. *On What Matters*, Vol. 2. Oxford: Oxford University Press.
- Parfit, Derek. 2013. Moral Agreement. In Hugh LaFollette and Ingmar Persson (eds.), *The Blackwell Guide to Ethical Theory*, 2nd edition. Hoboken: Blackwell.
- Parfit, Derek. 2015a. *On What Matters*, Vol. 3. MS (on file with author), pp. 251.
- Parfit, Derek. 2015b. *On What Matters*, Vol. 4. MS (on file with author), pp. 60.
- Paul VI. 1968. *Humane Vitae*. <http://w2.vatican.va/content/paul-vi/en/encyclicals/documents/hf_p-vi_enc_25071968_humanae-vitae.html>.

- Paul, John II. 1993. *Veritatis Splendor*. <http://w2.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_06081993_veritatis-splendor.html>.
- Paul, John II. 1995. *Evangelicum Vitae*. <http://w2.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_25031995_evangelium-vitae.html>.
- Persson, Ingmar. 2013. *From Morality to the End of Reason*. Oxford: Oxford University Press.
- Predelli, Stefano. 2004. Bombers: Some Comments on Double Effect and Harmful Involvement. *Journal of Military Ethics*, 3 (1): 16–26.
- Prichard, Harold A. 2002. *Moral Writings*, ed. by Jim MacAdam. Oxford: Oxford University Press.
- Pruss, Alexander R. 2013. The Accomplishment of Plans: A New Version of the Principle of Double Effect. *Philosophical Studies*, 165 (1): 49–69.
- Rachels, James. 1986. *The End of Life: Euthanasia and Morality*. Oxford: Oxford University Press.
- Rawls, John. 1999. *A Theory of Justice*, 2nd edition. Cambridge, MA: Harvard University Press.
- Raz, Joseph. 1999. *Practical Reason and Norms*. Oxford: Oxford University Press.
- Ridge, Michael. 2002. Mill's Intentions and Motives. *Utilitas*, 14 (1): 54–70.
- Ripstein, Arthur. 2006. Beyond the Harm Principle. *Philosophy and Public Affairs*, 34 (3): 215–245.
- Reese, Roger. 2011. *Why Stalin's Soldiers Fought: The Red Army's Military Effectiveness in World War II*. Lawrence: University Press of Kansas.
- Ross, David W. 1968 (1939). *Foundations of Ethics*. Oxford: Oxford University Press.
- Ross, David W. 2002 (1930). *The Right and the Good*. New York: Oxford University Press.
- Ross, Jacob. 2012. Actualism, Possibilism, and Beyond. In Mark Timmons (ed.), *Oxford Studies in Normative Ethics*, Vol. 2. Oxford: Oxford University Press.
- Roughley, Neil. 2007. The Double Failure of 'Double Effect'. In Christopher Lumer and Sandro Nannini (eds.), *Intentionality, Deliberation, and Autonomy: The Action-Theoretic Basis of Practical Philosophy*. Farnham: Ashgate.
- Saemi, Amir. 2009. Intention and Permissibility. *Ethical Perspectives*, 16 (1): 81–101.
- Schwartz, Daniel. 2015. Against Substitutive Harm. *Journal of Applied Philosophy*, 1–14.
- Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Scanlon, Thomas M. 2000. Intention and Permissibility I. *Proceedings of the Aristotelian Society*, Suppl. Vol., 74 (1): 301–317.
- Scanlon, Thomas M. 2003. Rawls on Justification. In Samuel Freeman (ed.), *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press.
- Scanlon, Thomas M. 2008. *Moral Dimensions: Permissibility, Meaning, and Blame*. Cambridge, MA: Harvard University Press.
- Scanlon, Thomas M. 2010. Ambiguity of "Intention". *Behavioral and Brain Sciences*, 33 (4): 348–349.
- Scanlon, Thomas M. 2011a. Précis on *Moral Dimensions: Permissibility, Meaning, and Blame*. *Philosophy and Phenomenological Research*, 83 (2): 459–463.
- Scanlon, Thomas M. 2011b. Reply to Hill, Mason, and Wedgwood. *Philosophy and Phenomenological Research*, 83 (2): 490–505.
- Scanlon, Thomas M. 2015. Forms and Limits of Responsibility. In Randolph Clarke, Michael McKenna, and Angela M. Smith (eds.), *The Nature of Moral Responsibility: New Essays*. New York: Oxford University Press.
- Schroeder, Severin. 2001. The Concept of Trying. *Philosophical Investigations*, 24 (3): 213–227.
- Schwitzgebel, Eric and Cushman, Fierry. 2012. Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind and Language*, 27 (2): 135–153.

- Searle, John R. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Shaw, Joseph. 2006a. Intentions and Trolleys. *Philosophical Quarterly*, 56 (222): 63–83.
- Shaw, Joseph. 2006b. Intention in Ethics. *Canadian Journal of Philosophy*, 36 (2): 187–224.
- Sidgwick, Henry. 1981 (1874). *The Methods of Ethics*, 7th edition. Indianapolis: Hackett Publishing.
- Singer, Peter. 1972. Famine, Affluence, and Morality. *Philosophy and Public Affairs*, 1 (3): 229–243.
- Singer, Peter. 2009. Theology's Unintended Consequences. *The Guardian*, <<http://www.theguardian.com/commentisfree/belief/2009/nov/17/katrina-double-effect-doctrine-catholic>>.
- Slote, Michael. 2001. *From Morals to Motives*. New York: Oxford University Press.
- Smith, Ian A. 2007. A New Defense of Quinn's Principle of Double Effect. *Journal of Social Philosophy*, 38 (2): 349–364.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Sorensen, Roy. 2007. Bald-faced Lies! Lying without the Intent to Deceive. *Pacific Philosophical Quarterly*, 88 (2): 251–264.
- Steinboff, Uwe. 2007. *On the Ethics of War and Terrorism*. New York: Oxford University Press.
- Stocker, Michael. 1970a. Morally Good Intentions. *Monist*, 54 (1): 124–141.
- Stocker, Michael. 1970b. Intentions and Act Evaluations. *Journal of Philosophy*, 67 (17): 589–602.
- Strawson, Peter F. 1962. Freedom and Resentment. *Proceedings of the British Academy*, 48 (1): 1–25.
- Steinbock, Bonnie and Nocross, Alastair. 1994. *Killing and Letting Die*, 2nd ed. New York: Fordham University Press.
- Stuchlik, Joshua. 2012. A Critique of Scanlon on Double Effect. *Journal of Moral Philosophy*, 9 (2): 178–199.
- Stump, Eleonore. 1999. Moral Responsibility Alternative Possibilities: The Flicker of Freedom. *Journal of Ethics*, 3 (4): 299–324.
- Suikkanen, Jussi. 2011. Intentions, Blame, and Contractualism. *Jurisprudence*, 2 (2): 561–573.
- Sulmasy, Daniel P. 2007. 'Reinventing the Rule of Double Effect'. In Bonnie Steinbock (ed.), *The Oxford Handbook of Bioethics*. New York: Oxford University Press.
- Sverdlik, Stephen. 2011. *Motive and Rightness*. New York: Oxford University Press.
- Tadros, Victor. 2011a. Wrongdoing and Motivation. In R. A. Duff and Stuart P. Greene (eds.), *Philosophical Foundations of Criminal Law*. New York: Oxford University Press.
- Tadros, Victor. 2011b. *The Ends of Harm: Moral Foundations of Criminal Law*. New York: Oxford University Press.
- Tadros, Victor. 2012. Replies. *Jerusalem Review of Legal Studies*, 5 (1): 89–109.
- Tadros, Victor. 2014a. What Might Have Been. In John Oberdiek (ed.), *Philosophical Foundations of the Law of Torts*. Oxford: Oxford University Press.
- Tadros, Victor. 2014b. Intentions and Permissibility: The Overdetermination Challenge. MS (on file with author) pp. 43.
- Tadros, Victor. 2014c. Intentions and Inchoate Wrongdoing. MS (on file with author) pp. 36.
- Tadros, Victor. 2015a. Wrongful Intentions Without Closeness. *Philosophy and Public Affairs*, 43 (1): 52–74.
- Tadros, Victor. 2015b. Answers. *Criminal Law and Philosophy*, 9 (1): 73–102.
- Thomson, Judith Jarvis. 1976. Killing, Letting Die, and the Trolley Problem. *Monist*, 59 (2): 204–217.
- Thomson, Judith Jarvis. 1985. The Trolley Problem. *Yale Law Journal*, 94 (6): 1395–1415.
- Thomson, Judith Jarvis. 1986. Imposing Risks. In her *Rights, Restitution, and Risk*. Cambridge, MA: Harvard University Press.

- Thomson, Judith Jarvis. 1990. *The Realm of Rights*. Harvard University Press, Cambridge.
- Thomson, Judith Jarvis. 1991. Self-Defense. *Philosophy and Public Affairs*, 20 (4): 283–310.
- Thomson, Judith Jarvis. 1999. Physician Assisted Suicide: Two Moral Arguments. *Ethics*, 109 (3): 497–518.
- Thomson, Judith Jarvis. 2008. Turning the Trolley. *Philosophy and Public Affairs*, 36 (4): 201–212.
- Tobia, Kevin P., Buckwalter, Wesley, Stich, Stephen. 2011. Moral Intuitions: Are Philosophers Experts? *Philosophical Psychology*, 26 (5): 629–638.
- Trankle, Steven A. 2014. Decisions that Hasten Death: Double Effect and the Experiences of Physicians in Australia. *BMC Medical Ethics*, 15 (1): 1–26.
- Uniacke, Suzanne. 1994. *Permissible Killing: Self-Defense Justification of Homicide*. Cambridge: Cambridge University Press.
- Uniacke, Suzanne. 2010. Responsibility, Intention and Consequence. In John Skorupski (ed.), *The Routledge Companion to Ethics*. New York: Routledge.
- Uniacke, Suzanne. 2014. Opportunistic Terrorism. *Journal of Moral Philosophy*, 11 (4): 395–410.
- Vranas, Peter B. M. 2007. I Ought, Therefore I Can. *Philosophical Studies*, 136 (2): 167–216.
- Walen, Alec and Wasserman, David. 2012. Agents, Impartiality, and the Priority of Claims over Duties: Diagnosing Why Thomson Still Gets the Trolley Problem Wrong by Appeal to the “Mechanics of Claims”. *Journal of Moral Philosophy*, 9 (4): 545–571.
- Walen, Alec. 2004. Permissibly Encouraging the Impermissible. *Journal of Value Inquiry*, 38 (3): 341–354.
- Walen, Alec. 2006. The Doctrine of Illicit Intentions. *Philosophy and Public Affairs*, 34 (1): 39–67.
- Walen, Alec. 2009. Comments on Doug Husak: The Low Cost of Recognizing (and of Ignoring) the Limited Relevance of Intentions to Permissibility. *Criminal Law and Philosophy*, 3 (1): 71–78.
- Walen, Alec. 2011. Choosing Your Reasons for an Action: Moral Permissibility, Willful Action, and the Doctrine of Double Effect. MS (on file with author), pp. 28.
- Walen, Alec. 2013. Wrongdoing Without Motives: Why Victor Tadros is Wrong About Wrongdoing and Motivation. *Law and Philosophy*, 32 (2-3): 217–240.
- Walen, Alec. 2014. Transcending the Means Principle. *Law and Philosophy*, 33 (4): 427–466.
- Walzer, Michael. 1977. *Just and Unjust Wars*. New York: Basic Books.
- Wedgwood, Ralph. 2009a. Intrinsic Values and Reasons for Action. *Philosophical Issues*, 19 (1): 321–342.
- Wedgwood, Ralph. 2009b. Defending Double Effect. *PEA Soup*, <<http://peasoup.typepad.com/peasoup/2009/03/defending-double-effect.html>>.
- Wedgwood, Ralph. 2009c. Against Actualism. *PEA Soup*, <<http://peasoup.typepad.com/peasoup/2009/09/against-actualism.html>>.
- Wedgwood, Ralph. 2010. Moral Obligation Applies to Act-Types, Not Act-Tokens. *PEA Soup*, <<http://peasoup.typepad.com/peasoup/2010/02/moral-obligation-applies-to-acttypes-not-acttokens.html>>.
- Wedgwood, Ralph. 2011a. Defending Double Effect. *Ratio*, 24 (4): 384–401.
- Wedgwood, Ralph. 2011b. Scanlon on Double Effect. *Philosophy and Phenomenological Research*, 83 (2): 464–472.
- Wedgwood, Ralph. 2013. Rational ‘Ought’ Implies ‘Can’. *Philosophical Issues*, 23 (1): 70–92.
- Wedgwood, Ralph. 2014. Defending “Defending Double Effect”. *Ralph Wedgwood’s Blog*, <<http://ralphwedgwood.typepad.com/blog/2014/06/defending-defending-double-effect.html>>.

- Westen, Peter. 2011. Is Intent Constitutive of Wrongdoing? In Matthew H. Kramer and Mark R. Reiff (eds.), *Crime, Punishment, and Responsibility: The Jurisprudence of Antony Duff*. New York: Oxford University Press.
- Widerker, David. 1991. Frankfurt on 'Ought Implies Can' and Alternative Possibilities. *Analysis*, 51 (4): 222–224.
- Williams, Bernard. 1973. A Critique of Utilitarianism. In his *Utilitarianism: For and Against*, with J. J. C. Smart. Cambridge: Cambridge University Press.
- Williams, Bernard. 1981. *Moral Luck*. Cambridge: Cambridge University Press.
- Williams, Bernard. 2006 (1985). *Ethics and the Limits of Philosophy*. Abingdon: Routledge.
- Williams, Glanville. 1957. *The Sanctity of Life and the Criminal Law*. New York: Knopf.
- Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Translated by G. E. M. Anscombe. Oxford: Blackwell.
- Woodward, Paul A. 1997. The Importance of the Proportionality Condition to the Doctrine of Double Effect: A Response to Fischer, Ravizza, and Copp. *Journal of Social Philosophy*, 28 (2): 140–152.
- Woodward, Paul A. ed. 2001. *The Doctrine of Double Effect: Philosophers Debate a Controversial Moral Principle*. Notre Dame: University of Notre Dame Press.
- Woollard, Fiona. 2015. *Doing and Allowing Harm*. New York: Oxford University Press.
- Yaffe, Gideon. 2010. *Attempts: In the Philosophy of Action and Criminal Law*. New York: Oxford University Press.
- Zimmerman, Michael J. 1993. Obligation, Responsibility, and Alternative Possibilities. *Analysis*, 53 (1): 51–53.
- Zimmerman, Michael J. 1996. *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.
- Zimmerman, Michael J. 2008. *Living With Uncertainty: The Moral Significance of Ignorance*. Cambridge: Cambridge University Press.